Econometrics I Lecture Notes

Liu Mingyang

September 10, 2025

Contents

1 Econometrics

2 The Linear Regression Model

2.1 The Linear Regression Model

The **multiple linear regression model** is used to study the relationship between a **dependent variable** and one or mode **independent variables**. The generic form is:

$$
\begin{aligned}
y &= f(x_1, x_2, \ldots, x_K) + \varepsilon \\
&= x_1\beta_1 + x_2\beta_2 + \ldots + x_K\beta_K + \varepsilon
\end{aligned}
$$

where:

y: Dependent variable

$x_1, \ldots, x_K$: Explanatory variables.

$\beta_1, \ldots, \beta_K$: Unknown paremeters.

ε: Disturbance.

The function $f(x_1, x_2, \ldots, x_K)$ is called the **population regression equation**.

2.1.1 Some Notations

2.1.1.1 Some Clarification between n and K

Usually, we assume K is the number of the explanatory variables. n is the number of the sample. So we ca say that the **data** form:

$$
(y_i, x_{i1}, x_{i2}, \ldots, x_{iK}), \text{ for } i = 1, 2, \ldots, n
$$

2.1.1.2 Matrix Form for Linear Regression Model

We require that all vectors are column vectors. For observation i, we define:

$$
x_i = \begin{pmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{iK} \end{pmatrix}; \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_K \end{pmatrix}
$$

The model of this observation:

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

As for sample, we define:

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad X = \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_K' \end{pmatrix}, \qquad \varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_K \end{pmatrix}$$

The model of this sample set is:

$$Y = X\beta + \varepsilon$$

2.2 Assumptions of the Linear Regression Model

2.2.1 Assumption 1: Linearity

A.1

The data satisfy:

$$Y = X\beta + \varepsilon$$

which means that y_i is a linear function of β_j. The linearity refers to the manner in which the parameters and the disturbance enter the equation, not necessarily to the relationship among the variables. For example:

$$y = \beta_1 + \beta_2 x + \varepsilon$$
$$\ln y = \beta_1 + \beta_2 \ln x + \varepsilon$$

are all linear model.

2.2.2 Assumption 2: Full Rank

A.2

There are no exact linear relationships among the variables:

   X is an $n \times K$ matrix with rank K.

which means there are at least K observations.

2.2.3 Assumption 3: Exogeneity of the explanatory variables

A.3

The disturbance is assumed to have conditional expected value zero at every observation:

$$\mathbb{E}[\boldsymbol{\varepsilon}|X] = \begin{bmatrix} \mathbb{E}[\varepsilon_1|X] \\ \mathbb{E}[\varepsilon_2|X] \\ \vdots \\ \mathbb{E}[\varepsilon_n|X] \end{bmatrix} = \mathbf{0}$$

which means that there is no information $\mathbb{E}[\varepsilon_i|\cdot]$ contained in any observation on $\boldsymbol{x}_j$.

And also, we assume that the disturbances convey no information about each other:

$$\mathbb{E}[\varepsilon_i|\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{i-1}, \varepsilon_{i+2}, \ldots, \varepsilon_n] = 0$$

Proposition 2.2.1

The unconditional expectation of disturbance is also 0:

$$\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbf{0}$$

which means that:

$$\mathbb{E}[Y|X] = X\boldsymbol{\beta}$$

Proof. Here, we use the **law of iterated expectation**:

$$\mathbb{E}[\boldsymbol{\varepsilon}] = \mathbb{E}_X\{\mathbb{E}[\boldsymbol{\varepsilon}|X]\} = \mathbb{E}_X[0] = \mathbf{0}$$

Then, we have:

$$\begin{aligned} \mathbb{E}[Y|X] &= \mathbb{E}[X\boldsymbol{\beta} + \boldsymbol{\varepsilon}|X] \\ &= \mathbb{E}[X\boldsymbol{\beta}|X] + \mathbb{E}[\boldsymbol{\varepsilon}|X] \\ &= \mathbb{E}[X\boldsymbol{\beta}|X] \\ &= X\boldsymbol{\beta} \end{aligned}$$

$\square$

2.2.4 Assumption 4: Homoscedastic and Nonautocorrelated Disturbance

A.4

The variances and covariances of the disturbances satisfy:

$$Var[\varepsilon_i|X] = \sigma^2, \ \forall \, i$$
$$Cov[\varepsilon_i, \varepsilon_j|X] = 0, \ \forall \, i \neq j$$

The **Assumption 4** can be rewritten as:

$$\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X] = \sigma^2 \boldsymbol{I}$$

which means:

$$Var[\boldsymbol{\varepsilon}|X] = \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X] - \mathbb{E}[\boldsymbol{\varepsilon}|X]\mathbb{E}[\boldsymbol{\varepsilon}'|X]$$
$$= \mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X]$$

Then we can have:

$$Var[\boldsymbol{\varepsilon}] = \mathbb{E}\{Var[\boldsymbol{\varepsilon}|X]\} + Var\{\mathbb{E}[\boldsymbol{\varepsilon}|X]\} = 0$$

2.2.5 Assumption 5: Data Generating Process

A.5

We allow X can be fixed or random.

2.2.6 Assumption 6: Normality

A.6

We assume that the disturbances are normally distributed, with zero mean and constant variance:

$$\boldsymbol{\varepsilon}|X \sim N[\boldsymbol{0}, \sigma^2 \boldsymbol{I}]$$

3 Least Square Regression

3.1 Least Square Regression

A normal model:

$$Y = X'\beta + \varepsilon$$

The difference of some notions:

β, ε: population regression:

$$E[Y|X] = X\beta$$
$$\varepsilon = Y - X\beta$$

b, e: sample estimates:

$$\hat{Y} = Xb$$
$$e = Y - Xb$$

How to measure the closeness of the sample estimates, a usual tools is **least square**.

3.2 The Least Square Coefficient Vector

The criterion is the sum of squared residuals:

$$\sum_{i=1}^{n} e_{i0}^2 = \sum_{i=1}^{n} (y_i - x_i'b_0)$$

The least square coefficient vector b_0 minimizes the sum of squared residuals in the form of matrix:

$$\min_{b_0} S(b_0) = e_0'e_0 = (Y - Xb_0)'(Y - Xb_0)$$

3.2.1 Matrix Derivatives

Definition 3.2.1

Let z be a $k \times 1$ column vector and $f(z)$ be a scalar. Then:

$$\frac{\partial f(z)}{\partial z} = \begin{pmatrix} \frac{\partial f(z)}{\partial z_1} \\ \frac{\partial f(z)}{\partial z_2} \\ \vdots \\ \frac{\partial f(z)}{\partial z_k} \end{pmatrix}$$

Proposition 3.2.2

Let a be a $k \times 1$ column vector and A be a $k \times k$ matrix. Then:

$$\frac{\partial a'z}{\partial z} = a; \qquad \frac{\partial z'Az}{\partial z} = (A + A')z$$

Proof.

i) We first consider the column vector:

$$a'z = \sum_{i=1}^{k} a_i z_i$$

Then:

$$\frac{\partial a'z}{\partial z} = \begin{pmatrix} \frac{\partial \sum_{i=1}^{k} a_i z_i}{\partial z_1} \\ \frac{\partial \sum_{i=1}^{k} a_i z_i}{\partial z_2} \\ \vdots \\ \frac{\partial \sum_{i=1}^{k} a_i z_i}{\partial z_k} \end{pmatrix} = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_k \end{pmatrix} = a$$

i) Now we consider the matrix:

$$z'Az = \sum_{n=1}^{k} z_n \sum_{i=1}^{k} a_{ni} z_i = \sum_{n=1}^{k} \sum_{i=1}^{k} z_n a_{ni} z_i$$

Then we have:

$$\frac{\partial z' A z}{\partial z} = \begin{pmatrix} \frac{\partial \sum_{n=1}^{k} \sum_{i=1}^{k} z_n a_{ni} z_i}{\partial z_1} \\ \frac{\partial \sum_{n=1}^{k} \sum_{i=1}^{k} z_n a_{ni} z_i}{\partial z_2} \\ \vdots \\ \frac{\partial \sum_{n=1}^{k} \sum_{i=1}^{k} z_n a_{ni} z_i}{\partial z_k} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{n=1}^{k} z_n a_{n1} + \sum_{i=1}^{k} z_i a_{1i} \\ \sum_{n=1}^{k} z_n a_{n2} + \sum_{i=1}^{k} z_i a_{2i} \\ \vdots \\ \sum_{n=1}^{k} z_n a_{nk} + \sum_{i=1}^{k} z_i a_{ki} \end{pmatrix}$$

$$= \begin{pmatrix} \sum_{n=1}^{k} z_n a_{n1} \\ \sum_{n=1}^{k} z_n a_{n2} \\ \vdots \\ \sum_{n=1}^{k} z_n a_{nk} \end{pmatrix} + \begin{pmatrix} \sum_{i=1}^{k} z_i a_{1i} \\ \sum_{i=1}^{k} z_i a_{2i} \\ \vdots \\ \sum_{i=1}^{k} z_i a_{ki} \end{pmatrix}$$

$$= A' z + A z$$

$$= (A' + A) z$$

$\square$

3.2.2   Solution of this optimal problem

$$e_0' e_0 = Y'Y - b_0' X'Y - Y'X b_0 + b_0' X'X b_0$$
$$= Y'Y - 2Y'X b_0 + b_0' X'X b_0 \quad (Y'X b_0 \text{ is a number.})$$

The necessary condition:

$$\frac{\partial S(b_0)}{\partial b_0} = -2X'Y + 2X'X b_0 = 0$$

11

If the inverse of $X'X$ exists, then:

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

The second derivatives matrix:

$$\frac{\partial^2 S(b_0)}{\partial b_0 \partial b_0'} = 2X'X$$

If X is full rank, the second derivatives is positive. Then b is unique.

3.2.3 Algebraic aspects of the least squares solution

Considering the optimal condition:

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

For the first column of X which is $\mathbf{1}$s, we have that:

$$x_1'e = \sum_i e_i = 0$$

There are 3 implications:

    i) The least squares residuals sum to zero.

    ii) The regression hyperplane passes through the point of means of the data:

$$y_i = x_i'b + e_i$$
$$\sum_i y_i = \sum_i x_i'b + \sum_i e_i$$
$$\bar{y} = \bar{x}'b$$

    iii) The mean of the fitted values from the regression equals the mean of the actual values.

3.2.4 Projection

From the least squares residuals:

$$e = Y - Xb$$

From the optimal condition:

$$e = Y - X(X'X)^{-1}X'Y$$
$$= [I - X(X'X)^{-1}X']Y$$

Here we can give a definition.

Definition 3.2.3 (*Residual Maker*)

Let the $n \times K$ full column rank matrix, X be composed of columns $(x_1, x_2, \ldots, x_K)$, and let Y be an $n \times 1$ column vector. The matrix:

$$M = I - X(X'X)^{-1}X'$$

is a 'residual maker' in that when M premultiplies a vector, Y, the result, MY, is the column vector of residuals in the least squares regression of Y on X.

Proposition 3.2.4

The residual maker is symmetric and idempotent.

$$M' = [I - X(X'X)^{-1}X']' = I - X(X'X)^{-1}X'$$
$$= M$$
$$M^2 = [I - X(X'X)^{-1}X']'[I - X(X'X)^{-1}X']$$
$$= I - X(X'X)^{-1}X' - X(X'X)^{-1}X' + X(X'X)^{-1}X'$$
$$= M$$

Corollary 3.2.5

From the definition of residual maker, we know that:

$$MX = 0$$

The intuition is that if we regress X on X, the residual will be zero.

Now we can divide Y into 2 parts, the fitted value and residuals:

$$Y = \hat{Y} + e$$
$$= Xb + MY$$

Proposition 3.2.6

The fitted value and residuals are orthogonal. Proof as follows:

$$(MY)' \times Xb = Y'M'Xb = Y'MXb = 0$$

Definition 3.2.7 (*Projection Matrix*)

From the definition we have that:

$$\hat{Y} = Y - e = IY - MY = (I - M)Y = X(X'X)^{-1}X'Y$$

so, we can define the projection matrix as follows:

$$P = X(X'X)^{-1}X'$$

Proposition 3.2.8

P is symmetric and idempotent.

$$P' = [X(X'X)^{-1}X']' = X(X'X)^{-1}X'$$
$$= P$$
$$P^2 = [X(X'X)^{-1}X']'[X(X'X)^{-1}X']$$
$$= X(X'X)^{-1}X'$$
$$= P$$

Corollary 3.2.9

Similar as the residual maker, we have that:

$$PX = X$$

Corollary 3.2.10

If we combine P and M together:

$$PM = MP = 0$$

Corollary 3.2.11

Because $\hat{Y} = PY$ and $e = MY$, form **Corollary 3.2.10**, we have:

$$\begin{aligned}
\hat{Y}'e &= (PY)'MY \\
&= Y'PMY \\
&= 0
\end{aligned}$$

which means that the fitted values and the residuals are orthogonal. That is to say, P maps Y to a vector in the column space of X, $\mathscr{C}(X)$, that is closest to Y. And this is along with the **Pythagorean theorem**:

$$\begin{aligned}
||Y||^2 &= Y'Y \\
&= (PY + MY)'(PY + MY) \\
&= Y'P'PY + Y'P'MY + Y'M'PY + Y'M'MY \\
&= Y'P'PY + Y'M'MY \\
&= ||\hat{Y}||^2 + ||e||^2
\end{aligned}$$

3.3 Partitioned Regression and Partial Regression

Suppose we divide the variables X into two sets X_1 and X_2:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

To solve the least squares problem:

$$X_1'X_1b_1 + X_1'X_2b_2 = X_1'Y \tag{1}$$
$$X_2'X_1b_1 + X_2'X_2b_2 = X_2'Y \tag{2}$$

From (1) we have that:

$$\begin{aligned}
b_1 &= (X_1'X_1)^{-1}X_1'Y - (X_1'X_1)^{-1}X_1'X_2b_2 \\
&= (X_1'X_1)^{-1}X_1'(Y - X_2b_2)
\end{aligned}$$

If $X'_1 X_2 = 0$, then

$$b_1 = (X'_1 X_1)^{-1} X'_1 Y \qquad (3)$$

which is the same as the regression which only regress on X_1, so we can give the theorem.

Theorem 3.3.1
If $X'_1 X_2 \neq 0$, then we consider (2), insert it into (3):

$$X'_2 X_1 (X'_1 X_1)^{-1} X'_1 (Y - X_2 b_2) + X'_2 X_2 b_2 = X'_2 Y$$

So we can have:

$$b_2 = \{X'_2 [I - X_1 (X_1 X_1)^{-1} X'_1] X_2\}^{-1} \{X'_2 [I - X_1 (X'_1 X_1)^{-1} X'_1] Y\}$$
$$= (X'_2 M_1 X_2)^{-1} (X'_2 M_1 Y)$$

If we define $X^\star_2 = M_1 X_2$ and $Y^\star = M_1 Y$, then we can rewrite it:

$$b_2 = (X^{\star\prime}_2 X^\star_2)^{-1} X^{\star\prime}_2 Y^\star$$

From this process, we can give the theorem.

Theorem 3.3.2 (*Frisch − Waugh − Lovell Theorem*) In the linear least squares regression of vector Y on two sets of variables, X_1 and X_2, the subvector b_2 is the set of coefficients obtained when the residuals from a regression of Y on X_1 alone are regressed on the set of residuals obtained when each column of X_2 is regressed on X_1.

Here we give a proof.

Proof.
We set:

$$P = X(X'X)^{-1} X'$$
$$P_1 = X_1 (X_1' X_1)^{-1} X_1'$$
$$M = I - M$$
$$M_1 = I - M_1$$

And we have that:

$$P_1 P = PP_1 = P_1$$
$$MM_1 = M_1 M = M$$

Thus, we know that:

$$Y = Xb + e = X_1 b_1 + X_2 b_2 + MY$$

Multiply both side by $X_2' M_1$:

$$X_2' M_1 Y = X_2' M_1 X_1 b_1 + X_2' M_1 X_2 b_2 + X_2' M_1 MY$$
$$= X_2' M_1 X_2 b_2$$

Thus:

$$b_2 = (X_2' M_1 X_2)^{-1} X_2' M_1 Y$$
$$= (X_2' M_1' M_1 X_2)^{-1} X_2' M_1' M_1 Y$$
$$= (X_2^{\star'} X_2^{\star})^{-1} X_2^{\star'} Y^{\star}$$

$\square$

With **Theorem 3.3.2**, we can give corollary.

Corollary 3.3.3 (*Individual Regression Coefficient*)

The coefficient on z in a multiple regression of Y on $W = [X, z]$ is computed as:

$$c = (z' M_X z)^{-1} (z' M_X Y) = (z^{\star'} z^{\star})^{-1} z^{\star'} Y^{\star}$$

here $z^{\star}$ and $Y^{\star}$ are the residual vectors from least squares regressions of z and y on X; $z^{\star} = M_X z, y^{\star} = M_X Y$.

Corollary 3.3.4
Theorem 3.3.5

3.4 Partial Regression and Partial Correlation Coefficients

Theorem 3.4.1 (*Diagonal Elements of the Inverse of a Moment Matrix*)

Let W denote the partitioned matrix $[X, z]$, that is the K columns of X plus an additional

column labeled z. The last diagonal element of $(W'W)^{-1}$ is:

$$(W'W)^{-1}_{K+1,K+1} = (z'M_X z)^{-1} = (z^{\star\prime}z^{\star})^{-1}$$

where:

$$z^{\star} = M_X z$$
$$M_X = I - X(X'X)^{-1}X'$$

The proof of **Theorem 3.4.1** is that:

Theorem 3.4.2

If $e'e$ is the sum of squared residual when Y is regressed on X and $u'u$ is the sum of squared residuals when Y is regressed on X and z, then

$$u'u = e'e - c^2(z^{\star\prime}z^{\star}) \leq e'e$$

where c is the coefficient on z in the long regression of Y on $[X, z]$ and $z^{\star} = Mz$ is the vector of residuals when z is regressed on X

Proof. In the long regression, we have that:

$$u = Y - Xd - zc \tag{4}$$

And we have that:

$$d = (X'X)^{-1}X'(Y - zc)$$
$$= b - (X'X)^{-1}X'zc \tag{5}$$

If we insert (5) into (4):

$$u = Y - X[b - (X'X)^{-1}X'zc] - zc$$
$$= e - M_X zc$$
$$= e - z^{\star}c$$

So we have that:

$$u'u = e'e + c^2(z^{\star\prime}z^{\star}) - 2c(z^{\star\prime}e)$$

18

From **Corollary 3.3.3**, we know that:

$$z^{\star\prime}e = z^{\star\prime}M_X Y$$
$$= c(z^{\star\prime}z^{\star})$$

So we have:

$$u'u = e'e - c^2(z^{\star\prime}z^{\star}) \leq e'e$$

$\square$

Let c and u denote the coefficient on z and the vector of the residuals in the multiple regression of Y on $W = [X, z]$.

From the definition of t ratio we know that:

$$t_z = \frac{c}{SD(c)}$$

Here because we don't know $Var(b)$, so we use the sum residuals:

$$Var(u) = \frac{u'u}{n - K + 1}$$

And we know that:

$$Var(b) = Var[(W'W)^{-1}W'Y]$$
$$= Var[(W'W)^{-1}W'(Wb + u)]$$
$$= Var[(W'W)^{-1}W'u]$$
$$= \frac{u'u}{n - K + 1}[(W'W)^{-1}W'][(W'W)^{-1}W']'$$
$$= \frac{u'u}{n - K + 1}(W'W)^{-1}$$

So we have:

$$Var(c) = \frac{u'u}{n - K + 1}(W'W)^{-1}_{K+1, K+1}$$

And because of **Theorem 3.4.1**, we know that:

$$Var(c) = \frac{u'u}{n - K + 1}(z^{\star\prime}z^{\star})^{-1}$$

19

From **Corollary 3.3.3** we know that:

$$c = (z^{\star\prime}z^{\star})^{-1}z^{\star\prime}Y^{\star}$$

So we can have:

$$
\begin{aligned}
t_z^2 &= \frac{[(z^{\star\prime}z^{\star})^{-1}z^{\star\prime}Y^{\star}]^2}{\frac{u'u}{n-K+1}(z^{\star\prime}z^{\star})^{-1}} \\
&= \frac{(z^{\star\prime}Y^{\star})^2 DF}{(u'u)(z^{\star\prime}z^{\star})}
\end{aligned}
$$

So we have that:

$$
\begin{aligned}
\frac{t_z^2}{t_z^2 + DF} &= \frac{(z^{\star\prime}Y^{\star})^2}{(z^{\star\prime}Y^{\star})^2 + (u'u)(z^{\star\prime}z^{\star})} \\
&= \frac{\frac{(z^{\star\prime}Y^{\star})^2}{(z^{\star\prime}z^{\star})(Y^{\star\prime}Y^{\star})}}{\frac{(z^{\star\prime}Y^{\star})^2}{(z^{\star\prime}z^{\star})(Y^{\star\prime}Y^{\star})} + \frac{(u'u)(z^{\star\prime}z^{\star})}{(z^{\star\prime}z^{\star})(Y^{\star\prime}Y^{\star})}} \\
&= \frac{r_{yz}^{\star 2}}{r_{yz}^{\star 2} + \frac{u'u}{Y^{\star\prime}Y^{\star}}}
\end{aligned}
$$

And from **Theorem 3.4.2** we know that:

$$
\begin{aligned}
\frac{u'u}{Y^{\star\prime}Y^{\star}} &= \frac{e'e - c^2(z^{\star\prime}z^{\star})}{Y^{\star\prime}Y^{\star}} \\
&= \frac{Y^{\star\prime}Y^{\star} - \frac{(z^{\star\prime}Y^{\star})^2}{z^{\star\prime}z^{\star}}}{Y^{\star\prime}Y^{\star}} \\
&= 1 - r_{yz}^{\star 2}
\end{aligned}
$$

So we have:

$$r_{yz}^{\star 2} = \frac{t_z^2}{t_z^2 + DF}$$

3.5 Goodness of Fit and the Analysis of Variance

One problem about the sum of squared residuals is that it can be influenced by the scale of the data, so we need another method to measure the goodness of fit.

The variation of the depend variable is defined in terms of deviations from its mean.

We can compute the total variation:

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

If the regression contains a constant term, from **3.2.2** we know that the regression go through the mean of the data:

$$\bar{y} = \bar{x}'b$$

Definition 3.5.1
Here we give the definition of the matrix, which can compute deviation from the mean of a vector:

$$M_0 = I - \iota(\iota'\iota)^{-1}\iota', \qquad \iota = (1, 1, \ldots, 1)'$$

So we have:

$$M_0Y = M_0Xb + M_0e$$

We know that $M_0e = e$. Then:

$$e'M_0X = e'X = 0$$

So we have:

$$(M_0Y)'M_0Y = b'X'M_0Xb + e'e$$

This equation can be written as:

$$SST = SSR + SSE$$

Definition 3.5.2 (*Coefficient of Determination*)
The coefficient of determination R^2 is defined as:

$$\frac{SSR}{SST} = \frac{b'X'M_0Xb}{Y'M_0Y} = 1 - \frac{e'e}{Y'M_0Y} = 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Also, we know that:

$$\frac{SSR}{SST} = \frac{\hat{Y}' M_0 (Y - e)}{Y' M_0 Y}$$

$$= \frac{\hat{Y}' M_0 Y - b' X' M_0 e}{Y' M_0 Y}$$

$$= \frac{\hat{Y}' M_0 Y - b' X' e}{Y' M_0 Y}$$

$$= \frac{\hat{Y}' M_0 Y}{Y' M_0 Y} \times \frac{\hat{Y}' M_0 Y}{\hat{Y}' M_0 \hat{Y}}$$

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

3.5.1   The Adjusted R-squared and A Measure of Fit

Theorem 3.5.3

Let R^2_{Xz} be the coefficient of determination in the regression of Y on X and an additional variable z, let R^2_X be the same regression of Y on X alone, and $r^{\star 2}_{yz}$ be the partial correlation between Y and z, controlling for X. Then:

$$R^2_{Xz} = R^2_X + (1 - R^2_X) r^{\star 2}_{yz}$$

Proof. We know that:

$$u' u = e' e - \frac{(z^{\star\prime} Y^\star)^2}{z^{\star\prime} z^\star}$$

$$= e' e (1 - r^{\star 2}_{yz})$$

Then we divide both side by $Y' M_0 Y$:

$$(1 - R^2_{Xz}) = (1 - R^2_X)(1 - r^{\star 2}_{yz})$$

Then we have:

$$R^2_{Xz} = R^2_X + (1 - R^2_X) r^{\star 2}_{yz}$$

$\square$

So we can see that R^2 cannot be smaller in a longer regression. It is tempting to exploit R^2 by adding variables until $R^2 = 1$. One way to avoid this problem is **adjusted** R^2,

which is[1]:

$$\bar{R}^2 = 1 - \frac{e'e/(n-K)}{Y'M_0Y/(n-1)}$$
$$= 1 - \frac{n-1}{n-K}(1 - R^2)$$

Theorem 3.5.4

In a multiple regression, $\bar{R}^2$ will fall (rise) when the variable x is deleted from the regression if the square of the t ratio associated with this variable is greater (less) than 1.

3.5.2   R-squared and the Constant term in the model

3.5.3   Comparing Models

3.6   Linearly Transformed Regression

Theorem 3.6.1

In the linear regression of Y on $Z = XA$ where A is a nonsingular matrix that transforms the columns of X, the coefficient will equal $A^{-1}b$ where b is the vector of coefficients in the linear regression of Y on X, and the R^2 will be identical.

Proof. Here we can give the new coefficients:

$$d = (Z'Z)^{-1}Z'Y$$
$$= [(XA)'(XA)]^{-1}(XA)'Y$$
$$= A^{-1}(X'X)^{-1}A'^{-1}A'X'Y$$
$$= A^{-1}(X'X)^{-1}X'Y$$
$$= A^{-1}b$$

So we can find that:

$$u = Y - Z(A^{-1}b)$$
$$= Y - XAA^{-1}b$$
$$= Y - Xb$$
$$= e$$

---

[1]Here K includes the constant term.

Since the residuals are the same we can say that R^2 is identical. $\qquad\square$

4 The Asymptotic Property

4.1 Motivation

With the assumption **A.1, A.2, A.3, A.4**, from the previous lecture, we know that the least squares estimator, b, is **B.L.U.E**.

In particular, the unbiasedness gives us that b is 'on average' correctly estimates β. And we also know that the variance of b is $\sigma^2(X'X)^{-1}$, which gives us the accuracy of b. However, if we want to conduct statistical inference on β, we need to know the distribution of b. If we have assumption **A.6**, we can see that:

$$b = \beta + (X'X)^{-1}X'\varepsilon \sim N(\beta, \sigma^2(X'X)^{-1})$$

In this lecture, we can see that **under the large number assumption**, we don't need **A.6** to get the distribution of b.

4.2 Mathematic Basis

Definition 1.2.1 (Convergence in Probability)

A sequence of random variables $x_1, x_2, \ldots$ is said to converge to c in probability if, for every $\varepsilon > 0$:

$$\mathbb{P}(|x_n - c| < \varepsilon) \to 1, \text{ as } n \to \infty$$

which is written as $x_n \xrightarrow{p} c$, or $\text{plim}_{n\to\infty} x_n = c$.

Theorem 1.2.2 (Chevychev's inequality)

If x is a random variable and c and ε are constant, then:

$$\mathbb{P}(|x - c| \geq \varepsilon) \leq \mathbb{E}[(x - c)^2]/\varepsilon^2$$

Lemma 1.2.3

If $E(x_n) \to c$ and $Var(x_n) \to 0$ as $n \to \infty$, then $x_n \xrightarrow{p} c$.

Definition 1.2.4 (Consistency)

An estimator $\hat{\theta}_n$ of θ is called consistent if:

$$\hat{\theta}_n \xrightarrow{p} \theta, \text{ as } n \to \infty$$

Lemma 1.2.5

Let $x_i \sim i.i.d.(\mu, \sigma^2)$, for $i = 1, \ldots, n$, and assume μ is the parameter of interest. Then, $\hat{\mu} = (1/n) \sum_{i=1}^{n} x_i$ is a consistent estimator of μ.

Proof.

We use **Lemma 1.2.3**. First for the expectation:

$$\mathbb{E}[\hat{\mu}] = \mathbb{E}\left[(1/n) \sum_{i=1}^{n} x_i \right]$$

$$= (1/n) \sum_{i=1}^{n} \mathbb{E}[x_i]$$

$$= \mu$$

Then we consider the variance:

$$Var[\hat{\mu}] = Var\left[(1/n) \sum_{i=1}^{n} x_i \right]$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} Var[x_i]$$

$$= \frac{1}{n} \sigma^2$$

As $n \to \infty$, $Var[\hat{\mu}] \to 0$.

Then $\hat{\mu} = (1/n) \sum_{i=1}^{n} x_i$ is a consistent estimator of μ. $\qquad\qquad \square$

Theorem 1.2.6 (i.i.d. Law of Large Number)

If x_i, for $i = 1, 2, \ldots, n$ is i.i.d. with $\mathbb{E}(x_i) = \mu$, then $(1/n) \sum_{i=1}^{n} x_i \overset{p}{\to} \mu$ as $n \to \infty$.

Theorem 1.2.7 (Slutsky's Theorem)

If $x_n \overset{p}{\to} c$ and $f(x)$ is continuous at c, then $f(x_n) \overset{p}{\to} f(c)$.

Theorem 1.2.8 (Rules for Probability Limits)

Suppose X_n, Y_n, W_n are random matrices such that $X_n \overset{p}{\to} A$, $Y_n \overset{p}{\to} B$, $W_n \overset{p}{\to} C$, and provided C^{-1} exists:

$$X_n + Y_n \overset{p}{\to} A + B$$

$$X_n Y_n \overset{p}{\to} AB$$

$$W^{-1} \overset{p}{\to} C^{-1}$$

Definition 1.2.9 (Convergence in Distribution)

Let $x_1, x_2, \ldots$ be a sequence of random variables with distribution function $F_n(z) = Pr(x_n \leq z)$. Let x be a random variable with distribution function $F(z)$. Then x_n is said to converge to x in distribution if:

$$\lim_{n \to \infty} |F_n(z) - F(z)| = 0$$

for every z at which $F(z)$ is continuous.

Theorem 1.2.10 (Rules for Limiting Distributions)

If $x_n \xrightarrow{d} x$ and $y_n \xrightarrow{p} c$ (non-random), then:

$$x_n y_n \xrightarrow{d} cx$$

$$x_n + y_n \xrightarrow{d} x + c$$

$$x_n / y_n \xrightarrow{d} x/c, \text{ if } c \neq 0$$

If $x_n \xrightarrow{d} x$ and $g(x_n)$ is a continuous function, then:

$$g(x_n) \xrightarrow{d} g(x)$$

Theorem 1.2.11 (Central Limit Theorem for i.i.d. observations)

Suppose $x_1, x_2, \ldots$ are i.i.d. $(k \times 1)$ vectors with mean μ and variance Σ. Let $\bar{x}_n = (1/n) \sum_{i=1}^{n} x_i$. Then:

$$\sqrt{n}(\bar{x}_n - \mu) \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

4.3   Consistency of the Least Squares Estimator

Now we come to the analysis about the consistency of the least squares estimator. We only use **A.1**, **A.3**, and **A.4**, together with the following assumption:

Assumption A.7 (Random Sampling)

$(y_i, X_i, \varepsilon_i)$, $i = 1, 2, \ldots, n$ is a sequence of independent and identically distributed observations with $\mathbb{E}[x_i x_i'] = Q$, and Q^{-1} exists. Further, $\mathbb{E}|x_{i\iota}|^4 < \infty$, for $1 \leq \iota \leq k$, and $\mathbb{E}|\varepsilon_i|^4 < \infty$.

Theorem 1.3.1

Under the assumption of **A.1**, **A.3**, **A.4**, and **A.7**, b is consistent for β.

Proof.

First, we know that:

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

And we have that:

$$
X'X = \begin{pmatrix} x_1 & x_2 & \dots & x_n \end{pmatrix} \begin{pmatrix} x_1' \\ x_2' \\ \vdots \\ x_n' \end{pmatrix} = \sum_{i=1}^{n} x_i x_i'
$$

$$
X'\varepsilon = \sum_{i=1}^{n} x_i \varepsilon_i
$$

Then, we can rewrite it:

$$
\begin{aligned}
b &= \beta + \left(\frac{X'X}{n}\right)^{-1} \frac{X'\varepsilon}{n} \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \left(\frac{1}{n}\sum_{i=1}^{n} x_i \varepsilon_i\right)
\end{aligned}
$$

First, with **A.7**, we can see that x_i is i.i.d., with $\mathbb{E}[x_i x_i'] = Q$. Then by the **Theorem 1.2.6 (i.i.d. Law of Large Number)**, we can see that:

$$
\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i'\right)^{-1} \xrightarrow{p} Q^{-1}
$$

Second, with **A.7**, we can see that $x_i \varepsilon_i$ is i.i.d., and from the **A.3**, we can see that:

$$
\begin{aligned}
\mathbb{E}[\varepsilon_i^2 | x_i] &= \sigma^2 \\
\mathbb{E}[\varepsilon_i^2] &= \sigma^2
\end{aligned}
$$

For $l = 1, 2, \ldots, k$, we have that:

$$\mathbb{E}|x_{il}\varepsilon_i| \le (\mathbb{E}|x_{il}|^2)^{1/2}(\mathbb{E}|\varepsilon_i^2|^2)^{1/2} < \infty$$

Then we can say that $\mathbb{E}[\boldsymbol{x_i}\varepsilon_i]$ exists, as $\mathbb{E}|\boldsymbol{x_i}\varepsilon_i| < \infty$.
Then, from **A.3** we have that:

$$\begin{aligned}
\mathbb{E}[\varepsilon_i|\boldsymbol{x}_i] &= 0 \\
\mathbb{E}[\boldsymbol{x_i}\varepsilon_i] &= \mathbb{E}_x[\boldsymbol{x}_i\mathbb{E}[\varepsilon_i|\boldsymbol{x}_i]] \\
&= \mathbb{E}_x[\boldsymbol{x}_i 0] \\
&= 0
\end{aligned}$$

Thus, by the **Theorem 1.2.6 (i.i.d. Law of Large Number)**, we can see that:

$$\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i}\varepsilon_i \xrightarrow{p} 0$$

Then, by **Theorem 1.2.6 (Rules for Probability Limits)**, we can say that:

$$\boldsymbol{b} \xrightarrow{p} \boldsymbol{\beta}$$

$\square$

4.4 Asymptotic Property of the Least Squares Estimator

Theorem 1.4.1

Suppose assumptions **A.1**, **A.3**, **A.4**, and **A.7** hold. Then, as $n \to \infty$,

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q}^{-1})$$

where $\boldsymbol{Q} = \mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i']$.

Proof.
Recall that:

$$\boldsymbol{b} = \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon}$$

Then we have that:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) = \left(\frac{X'X}{n}\right)^{-1} \frac{X'\boldsymbol{\varepsilon}}{\sqrt{n}}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i x_i}'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{x_i}\varepsilon_i$$

From the proof of **Theorem1.3.1**, we know that $\boldsymbol{x_i}\varepsilon_i$ is i.i.d., and:

$$\left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x_i x_i}'\right)^{-1} \xrightarrow{p} (\mathbb{E}[\boldsymbol{x_i x_i}'])^{-1} = \boldsymbol{Q}^{-1}$$

$$\mathbb{E}[\boldsymbol{x_i}\varepsilon_i] = 0$$

Then, from assumption **A.7**, we have that:

$$\mathbb{E}[x_{il}]^4 < \infty, \ \forall \ 1 \leq l \leq k$$

$$\mathbb{E}|\varepsilon_i|^4 < \infty$$

Then, by Cauchy-Schwartz inequality, we have:

$$\mathbb{E}|\varepsilon_i^2 \boldsymbol{x_i x_i}'| \leq \sqrt{\mathbb{E}[x_{il}]^4}\sqrt{\mathbb{E}|\varepsilon_i|^4} < \infty$$

Then, we can say that $\mathbb{E}[\varepsilon_i^2 \boldsymbol{x_i x_i}']$ exists. With assumption **A.3**, we have:

$$Var[\boldsymbol{x_i}\varepsilon_i] = \mathbb{E}[\varepsilon_i^2 \boldsymbol{x_i x_i}']$$

$$= \mathbb{E}[\boldsymbol{x_i x_i}'\mathbb{E}[\varepsilon_i^2|\boldsymbol{x_i}]]$$

$$= \sigma^2\mathbb{E}[\boldsymbol{x_i x_i}']$$

$$= \sigma^2\boldsymbol{Q}$$

Thus, by **Theorem 1.2.11 (Central Limit Theorem for i.i.d. observations)**, we can say that:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} \boldsymbol{x_i}\varepsilon_i \xrightarrow{d} N(\boldsymbol{0}, \sigma^2\boldsymbol{Q})$$

Then, by **Theorem 1.2.10 (Rules for Limiting Distributions)**, we can see that:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2\boldsymbol{Q}^{-1})$$

$\square$

4.4.1 The Estimator of the Asymptotic Variance of b

We want to use $s^2(X'X/n)^{-1}$ can be a consistent estimator of $\sigma^2 Q^{-1}$. Here, we give the proof:

Proof.

$$
\begin{aligned}
s^2(X'X/n)^{-1} &= \frac{1}{n-k} \boldsymbol{\varepsilon}' \boldsymbol{M} \boldsymbol{\varepsilon} (X'X/n)^{-1} \\
&= \frac{1}{n-k} [\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon} - \boldsymbol{\varepsilon}' X (X'X)^{-1} X' \boldsymbol{\varepsilon}] (X'X/n)^{-1} \\
&= \frac{n}{n-k} \left[\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n} - \frac{\boldsymbol{\varepsilon}' X}{n} \left(\frac{X'X}{n} \right)^{-1} \frac{X'\boldsymbol{\varepsilon}}{n} \right] (X'X/n)^{-1}
\end{aligned}
$$

Then, we can see that:

$$
\begin{aligned}
\frac{n}{n-k} &\to 1 \\
\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{n} &\xrightarrow{P} \sigma^2 \\
\frac{X'X}{n} &\xrightarrow{P} Q \\
\frac{\boldsymbol{\varepsilon}' X}{n} &\xrightarrow{P} 0 \\
X'X/n &\xrightarrow{P} Q
\end{aligned}
$$

Then, by **Theorem 1.2.8 (Rules for Probability Limits)**, we have:

$$
s^2(X'X/n)^{-1} \xrightarrow{P} \sigma^2 Q^{-1}
$$

$\square$

4.4.2 Asymptotic Distribution of a Function of b

Theorem 1.4.2 (Delta-method) (Need Proof)
Suppose z is a sequence of k-vector random variables such that $\sqrt{n}(z - \mu) \xrightarrow{d} N(0, \Sigma)$, as $n \to \infty$. Let $c(z)$ be a J-vector valued continuously differentiable function, and let

$C(z)$ be the $J \times k$ matrix $\partial c(z)/\partial z'$. Then

$$\sqrt{n}[c(z) - c(\mu)] \xrightarrow{d} N(0, C(\mu)\Sigma C(\mu)')$$

Then, if we combine **Theorem 1.4.2** and **Theorem 1.4.1**, we can have that:

Theorem 1.4.3 (Asymptotic Distribution of a Function of b)

For a $k \times 1$ vector $\boldsymbol{b}$, let $f(\boldsymbol{b})$ be a J-vetor valued continuously differentiable function, and let $C(\boldsymbol{b}) = \partial f(\boldsymbol{b})/\partial \boldsymbol{b}'$. Then;

$$\sqrt{n}[f(\boldsymbol{b}) - f(\boldsymbol{\beta})] \xrightarrow{d} N(0, C(\boldsymbol{\beta})\sigma^2 \boldsymbol{Q}^{-1} C(\boldsymbol{\beta})^{-1})$$

and we can estimate this variance consistently by:

$$C(\boldsymbol{b})s^2(\boldsymbol{X}'\boldsymbol{X}/n)^{-1}C(\boldsymbol{b})'$$

4.5   Interval Estimation

From **Theorem 1.4.1**, we know that:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \sigma^2 \boldsymbol{Q}^{-1})$$

which means that for the jth parameter $\boldsymbol{b}_j$, we have that:

$$\sqrt{n}(b_j - \beta_j) \xrightarrow{d} N(0, \sigma^2 \boldsymbol{Q}_j^{-1})$$

From **1.4.1**, we know that:

$$s^2(\boldsymbol{X}'\boldsymbol{X}/n)^{-1} \xrightarrow{p} \sigma^2 \boldsymbol{Q}^{-1}$$

which means that:

$$ns^2(\boldsymbol{X}'\boldsymbol{X})_j^{-1} \xrightarrow{p} \sigma^2 \boldsymbol{Q}_j^{-1}$$

Thus, because of the **Rules for Limiting Distributions**, we can have that:

$$\frac{\sqrt{n}(b_j - \beta_j)}{\sqrt{ns^2(\boldsymbol{X}'\boldsymbol{X})_j^{-1}}} \xrightarrow{d} N(0, 1)$$

4.6 Prediction and Forecasting

4.7 Data Problem

5 Hypothesis Test

5.1 Test

5.1.1 The F Statistic

For the model:

$$Y = X\beta + \varepsilon$$

Our hypothesis is that:

$$H_0 : \ R\beta - q = 0$$
$$H_1 : \ R\beta - q \neq 0$$

where R is a $J \times K$ matrix, q is a $J \times 1$ matrix.

Now, we can regress Y on X with and without the hypothesis:

$$Y = X\beta + \varepsilon$$
$$s.t.$$
$$R\beta = q$$

And we can partition R into $J \times (K - J)$ and $J \times J$ block, and β into $(K - J) \times 1$ and $J \times 1$ block:

$$R = \begin{pmatrix} R_1 & R_2 \end{pmatrix}$$
$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}$$

Then we can define:

$$\gamma = R_1\beta_1 + R_2\beta_2 - q$$

Then, we can solve backwards to get the β_2:

$$\beta_2 = R_2^{-1}(\gamma + q - R_1\beta_1)$$

Then, we take $\boldsymbol{\beta_2}$ into the model:

$$\begin{aligned}
Y &= X_1\beta_1 + X_2R_2^{-1}(\gamma + q - R_1\beta_1) + \varepsilon \\
&= X_1\beta_1 + X_2R_2^{-1}\gamma + X_2R_2^{-1}q - X_2R_2^{-1}R_1\beta_1 + \varepsilon
\end{aligned}$$

Then, we define:

$$\begin{aligned}
\tilde{Y} &= Y - X_2R_2^{-1}q \\
Z_1 &= X_1 - X_2R_2^{-1}R_1 \\
Z_2 &= X_2R_2^{-1}
\end{aligned}$$

The model can be rewrite as:

$$\tilde{Y} = Z_1\beta_1 + Z_2\gamma + \varepsilon$$

For the unrestricted residuals, $\gamma \neq 0$, the unrestricted residuals is:

$$e = M\tilde{Y}$$

where:

$$M = I - P = I - Z(Z'Z)^{-1}Z'$$

Then for the restricted residuals, $\gamma = 0$, the restricted model is:

$$\tilde{Y} = Z_1\beta_1 + \varepsilon$$

and the restricted residuals is:

$$e_\star = M_\star\tilde{Y} = M_\star\varepsilon$$

where:

$$M_\star = I - P_\star = I - Z_1(Z_1'Z_1)^{-1}Z_1'$$

Now we consider the statistic:

$$\frac{(e_\star'e_\star - e'e)/J}{e'e/(n-K)} = \frac{(\varepsilon/\sigma)'(M_\star - M)(\varepsilon/\sigma)/J}{(\varepsilon/\sigma)'M(\varepsilon/\sigma)/(n-K)}$$

and we can prove that:

$$\frac{(e'_\star e_\star - e'e)/J}{e'e/(n-K)} \sim F[J, n-K]$$

5.2 Large-sample Tests and Robust Inference

For the model

$$Y = X\beta + \varepsilon$$

With assumption **A.1, A.3, A.4, A.7**, we know that the OLS estimator is asymptotically normal.

5.2.1 Asymptotic t-test

Consider the hypothesis:

$$H_0 : \beta_j = \beta_j^0$$

Now we consider a statistic:

$$\frac{b_j - \beta_j^0}{se_j}$$

where $se_j = \sqrt{s^2(X'X)_j^{-1}}$
From **4.5** we know that:

$$t_j = \frac{b_j - \beta_j^0}{se_j} \xrightarrow{d} Z \sim N(0, 1)$$

We set the reject size α, and the rejection rule:

$$\text{Reject } H_0 : \ \beta_j = \beta_j^0, \text{ if } |t_j| \geq z_{1-\alpha/2}$$

Then, because the convergence in distribution:

$$\lim_{n \to \infty} |F_t(t_j \leq z_{1-\alpha/2}|H_0 \text{ is correct}) - F_z(Z \leq z_{1-\alpha/2})| = 0$$

$$\lim_{n \to \infty} |F_t(t_j \leq -z_{1-\alpha/2}|H_0 \text{ is correct}) - F_z(Z \leq -z_{1-\alpha/2})| = 0$$

which means that:

$$\lim_{n \to \infty} |F_t(|t_j| \geq z_{1-\alpha/2}|H_0 \text{ is correct}) - F_z(|Z| \geq z_{1-\alpha/2})|$$

$$= \lim_{n \to \infty} |F_t(|t_j| \geq z_{1-\alpha/2}|H_0 \text{ is correct}) - \alpha|$$

$$= 0$$

which means that:

$$Pr(\text{reject } H_0|H_0 \text{ is correct}) \to \alpha$$

This test is called the (asymptotic) t-test

5.2.2   Wald test

Consider the hypothesis:

$$H_0 : \boldsymbol{R\beta} = \boldsymbol{q}$$

If H_0 is correct, from **4.5** we know that:

$$\sqrt{n}(\boldsymbol{Rb} - \boldsymbol{q}) = \sqrt{n}\boldsymbol{R}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{R}\sigma^2 \boldsymbol{Q}^{-1} \boldsymbol{R}')$$

And we know that:

$$s^2 \left(\frac{\boldsymbol{X'X}}{n} \right)^{-1} \xrightarrow{p} \sigma^2 \boldsymbol{Q}^{-1}$$

which means that that:

$$\frac{\sqrt{n}(\boldsymbol{Rb} - \boldsymbol{q})}{\sqrt{\boldsymbol{R}s^2 \left(\frac{\boldsymbol{X'X}}{n} \right)^{-1} \boldsymbol{R}}} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{1})$$

Thus, we can construct the Wald statistic:

$$W = \sqrt{n}(\boldsymbol{Rb} - \boldsymbol{q})' \left[\boldsymbol{R}s^2 \left(\frac{\boldsymbol{X'X}}{n} \right)^{-1} \boldsymbol{R} \right]^{-1} \sqrt{n}(\boldsymbol{Rb} - \boldsymbol{q})$$

$$= (\boldsymbol{Rb} - \boldsymbol{q})' \left[\boldsymbol{R}s^2 (\boldsymbol{X'X})^{-1} \boldsymbol{R} \right]^{-1} (\boldsymbol{Rb} - \boldsymbol{q}) \xrightarrow{d} \chi^2(J)$$

5.3 Testing Nonlinear Restrictions

5.3.1 Construction of Wald Test

Now we consider the nonlinear hypothesis:

$$H_0 : c(\beta) = 0$$

where $c(\beta)$ is a set of J functions of β.

Then let :

$$G(\beta) = \frac{\partial c(\beta)}{\partial \beta'}$$

which is a $J \times K$ Jacobian matrix.

From **A.1, A.3, A.4, A.7**, we know that:

$$\sqrt{n}(b - \beta) \xrightarrow{d} N(0, \sigma^2 Q^{-1})$$

If H_0 is correct, then by **Theorem 4.4.2 (Delta-method)**, we know that:

$$\sqrt{n}(c(b) - c(\beta)) \xrightarrow{d} N(0, G(\beta)\sigma^2 Q^{-1}G(\beta)')$$

Then, similar to **5.2.2**, we can construct the Wald statistic:

$$W = \sqrt{n}c(b)' \left[G(b)s^2 \left(\frac{X'X}{n} \right)^{-1} G(b)' \right]^{-1} \sqrt{n}c(b)$$

$$= c(b)' \left[G(b)s^2 (X'X)^{-1} G(b)' \right]^{-1} c(b) \xrightarrow{d} \chi^2(J)$$

5.3.2 Drawbacks of Nonlinear Wald Test

5.3.2.1 Over-Rejection

If we use the $100(1 - \alpha)$-percentile of $\chi^2(J)$ distribution, from the asymptotic property we know that as $n \to \infty$:

$$Pr(\text{reject } H_0 | H_0 \text{ is correct}) \to \alpha$$

However, if n is finite, the rejection probability may exceed α. For example, for $n = 200$, $\alpha = 0.05$, the real probability will exceed 0.15.

5.3.2.2 Multiple Restriction

For a nonlinear hypothesis, we may have multiple way to construct the continuous function, which lead to multiple Jacobian matrix. For example:

$$H_0 : \beta_1 \beta_2 - 1 = 0$$

The restriction can be rewritten as:

$$H_{01} : \beta_1 - \frac{1}{\beta_2} = 0$$

$$H_{01} : \beta_2 - \frac{1}{\beta_1} = 0$$

which lead to different Jacobian matrix:

$$G_1(b) = \begin{pmatrix} 1 & \frac{1}{\beta_2^2} \end{pmatrix}$$

$$G_2(b) = \begin{pmatrix} \frac{1}{\beta_1^2} & 1 \end{pmatrix}$$

5.4 Choosing Between Nonnested Model

All the hypotheses above are just special cases of the origin model. If our hypothesis is to compare two different models:

$$H_0 : Y = X\beta + \varepsilon_0$$

$$H_1 : Y = Z\gamma + \varepsilon_1$$

5.4.1 Encompassing Model

One way is to construct another model:

$$Y = \bar{X}\bar{\beta} + \bar{Z} + \bar{\gamma} + W\delta + \varepsilon$$

where $\bar{X}$ be the set of variables in X that are not in Z, $\bar{Z}$ be the set of variables in Z that are not in X, W are the variables that the two models have in common.

Then H_1 could be reject if we use test to verify $\bar{\gamma} = 0$, and H_0 could be reject if we use test to verify $\bar{\beta} = 0$

5.4.2 Drawbacks of Encompassing Model

5.4.2.1 Model Specification

The first is that even we test that $\bar{\gamma}$ or $\bar{\beta}$ is 0, δ is a hybrid term of these two terms, which means we still can not exclude β or γ.

What we really test is between H_0 or H_1 with the encompassing model.

5.4.2.2 Variable Numbers

If we use the encompassing model, there are two many variables, which may lead to collinearity problem.

5.4.3 Comprehensive Approach: the J test

We construct a new model:

$$Y = (1 - \lambda)X\beta + \lambda Z\gamma + \varepsilon$$

Then our test is to test whether λ equals 0 or 1, and the steps are as follow:

First we regress Y on Z to get $\hat{\gamma}$.

Then we regress Y on X and $Z\hat{\gamma}$ to get $\hat{\beta}$ and $\hat{\lambda}$.

Then if H_0 is correct, we have:

$$\hat{\lambda} \xrightarrow{d} N(0, 1)$$

5.5 Model Building

6 Dummy Variables

6.1 Dummy Variables

A dummy variable only takes values of 1 or 0. It can be used to represent the state variables.

For n state, we use $n - 1$ dummy variables.

6.1.1 Difference in Differences Regression

We want to verify the influence of some policy. If we use the simple regression model with dummy variables:

$$Y = X\beta + D\delta + \varepsilon$$

where D is the dummy variable, and we use δ to measure the influence of the policy.

One problem is that D may correlate with the error term ε. Thus, we add more dummy variables to use DiD. The new model is:

$$Y = X\beta_1 + T\beta_2 + D\beta_3 + (TD)\delta + \varepsilon$$

T is the time dummy variable. $T = 0$ if before the treatment and $T = 1$ if after the treatment.

D is the treatment dummy variable. $D = 0$ if not in the treatment group $D = 1$ if in the treatment group.

DT equals 1 only for treated individuals after the treatment.

Thus, we first compute the treated group:

$$\mathbb{E}[y_{i1}|D_i = 1] - \mathbb{E}[y_{i0}|D_i = 1] = x_i{}'\beta_1 + \beta_2 + \beta_3 + \delta - x_i{}'\beta_1 - \beta_3 = \beta_2 + \delta$$

Then, we compute the control group:

$$\mathbb{E}[y_{i1}|D_i = 0] - \mathbb{E}[y_{i0}|D_i = 0] = x_i{}'\beta_1 + \beta_2 - x_i{}'\beta_1 = \beta_2$$

Thus, we can get the influence of the treatment by:

$$\delta = \mathbb{E}[\Delta y_{i1}|D_i = 1] - \mathbb{E}[\Delta y_{i1}|D_i = 0]$$

One important assumption is the exogeneity of the treatment, which mean that:

$$\mathbb{E}[\varepsilon_{it}|D_i] = 0$$

6.2 Structural Break and Parameter Variation

6.2.1 Chow Test

If we want to test whether the regression model are the same for two data set, we can use Chow test. Here we give the model:

$$Y_1 = X_1\beta_1 + \varepsilon_1$$
$$Y_2 = X_2\beta_2 + \varepsilon_2$$

The hypotheses are:

$$H_0 : \beta_1 = \beta_2$$
$$H_1 : \beta_1 \neq \beta_2$$

Then, for the unrestricted model:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

The unrestricted SSR is the sum of the two SSR:

$$e'e = e_1'e_1 + e_2'e_2$$

For the restricted model:

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}$$

The restricted SSR is the SSR of the regression:

$$e_\star = MY$$

Then, we the assumption $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$, we can construct the F test:

$$F = \frac{[\boldsymbol{e}'_\star \boldsymbol{e}_\star - (\boldsymbol{e_1}'\boldsymbol{e_1} + \boldsymbol{e_2}'\boldsymbol{e_2})]/K}{(\boldsymbol{e_1}'\boldsymbol{e_1} + \boldsymbol{e_2}'\boldsymbol{e_2})/(n_1 + n_2 - 2k)}$$

6.2.2 Wald Test for Structural Break

Chow test depends on the assumption that $\boldsymbol{\varepsilon} \sim N(0, \sigma^2 \boldsymbol{I})$. If the sample size is large, by asymptotic consistency, we can use the Wald test. For the unrestricted model:

$$\begin{bmatrix} \boldsymbol{Y_1} \\ \boldsymbol{Y_2} \end{bmatrix} = \begin{bmatrix} \boldsymbol{X_1} & 0 \\ 0 & \boldsymbol{X_2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta_1} \\ \boldsymbol{\beta_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \end{bmatrix}$$

We assume that $(\boldsymbol{X_1}, \boldsymbol{\varepsilon_1})$ and $(\boldsymbol{X_2}, \boldsymbol{\varepsilon_2})$ are independent. Under the assumptions **A.1, A.3, A.4, A.7**, we have that:

$$\sqrt{n} \begin{pmatrix} \boldsymbol{b_1} - \boldsymbol{\beta_1} \\ \boldsymbol{b_2} - \boldsymbol{\beta_2} \end{pmatrix} \xrightarrow{d} N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_1} & 0 \\ 0 & \boldsymbol{\Sigma_2} \end{pmatrix} \right)$$

where $\Sigma_1 = \sigma^2 \boldsymbol{Q_1}^{-1}$ and $\Sigma_2 = \sigma^2 \boldsymbol{Q_2}^{-1}$

Then, if H_0 is correct, we know that:

$$\sqrt{n}(\boldsymbol{b_1} - \boldsymbol{b_2}) \xrightarrow{d} N(0, \boldsymbol{\Sigma_1} + \boldsymbol{\Sigma_2})$$

Thus, we can construct the Wald test:

$$W = n(\boldsymbol{b_1} - \boldsymbol{b_2})'(\hat{\boldsymbol{\Sigma}}_1 + \hat{\boldsymbol{\Sigma}}_2)^{-1}(\boldsymbol{b_1} - \boldsymbol{b_2}) \xrightarrow{d} \chi^2(K)$$

6.3 Bias-variance Trade-off

In regression, we often don't know what the true independent variables set is. We will analysis the consequences of (i) omitting relevant variables and (ii) including superfluous variables.

Suppose the true model is:

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

where:

$$\mathbb{E}[\varepsilon|X] = \mathbf{0}$$
$$Var[\boldsymbol{\varepsilon}|X] = \sigma^2 I$$
$$X = [X_1 \ X_2]$$

If we omit X_2 and only regress Y on X_1, then by OLS, $\boldsymbol{b_1}$ is:

$$\boldsymbol{b_1} = (X_1'X_1)^{-1}X_1'Y$$
$$= \boldsymbol{\beta_1} + (X_1'X_1)^{-1}X_1'X_2\boldsymbol{\beta_2} + (X_1'X_1)^{-1}X_1'\boldsymbol{\varepsilon}$$

Then, the expectation of $\boldsymbol{b_1}$ is:

$$\mathbb{E}[\boldsymbol{b_1}|X] = \boldsymbol{\beta_1} + (X_1'X_1)^{-1}X_1'X_2\boldsymbol{\beta_2}$$

$\boldsymbol{b_1}$ is unbiased unless $X_1'X_2 = 0$ (they are orthogonal) or $\boldsymbol{\beta_2} = 0$ (X_2 is superfluous).

Now we consider the variance of $\boldsymbol{b_1}$:

$$Var[\boldsymbol{b_1}|X] = \mathbb{E}[(\boldsymbol{b_1} - \mathbb{E}[\boldsymbol{b_1}|X])(\boldsymbol{b_1} - \mathbb{E}[\boldsymbol{b_1}|X])'|X]$$
$$= \mathbb{E}[(X_1'X_1)^{-1}X_1'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'X_1(X_1'X_1)^{-1}|X]$$
$$= (X_1'X_1)^{-1}X_1'\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X]X_1(X_1'X_1)^{-1}$$
$$= (X_1'X_1)^{-1}X_1'\sigma^2 X_1(X_1'X_1)^{-1}$$
$$= \sigma^2(X_1'X_1)^{-1}$$

If we regress Y on X, by FWL theorem, $\boldsymbol{b_{1,2}}$ is:

$$\boldsymbol{b_{1,2}} = (X_1'M_2X_1)^{-1}X_1'M_2Y$$
$$= \boldsymbol{\beta} + (X_1'M_2X_1)^{-1}X_1'M_2\boldsymbol{\varepsilon}$$

Then the variance is:

$$Var[\boldsymbol{b_{1,2}}|X] = \sigma^2(X_1'M_2X_1)^{-1}$$

Note that:

$$(Var[\boldsymbol{b_1}|X])^{-1} - (Var[\boldsymbol{b_{1,2}}|X])^{-1} = \sigma^2(X_1'X - X_1'M_2X_1) = X_1'P_2X_1 = (P_2X_1)'P_2X_1$$

which means that $(Var[\boldsymbol{b_1}|X])^{-1} - (Var[\boldsymbol{b_{1,2}}|X])^{-1}$ is positive semidefinite. Thus, $Var[\boldsymbol{b_{1,2}}|X] - Var[\boldsymbol{b_1}|X]$ is positive semidefinite.

We can see that the true model regression have larger variance.

In conclusion:

When $\boldsymbol{\beta_2} = 0$, $\boldsymbol{b_1}$ is more efficient than $\boldsymbol{b_{1,2}}$.

When $\boldsymbol{\beta_2} \neq 0$, $\boldsymbol{b_{1,2}}$ is unbiased and $\boldsymbol{b_1}$ is biased.

$\boldsymbol{b_{1,2}}$ always has larger variance than $\boldsymbol{b_1}$.

However, as $n \rightarrow \infty$, both variance converge to 0, but $\boldsymbol{b_1}$ is still biased. Thus in large samples, we prefer $\boldsymbol{b_{1,2}}$

7 Instrument Variables

The consistency of the OLS requires that:

$$\mathbb{E}[\varepsilon_i | \boldsymbol{x_i}] = 0$$

This assumption cannot always be satisfied, which means endogenous.

7.1 Assumptions of the Extended Model

To solve the problem, we slightly change the previous assumptions in OLG:

A.1 (Linearity)

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$$

A.2 (Full Rank)

$$rank(\boldsymbol{X}) = K$$

A.I3 (Endogeneity)

The regressors provide information about the expectation of the error term:

$$\mathbb{E}[\varepsilon_i | \boldsymbol{x_i}] = \eta(\boldsymbol{x_i}) \neq 0$$

We add additional vector of $L \times 1$ variables $\boldsymbol{z_i}$, which satisfies:

    Exogeneity: $\boldsymbol{z_i}$ is uncorrelated with ε_i

    Relevance: $\boldsymbol{z_i}$ is correlated with $\boldsymbol{x_i}$

About $\boldsymbol{z_i}$, we add some assumptions:

A.I4 (Homoskedasticity)

$$\mathbb{E}[\varepsilon_i^2 | \boldsymbol{z_i}] = \sigma^2$$

A.I5 (Random sampling)

$$(\boldsymbol{x_i}, \boldsymbol{z_i}, \varepsilon_i) \sim i.i.d.$$

A.I6 (Moments of x_i and z_i)

$$\mathbb{E}[x_i x_i'] = Q_{xx}, \, rank(Q_{xx}) = K$$
$$\mathbb{E}[z_i z_i'] = Q_{zz}, \, rank(Q_{zz}) = L$$
$$\mathbb{E}[z_i x_i'] = Q_{zx}, \, rank(Q_{zx}) = K$$

A.I7 (Exogeneity of instruments)

$$\mathbb{E}[\varepsilon_i | z_i] = 0$$

7.2   Instrumental Variable Estimation

7.2.1   The Biasedness and Inconsistency of OLS

Suppose we have that:

$$\eta(X) = \begin{pmatrix} \eta(x_1) & \eta(x_2) & \dots & \eta(x_n) \end{pmatrix}'$$

Then, we have that:

$$\begin{aligned} \mathbb{E}[b|X] &= \mathbb{E}[(X'X)^{-1} X'(X\beta + \varepsilon)|X] \\ &= \beta + (X'X)^{-1} X' \eta(X) \\ &\neq \beta \end{aligned}$$

Thus, the OLG is biased under endogeneity.

With **A.I3**, we have that;

$$\begin{aligned} \mathbb{E}[x_i \varepsilon_i] &= \mathbb{E}[x_i \mathbb{E}[\varepsilon_i | x_i]] \\ &= \mathbb{E}[x_i \eta(x_i)] \\ &\neq 0 \end{aligned}$$

Then, we have that:

$$
\begin{aligned}
\boldsymbol{b} =& \boldsymbol{\beta} + (X'X)^{-1}X' \\
&\xrightarrow{p} \boldsymbol{\beta} + \plim_{n\to\infty}\left(\frac{X'X}{n}\right)^{-1}\plim_{n\to\infty}\left(\frac{X'\boldsymbol{\varepsilon}}{n}\right) \\
=& \boldsymbol{\beta} + \boldsymbol{Q_{xx}}^{-1}\mathbb{E}[\boldsymbol{x_i}\varepsilon_i] \\
\neq& \boldsymbol{\beta}
\end{aligned}
$$

Thus, the OLG is inconsistent under endogeneity.

7.2.2 Method of Moments

Besides OLG, we introduce a new estimating method. First, we go back to the ordinary assumption with exogeneity. Thus, we have that:

$$
\mathbb{E}[\boldsymbol{x_i}\varepsilon_i] = \mathbb{E}[\boldsymbol{x_i}(y_i - \boldsymbol{x_i}'\boldsymbol{\beta})] = \boldsymbol{0}
$$

Then, we can view this as the condition that $(y_i, \boldsymbol{x_i}, \boldsymbol{\beta})$ must satisfies.
Then, for the sample estimator, we have that:

$$
\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x_i}(y_i - \boldsymbol{x_i}'\boldsymbol{b}_{MOM}) = \boldsymbol{0}
$$

We get:

$$
\boldsymbol{b}_{MOM} = \left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x_i}\boldsymbol{x_i}'\right)\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{x_i}y_i\right) = (X'X)^{-1}X'Y = \boldsymbol{b}_{OLS}
$$

Now, we go back to the endogenous model, we have that;

$$
\mathbb{E}[\boldsymbol{z_i}\varepsilon_i] = 0
$$

Then, the MOM sample estimator satisfies:

$$
\frac{1}{n}\sum_{i=1}^{n}[\boldsymbol{z_i}(y_i - \boldsymbol{x_i}'\boldsymbol{b}_{MOM})] = \boldsymbol{0}
$$

If we have that $L = K$, we can solve it back to get:

$$\boldsymbol{b}_{MOM} = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i{}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i y_i = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$$

8.2.2.1 Consistency and Biasedness

We can verify that $\boldsymbol{b}_{MOM}$ is consistent but maybe not unbiased.

First, we verify the unbiasedness:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{b}_{MOM} - \boldsymbol{\beta}|\boldsymbol{X}, \boldsymbol{Z}] &= \mathbb{E}[(\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}] \\
&= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}, \boldsymbol{Z}] \\
&\neq 0
\end{aligned}$$

which means that it is not unbiased.

Second, we verify the consistency:

$$\begin{aligned}
\boldsymbol{b}_{MOM} &= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y} \\
&= (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} z_i \varepsilon_i \\
&\xrightarrow{p} \boldsymbol{\beta} + (\mathbb{E}[z_i x_i{}'])^{-1}\mathbb{E}[z_i \varepsilon_i] \\
&= \boldsymbol{\beta} + \boldsymbol{Q}_{zx}^{-1} \times 0 \\
&= \boldsymbol{\beta}
\end{aligned}$$

which means it is consistent.

8.2.2.2 Asymptotical Normality

For the homoscedastic case, we have that:

$$\sqrt{n}(\boldsymbol{b}_{MOM} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q}_{zx}^{-1}\boldsymbol{Q}_{zz}\boldsymbol{Q}_{xz}{}')$$

Here we give the proof:

Proof.

$$\sqrt{n}(\boldsymbol{b}_{MOM} - \boldsymbol{\beta}) = \left(\frac{1}{n}\sum_{i=1}^{n} z_i x_i{}'\right)^{-1} \frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i \varepsilon_i$$

By the **i.i.d. law of large number**, we have that:

$$\frac{1}{n}\sum_{i=1}^{n} z_i x_i{}' \xrightarrow{p} \boldsymbol{Q}_{zx}{}^{-1}$$

By the **i.i.d. law of large number**, we also have that:

$$\mathbb{E}[z_i \varepsilon_i] = 0$$
$$Var[z_i \varepsilon] = \mathbb{E}[z_i z_i{}' \varepsilon_i^2]$$
$$= \sigma^2 \mathbb{E}[z_i z_i{}']$$
$$= \sigma^2 \boldsymbol{Q}_{zz}$$

Thus, by **Central limit theorem**, we have that:

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} z_i \varepsilon_i \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q}_{zz})$$

And by the **rules of limiting distribution**, we have that:

$$\sqrt{n}(\boldsymbol{b}_{MOM} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q}_{zx}{}^{-1} \boldsymbol{Q}_{zz} \boldsymbol{Q}_{xz}{}^{-1})$$

$\square$

For the heteroskedastic case:
If we have:

$$\mathbb{E}[\varepsilon_i^2 | z_i] = \omega(z_i)$$

Then, we have:

$$Var[z_i \varepsilon_i] = \mathbb{E}[z_i z_i{}' \varepsilon_i^2] = \mathbb{E}[z_i z_i{}' \mathbb{E}[\varepsilon_i^2 | z_i]] = \mathbb{E}[z_i z_i{}' \omega(z_i)]$$

We set:

$$\boldsymbol{Q_{zz}}^\omega = \mathbb{E}[\boldsymbol{z_i z_i}' \omega(\boldsymbol{z_i})]$$

Then, we have:

$$\sqrt{n}(\boldsymbol{b}_{MOM} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q_{zx}}^{-1} \boldsymbol{Q_{zz}}^\omega \boldsymbol{Q_{xz}}^{-1})$$

7.3   Two-Stage Least Squares

In some cases, the number of the instruments, L, is larger than the number of independent variables, K. To fully use the instruments, we decide to use the projection of independent variable, X, onto the instrumental variables, Z:

$$\hat{X} = Z(Z'Z)^{-1} Z'X = P_z X$$

and because the instruments are exogenous. $\hat{X}$ is on the instruments space, thus $\hat{X}$ is exogenous.

Then, the two-stage least squares estimator is:

$$\begin{aligned}
\boldsymbol{b}_{2SLS} &= (\hat{X}'\hat{X})^{-1} \hat{X} Y \\
&= (X' P_z X)^{-1} X' P_z Y
\end{aligned}$$

7.3.1   Consistency and Biasedness

8.3.1.1 Consistency of Two-Stage Least Squares

First, we can prove that the 2-stage least squares estimator is consistent:

$$\begin{aligned}
\boldsymbol{b}_{2SLS} &= (X' P_z X)^{-1} X' P_z Y \\
&= (X' P_z X)^{-1} X' P_z (X\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta} + (X' P_z X)^{-1} X' P_z \boldsymbol{\varepsilon}
\end{aligned}$$

From the **i.i.d. law of large number**, we have that:

$$\left(\frac{X' P_z X}{n}\right)^{-1} = \left(\frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'X}{n}\right)^{-1} \xrightarrow{p} (\boldsymbol{Q_{xz}}\boldsymbol{Q_{zz}}^{-1}\boldsymbol{Q_{zx}})^{-1}$$

$$\frac{X' P_z \boldsymbol{\varepsilon}}{n} = \frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'\boldsymbol{\varepsilon}}{n} \xrightarrow{p} \boldsymbol{Q_{xz}}\boldsymbol{Q_{zz}}^{-1}\mathbb{E}[\boldsymbol{z_i}\varepsilon_i] = 0$$

51

Thus, we have that:

$$b_{2SLS} \xrightarrow{p} \beta$$

8.3.1.2 Biasedness of Two-Stage Least Squares

From above we know that:

$$b_{2SLS} = \beta + (X'P_z X)^{-1} X'P_z \varepsilon$$

We can not guarantee that:

$$\mathbb{E}[\varepsilon|XZ] = 0$$

Thus, the estimator may not be unbiased.

7.3.2 Asymptotic Normality of Two-Stage Least Squares

For the homoscedastic case, we have that:

$$\sqrt{n}(b_{2SLS} - \beta) \xrightarrow{d} N(0, \sigma^2(Q_{zx}Q_{zz}^{-1}Q_{xz})^{-1})$$

Proof.
We have that:

$$b_{2SLS} - \beta = (X'P_z X)^{-1} X'P_z \varepsilon$$

We know that:

$$\left(\frac{X'P_z X}{n}\right)^{-1} \xrightarrow{p} (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1}$$

$$\frac{X'P_z \varepsilon}{n} = \frac{X'Z}{n}\left(\frac{Z'Z}{n}\right)^{-1}\frac{Z'\varepsilon}{n}$$

From **central limit theorem**, we know that:

$$\sqrt{n}\frac{Z'\varepsilon}{n} \xrightarrow{d} N(0, \sigma^2 Q_{zz})$$

Thus:

$$\frac{X'P_z \varepsilon}{\sqrt{n}} \xrightarrow{d} N(0, \sigma^2 Q_{xz}Q_{zz}^{-1}Q_{zx})$$

Thus, we have:

$$\sqrt{n}(\boldsymbol{b}_{2SLS} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2(\boldsymbol{Q}_{zx}\boldsymbol{Q}_{zz}^{-1}\boldsymbol{Q}_{xz})^{-1})$$

$\square$

For the heteroskedastic case:

Suppose we have that:

$$\mathbb{E}[\varepsilon_i^2|\boldsymbol{z_i}] = \omega(\boldsymbol{z_i})$$

We set:

$$Var[\boldsymbol{z_i}\varepsilon_i] = \mathbb{E}[\boldsymbol{z_i}\boldsymbol{z_i}'\omega(\boldsymbol{z_i})] = \boldsymbol{Q}_{zz}^{\omega}$$

We have that:

$$\sqrt{n}(\boldsymbol{b}_{2SLS} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{A}^{-1}\boldsymbol{B}(\boldsymbol{A}^{-1})')$$

where:

$$\boldsymbol{A} = \boldsymbol{Q}_{xz}\boldsymbol{Q}_{zz}^{-1}\boldsymbol{Q}_{zx}$$
$$\boldsymbol{B} = \boldsymbol{Q}_{xz}\boldsymbol{Q}_{zz}^{-1}\boldsymbol{Q}_{zz}^{\omega}\boldsymbol{Q}_{zz}^{-1}\boldsymbol{Q}_{zx}$$

7.3.3 Asymptotic Variance Estimation and t-statistic

From above we know that:

$$\sqrt{n}(\boldsymbol{b}_{2SLS} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2(\boldsymbol{Q}_{zx}\boldsymbol{Q}_{zz}^{-1}\boldsymbol{Q}_{xz})^{-1})$$

If we want to conduct the statistical inference, we need to estimate the variance. $\boldsymbol{Q}_{xz}$ and $\boldsymbol{Q}_{zz}$ can be easily estimate by $n^{-1}\boldsymbol{X}'\boldsymbol{Z}$ and $n^{-1}\boldsymbol{Z}'\boldsymbol{Z}$, as for σ^2, we can prove that $1/n \sum_{i=1}^n (y_i - \boldsymbol{x_i}'\boldsymbol{b}_{2SLS})^2$ is a consistent estimator. We give the proof:

Proof.

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'b_{2SLS})^2$$

$$=\frac{1}{n}\sum_{i=1}^{n}[\varepsilon_i + x_i'(\beta - b_{2SLS})]^2$$

$$=\frac{1}{n}(\sum_{i=1}^{n}\varepsilon_i^2 + 2\sum_{i=1}^{n}\varepsilon_i x_i'(\beta - b_{2SLS}) + \sum_{i=1}^{n}[x_i'(\beta - b_{2SLS})]^2)$$

The last tow parts are converging to 0. Thus,

$$\frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'b_{2SLS})^2 \xrightarrow{p} \mathbb{E}[\varepsilon_i^2] = \sigma^2$$

$\square$

And with the asymptotic estimation of variance, we can conduct the t test:

$$H_0 = \beta_j = \beta_j^0$$

The t-statistics:

$$t_{b_j} = \frac{b_{j,2SLS} - \beta_j^0}{\sqrt{\hat{\sigma}^2 S^{jj}}}$$

7.4   Hypothesis Tests for Endogeneity

7.4.1   The Efficiency of Instrumental Variable

Suppose that we have assumption **A.1, A.3, A.4, A.7, A.I4, A.I5, A.I6, A.I7**, then both OLS and IV estimators are consistent. To compare which one is more efficient, we consider the asymptotic variance. From above we know that:

$$\sqrt{n}(b_{OLS} - \beta) \xrightarrow{d} N(0, \sigma^2 Q_{xx}^{-1})$$

$$\sqrt{n}(b_{MOM} - \beta) \xrightarrow{d} N(0, \sigma^2 (Q_{xz}Q_{zz}^{-1}Q_{zx})^{-1})$$

Then, we have that:

$$(\sigma^2 \boldsymbol{Q_{xx}}^{-1})^{-1} - [\sigma^2 (\boldsymbol{Q_{xz}} \boldsymbol{Q_{zz}}^{-1} \boldsymbol{Q_{zx}})^{-1}]^{-1}$$
$$= \sigma^{-2} (\boldsymbol{Q_{xx}} - \boldsymbol{Q_{xz}} \boldsymbol{Q_{zz}}^{-1} \boldsymbol{Q_{zx}})$$
$$= \sigma^{-2} \plim_{n \to \infty} (\frac{1}{n} \boldsymbol{X}' \boldsymbol{X} - \frac{1}{n} \boldsymbol{X}' \boldsymbol{P_z} \boldsymbol{X})$$
$$= \sigma^{-2} \plim_{n \to \infty} (\frac{1}{n} \boldsymbol{X}' \boldsymbol{M_z} \boldsymbol{X})$$

We know that $\sigma^{-2} \plim_{n \to \infty} (\frac{1}{n} \boldsymbol{X}' \boldsymbol{M_z} \boldsymbol{X})$ is positive semidefinite, which means that $\sigma^2 (\boldsymbol{Q_{xz}} \boldsymbol{Q_{zz}}^{-1} \boldsymbol{Q_{zx}})^{-1} - \sigma^2 \boldsymbol{Q_{xx}}^{-1}$ is positive semidefinite. Thus, the OLG estimator is efficient, when $\boldsymbol{X}$ is exogenous.

This result is along with the **Gauss-Markov Theorem**.

7.4.2 Hausman Test

Now, we want to test whether there is endogeneity in $\boldsymbol{X}$. We use the Hausman test, which is:

$$H_0 : \ \mathbb{E}[\varepsilon_i | \boldsymbol{x_i}] = 0$$
$$H_1 : \ \mathbb{E}[\varepsilon_i | \boldsymbol{x_i}] \neq 0$$

We set the difference of the two estimator as:

$$\boldsymbol{d} = \boldsymbol{b}_{MOM} - \boldsymbol{b}_{OLS}$$

If H_0 holds, then by the property of normal distribution, we know that:

$$\sqrt{n} \boldsymbol{d} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V})$$

Here, we need to compute $\boldsymbol{V}$ so that we can construct the Wald test. To simplify $\boldsymbol{V}$, we give the Hauseman's principle:

Proposition 8.4.1 (Hauseman's principle)
Let $\boldsymbol{b_E}$ an estimator of $\boldsymbol{\beta}$ with $\sqrt{n}(\boldsymbol{b_E} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{V_E})$. $\boldsymbol{b_E}$ is efficient if:
For any other estimator $\boldsymbol{b}$ of $\boldsymbol{\beta}$ such that $\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{V})$, $V - V_E$ is positive semidefinite.

Let $\boldsymbol{b_I}$ be an inefficient estimator of $\boldsymbol{\beta}$, then:

$$Asy.Cov(\boldsymbol{b_E}, \boldsymbol{b_I}) = Asy.Var(\boldsymbol{b_E})$$
$$Asy.Cov(\boldsymbol{b_E}, \boldsymbol{b_I} - \boldsymbol{b_E}) = 0$$

Thus, under H_0, $\boldsymbol{b_{OLS}}$ is efficient and $\boldsymbol{b_{MOM}}$ is inefficient. Thus, we can calculate $\boldsymbol{V}$:

$$\begin{aligned}
\boldsymbol{V} &= Var(\boldsymbol{b_{MOM}} - \boldsymbol{b_{OLS}}) \\
&= Var(\boldsymbol{b_{MOM}}) + Var(\boldsymbol{b_{OLS}}) - 2Cov(\boldsymbol{b_{MOM}}, \boldsymbol{b_{OLS}}) \\
&= Var(\boldsymbol{b_{MOM}}) + Var(\boldsymbol{b_{OLS}}) - 2Var(\boldsymbol{b_{OLS}}) \\
&= Var(\boldsymbol{b_{MOM}}) - Var(\boldsymbol{b_{OLS}})
\end{aligned}$$

Thus, we can construct the Wald statistic:

$$n\boldsymbol{d}'(\hat{\boldsymbol{V}}_{MOM} - \hat{\boldsymbol{V}}_{OLS})\boldsymbol{d} \xrightarrow{d} \chi^2(r)$$

Here r is the number of independent variables in $\boldsymbol{x_i}$ that is considered to be endogenous.

7.5   Weak Instruments and LIML

The key conditions for consistency of instrument variance is **A.I6**:

$$\mathbb{E}[\boldsymbol{z_i}\boldsymbol{x_i}'] = \boldsymbol{Q_{zx}}, \ rank(Q_{zx}) = K$$

When the instruments are only slightly correlated with $\boldsymbol{x_i}$, it is called 'weak'. Under the weak instruments, we have:

$$rank(\boldsymbol{Q_{zx}}) < K$$

7.5.1 The Method of Moment under Weak Instrument

Under the MOM, we know the estimator is given as:

$$\boldsymbol{b}_{MOM} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z_i}\boldsymbol{x_i}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z_i}y_i$$

$$= \boldsymbol{\beta} + \left(\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z_i}\boldsymbol{x_i}'\right)^{-1}\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{z_i}\varepsilon_i$$

As instruments only slightly correlated with $\boldsymbol{x_i}$

7.6 Measurement Error

7.6.1 Measurement Error of Dependent Variable

Suppose the true model is:

$$y^{\star} = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon$$

The reported y has error:

$$y = y^{\star} + v$$

where v is independent of $(y^{\star}, x)$, $\mathbb{E}[v] = 0$ and $Var[v] = \sigma_v^2$.
Then we have that:

$$y = \boldsymbol{x}'\boldsymbol{\beta} + \varepsilon + v$$

We regress $\boldsymbol{Y}$ on $\boldsymbol{X}$, the OLS estimator is:

$$\boldsymbol{b} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

$$= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \varepsilon + \boldsymbol{v})$$

$$= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{v}$$

It is still unbiased and consistent.

7.6.2 Measurement Error of Independent Variable

Suppose the true model is:

$$y^\star = x^{\star\prime}\beta + \varepsilon$$

The reported x has error:

$$x = x^\star + u$$

where u is independent of $(y, x^\star, \varepsilon)$, $\mathbb{E}[u] = 0$ and $Var[u] = \sigma_u^2$.
Then we have that:

$$y = x'\beta + [\varepsilon - u'\beta] = x'\beta + w$$

Then, there is a endogeneity problem:

$$Cov[x, w] = Cov[x^\star + u, \varepsilon - u'\beta] = -\beta\sigma_u^2$$

Then, the OLS estimator is:

$$
\begin{aligned}
b &= (X'X)^{-1}X'Y \\
&= (X'X)^{-1}X'(X\beta + w) \\
&= \beta + (X'X)^{-1}X'w
\end{aligned}
$$

It is still unbiased.
However:

$$b \xrightarrow{p} \beta + Cov(x, w) = \beta \frac{\sigma_{x^\star}^2}{\sigma_{x^\star}^2 + \sigma_u^2}$$

which means it is not consistent.

8 Generalized Least Squares

In the previous model, assumption **A.4** gives us homoskedasticity and non-autocorrelation assumption, in this chapter, we relax this assumption as:

Assumption A.H4 (Heteroskedasticity)

$$Var[\varepsilon_i|X] = \sigma_i^2, \; i = 1, 2, \ldots, n$$
$$Cov[\varepsilon_i, \varepsilon_j|X] = 0, \; i \neq j$$

8.1 The Least Squares under Heteroskedasticity

We still consider the least squares estimator:

$$\boldsymbol{b} = (X'X)^{-1}X'Y$$

We can see that it is still unbiased and consistent.

First, we gives a another assumption:

Assumption A.7 (Random Sampling with Heteroskedasticity)

$(y_i, \boldsymbol{x}_i, \varepsilon_i)$ are i.i.d., $Var[\varepsilon_i|\boldsymbol{x}_i] = g(\boldsymbol{x_i})$, where $g(\cdot)$ is unknown. We assume that $\mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i'\varepsilon_i^2] = \boldsymbol{Q}_x^\varepsilon < \infty$, $\mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i'] = \boldsymbol{Q}_x < \infty$, and $\boldsymbol{Q}_x^{-1}$ exists.

8.1.1 Unbiasedness and Consistency

We first verify the unbiasedness:

$$\begin{aligned}
\mathbb{E}[\boldsymbol{b}|X] &= \mathbb{E}[\boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon}|X] \\
&= \boldsymbol{\beta} + \mathbb{E}[(X'X)^{-1}X'\boldsymbol{\varepsilon}|X] \\
&= \boldsymbol{\beta}
\end{aligned}$$

The OLS estimator is unbiased.

Now, we verify the consistency:

$$\begin{aligned}
\boldsymbol{b} &= \boldsymbol{\beta} + (X'X)^{-1}X'\boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta} + (\frac{X'X}{n})^{-1}\frac{X'\boldsymbol{\varepsilon}}{n} \xrightarrow{p} \boldsymbol{\beta} + \boldsymbol{Q}_x'\mathbb{E}[\boldsymbol{x}_i'\varepsilon_i] = \boldsymbol{\beta}
\end{aligned}$$

8.1.2 Asymptotic Normality

We know that:

$$\mathbb{E}[x_i'\varepsilon_i] = \mathbf{0}$$
$$Var[x_i'\varepsilon_i] = \mathbb{E}[\varepsilon_i x_i x_i'\varepsilon_i] = \mathbf{Q}_x^\varepsilon$$

Then for the OLS estimator:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) = (\frac{X'X}{n})^{-1}\sqrt{n}\frac{X'\varepsilon}{n}$$

And:

$$\sqrt{n}\frac{X'\varepsilon}{n} = \sqrt{n}\frac{1}{n}\sum_{i=1}^{n}x_i'\varepsilon_i \xrightarrow{d} N(0, \mathbf{Q}_x^\varepsilon)$$

Thus, we have that:

$$\sqrt{n}(\boldsymbol{b} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \mathbf{Q}_x^{-1}\mathbf{Q}_x^\varepsilon\mathbf{Q}_x^{-1})$$

8.2 Robust Least Squares Inference

From **9.1.2** we know that:

$$\boldsymbol{b} - \boldsymbol{\beta} \xrightarrow{d} N(0, \mathbf{Q}_x^{-1}\mathbf{Q}_x^\varepsilon\mathbf{Q}_x^{-1})$$

We don't know the variance. Thus, we have to estimate the variance consistently:

$$\frac{1}{n}\sum_{i=1}^{n}x_i x_i' e_i^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - x_i'\boldsymbol{b})'x_i x_i'(y_i - x_i'\boldsymbol{b})$$

If we want to test the hypothesis:

$$H_0 : \beta_j = \beta_j^0$$

We can construct the t-statistic:

$$t_j^R = \frac{b_j - \beta_j^0}{se_j^R} = \frac{\sqrt{n}(b_j - \beta_j^0)}{\sqrt{n}se_j^R} \xrightarrow{d} N(0, 1)$$

where:

$$se_j^R = \sqrt{\frac{\hat{V}_{jj}}{n}}$$

Here, we want to find a $\hat{V}$ to estimate the variance, $Q_x^{-1} Q_x^\varepsilon Q_x^{-1}$ consistently.
The Q_x^{-1} is estimated by $(X'X/n)^{-1}$.
As for Q_x^ε, we use $(1/n) \sum_{i=1}^n x_i x_i' e_i^2$ to estimate it consistently. Here we give the proof:

Proof.
We have that

$$y_i = x_i' b + e_i = x_i' \beta + \varepsilon_i$$

which means that:

$$e_i^2 = \varepsilon_i^2 + [x_i'(\beta - b)]^2 + 2\varepsilon_i x_i'(\beta - b) \xrightarrow{p} \varepsilon_i^2$$

Then, we have that:

$$\frac{1}{n} \sum_{i=1}^n x_i x_i' e_i^2 = \frac{1}{n} \sum_{i=1}^n x_i x_i' \varepsilon_i^2 + \frac{1}{n} \sum_{i=1}^n x_i x_i' (e_i^2 - \varepsilon_i^2)$$

$$= \frac{1}{n} \sum_{i=1}^n x_i x_i' \varepsilon_i^2 + \frac{1}{n} \sum_{i=1}^n x_i x_i' \{[x_i'(\beta - b)]^2 + 2\varepsilon_i x_i'(\beta - b)\}$$

$$\xrightarrow{p} Q_x^\varepsilon$$

$\square$

8.3 Instrumental Variance Estimation

Similar to **Chapter 8**, here we assume **A.1 (Linearity)**, **A.I5 (Random sampling)**, **A.I6 (Moments of x_i and z_i)**, **A.I7 (Exogeneity of instruments)**. We also modify slightly the heteroskedasticity assumption:
A.I4H (Heteroskedasticity)

$$\mathbb{E}[\varepsilon_i^2 | z_i] = g(z_i)$$

where $g(\cdot)$ is unknown, and $Q_z^\varepsilon = \mathbb{E}[z_i z_i' \varepsilon_i^2]$ is finite.

Now for the 2SLS estimator:

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

We have some properties.

8.3.1 Consistency and Asymptotic Normality of 2SLS Estimator

For the estimator, we have that:

$$\sqrt{n}(b - \beta) = \left(\frac{X' P_z X}{n}\right)^{-1} \sqrt{n} \frac{X' P_z \varepsilon}{n}$$

$$= \left[\frac{X' Z}{n} \left(\frac{Z' Z}{n}\right)^{-1} \frac{Z' X}{n}\right]^{-1} \frac{X' Z}{n} \left(\frac{Z' Z}{n}\right)^{-1} \sqrt{n} \frac{Z' \varepsilon}{n}$$

We have that:

$$\mathbb{E}[z_i' \varepsilon_i] = 0$$
$$Var[z_i' \varepsilon_i] = \mathbb{E}[z_i z_i' \varepsilon_i^2] = Q_z^\varepsilon$$

Thus, we have that:

$$\sqrt{n} \frac{Z' \varepsilon}{n} \xrightarrow{d} N(0, Q_z^\varepsilon)$$

Then, we have that:

$$\sqrt{n}(b - \beta) \xrightarrow{d} N(0, A^{-1} B A^{-1})$$

where:

$$A = Q_{xz} Q_{zz}^{-1} Q_{zx}$$
$$B = Q_{xz} Q_{zz}^{-1} Q_z^\varepsilon Q_{zz}^{-1} Q_{xz}$$

Similarly, we can use $(1/n) \sum_{i=1}^n z_i' z_i e_i^2$ to estimate $\hat{Q}_z^\varepsilon$ consistently.

8.4 Efficient Estimator by Generalized Least Squares

For the heteroskedastic model, we assume that:

$$\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X] = \boldsymbol{\Sigma}$$

The normal OLS estimator is no BLUE. However, we can construct another estimator. We first define:

$$\boldsymbol{\Sigma}^{-1/2} = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma_2} & \cdots & 0 \\ & & \ddots & \\ 0 & 0 & \cdots & \frac{1}{\sigma_n} \end{bmatrix}$$

Premultiplying the model by $\boldsymbol{\Sigma}^{-1/2}$:

$$\begin{aligned} \boldsymbol{\Sigma}^{-1/2}Y &= \boldsymbol{\Sigma}^{-1/2}X\boldsymbol{\beta} + \boldsymbol{\Sigma}^{-1/2}\boldsymbol{\varepsilon} \\ Y_\star &= X_\star\boldsymbol{\beta} + \boldsymbol{\varepsilon}_\star \end{aligned}$$

Then, we can have that:

$$\begin{aligned} \mathbb{E}[\boldsymbol{\varepsilon}_\star|X_\star] &= \mathbf{0} \\ \mathbb{E}[\boldsymbol{\varepsilon}_\star\boldsymbol{\varepsilon}_\star'|X_\star] &= I \end{aligned}$$

It becomes a homoskedasticity model, then we can use OLS:

$$\boldsymbol{b}_{GLS} = (X_\star'X_\star)^{-1}X_\star'Y_\star$$

This is called the generalized least squares estimator. It satisfies the OLS assumptions, so this estimator is BLUE.

If $\boldsymbol{\Sigma}$ is diagonal, this is called the weighted least squares estimator.

8.4.1 Asymptotic Normality of GLS

From above, we know that:

$$\sqrt{n}(\boldsymbol{b}_{GLS} - \boldsymbol{\beta}) = \left(\frac{X'_\star X_\star}{n}\right)^{-1} \sqrt{n}\frac{X'_\star \boldsymbol{\varepsilon}_\star}{n}$$

$$\sqrt{n}\frac{X'_\star \boldsymbol{\varepsilon}_\star}{n} \xrightarrow{d} N(0, 1)$$

Then, we can have that:

$$\sqrt{n}(\boldsymbol{b}_{GLS} \xrightarrow{d} N(0, plim\left(\frac{X'_\star X_\star}{n}\right)^{-1})$$

8.4.2 Feasible Generalized Least Squares

If we want to compute the GLS estimator, we need to know Σ. If the Σ is parameterized as $\Sigma = \Sigma(\alpha)$ for an parameter α, and we have a consistent estimator $\hat{\alpha}$ of α.
The estimator $(X'\Sigma(\hat{\alpha})^{-1}X)^{-1}X'\Sigma(\hat{\alpha})^{-1}Y$ is called feasible GLS estimator if:

$$plim_{n\to\infty}\left(\frac{1}{n}X'\Sigma(\hat{\alpha})^{-1}X - \frac{1}{n}X'\Sigma^{-1}X\right) = \boldsymbol{0}$$

$$plim_{n\to\infty}\left(\frac{1}{n}X'\Sigma(\hat{\alpha})^{-1}\boldsymbol{\varepsilon} - \frac{1}{n}X'\Sigma^{-1}\boldsymbol{\varepsilon}\right) = \boldsymbol{0}$$

which means this estimator has the same asymptotic distribution.

9 Systems of Regression Equations

9.1 GLS Estimator

We consider a model with M equations, and each equations has T observations:

$$Y_1 = X_1\beta_1 + \varepsilon_1$$
$$Y_2 = X_2\beta_2 + \varepsilon_2$$
$$\vdots$$
$$Y_M = X_M\beta_M + \varepsilon_M$$

We assume that M is fixed and $T \to \infty$.

The first method is just regression each equation to get the parameters:

$$b_i = (X_i' X_i)^{-1} X_i' Y_i$$

The second method is to combine all the equations into one equation:

$$Y = X\beta + \varepsilon$$

where:

$$Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_M \end{bmatrix} \qquad X = \begin{bmatrix} X_1 & 0 & \dots & 0 \\ 0 & X_2 & \dots & \vdots \\ & & \ddots & \\ 0 & \dots & 0 & X_M \end{bmatrix}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \end{bmatrix} \qquad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_M \end{bmatrix}$$

Here, we assume that X is exogenous.

We assume the errors are serially uncorrelated (in different time they are uncorrelated):

$$\mathbb{E}[\varepsilon_{it}\varepsilon_{js}|X] = 0, \ \forall \ (i, j), \ \forall \ t \neq s$$

We assume the errors are correlated in the same time:

$$\mathbb{E}[\varepsilon_{it}\varepsilon_{jt}|X] = \sigma_{ij}$$

Then, we have that:

$$\mathbb{E}[\boldsymbol{\varepsilon_i}\boldsymbol{\varepsilon_j}'X] = \sigma_{ij}\boldsymbol{I}$$
$$\mathbb{E}[\boldsymbol{\varepsilon}|X] = 0$$
$$\mathbb{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'|X] = \boldsymbol{\Sigma} \otimes \boldsymbol{I} = \boldsymbol{\Omega}$$

Then, the GLS estimator is:

$$\boldsymbol{b}_{GLS} = (X'\boldsymbol{\Omega}^{-1}X)^{-1}X'\boldsymbol{\Omega}^{-1}Y$$

The GLS and OLS are identical under either of the following conditions:
1. $\sigma_{ij} = 0, \forall\, i \neq j$.
2. $\boldsymbol{X_i} = \boldsymbol{X_j}, \forall\, i, j$.

9.2 Feasible GLS

We need to estimate σ_{ij} so that we can estimate the GLS estimator.
Let $\boldsymbol{e_i}$ denotes the least squares residuals from the ith equations. Then the σ_{ij} can be estimated consistently by:

$$\hat{\sigma}_{ij} = \frac{\boldsymbol{e_i}'\boldsymbol{e_i}}{T}$$

Then, we can estimate the GLS estimator.

10 General Method of Moments

10.1 Identification of Model

The parameter β is **identified** if we can write β in terms of a function of population distribution, or we can pin down β uniquely when we have an infinite number of observations.

Back to the moment of method for instrument variable:

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

where x_i and β are $k \times 1$, and $(y_i, x_i, \varepsilon_i) \sim i.i.d.$

We assume that:

$$\mathbb{E}[z_i \varepsilon_i] = 0 \; \text{(A.I7)}$$
$$rank\left(\mathbb{E}[z_i x_i']\right) = k \; \text{(A.I6)}$$

Using the MOM:

$$\mathbb{E}[z_i(y_i - x_i'\beta)] = 0$$

which means that:

$$\mathbb{E}[z_i y_i] = \beta \mathbb{E}[z_i x_i']$$

If $k = L$, from **A.I6** we know that $\mathbb{E}[z_i x_i']$ is nonsingular, then the β is identified:

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

If $L < k$, we cannot identify β. It is not identified.
If $L > k$, β is overqualified.

10.2 GMM Estimator

We first define for a $k \times 1$ vector b:

$$m_i(b) = z_i(y_i - x_i'b)$$

Then, the MOM gives the population moment condition:

$$\mathbb{E}[m_i(\boldsymbol{\beta})] = \mathbf{0}$$

and the sample moment is:

$$\bar{m}_n(\boldsymbol{b}) = \frac{1}{n} \sum_{i=1}^{n} m_i(\boldsymbol{\beta})$$

by **i.i.d. law of large number**, we know that:

$$\bar{m}_n(\boldsymbol{b}) \xrightarrow{P} \mathbb{E}[m_i(\boldsymbol{b})]$$

Thus, the vector $\boldsymbol{b}$ that makes $\bar{m}_n(\boldsymbol{b})$ is a good estimator for $\boldsymbol{\beta}$.

Similar to **11.2**, if $L = k$, we can solve the problem $\bar{m}_n(\boldsymbol{b}) = \mathbf{0}$ and get:

$$\boldsymbol{b} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y}$$

However, if $L > k$, we can not solve the problem.

Now we construct a optimality problem to estimate $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{b}} Q_n(\boldsymbol{b}), \ Q_n(\boldsymbol{b}) = \bar{m}_n(\boldsymbol{b})'\hat{W}\bar{m}_n(\boldsymbol{b})$$

where $\hat{W}$ is an $L \times L$ positive definite weighting matrix.
We solve the problem by derivatives:

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

This is called GMM estimator.

10.3 The Property of GMM Estimator

10.3.1 Assumption for GMM

A.GMM1 (Random Sampling)

$$(y_i, \boldsymbol{x_i}, \boldsymbol{z_i}), \ i = 1, \ldots, n, \ \text{is i.i.d.}$$

A.GMM2 (Population Moment Condition)

$$\mathbb{E}[z_i(y_i - x_i'\beta)] = 0$$

A.GMM3 (Identification)

$$rank(\mathbb{E}[z_i x_i']) = k$$

A.GMM4 (Weighting Matrix)

$\hat{W} \xrightarrow{P} W$, which is nonrandom, symmetric, and positive definite.

10.3.2   Asymptotic Normality of GMM

From the GMM estimator, we have that:

$$\begin{aligned}
\hat{\beta} - \beta &= (X'Z\hat{W}Z'X)^{-1}X'Z\hat{W}Z'\varepsilon \\
&= \left(\frac{X'Z}{n}\hat{W}\frac{Z'X}{n}\right)^{-1}\frac{X'Z}{n}\hat{W}\frac{Z'\varepsilon}{n}
\end{aligned}$$

If we define:

$$\begin{aligned}
C &= \mathbb{E}[z_i x_i'] \\
\Lambda &= \mathbb{E}[z_i z_i' \varepsilon_i^2] = Var[z_i \varepsilon_i]
\end{aligned}$$

We have that:

$$\begin{aligned}
\frac{X'Z}{n} &\xrightarrow{P} C \\
\frac{Z'\varepsilon}{n} &\xrightarrow{d} N(0, \Lambda) \\
\hat{W} &\xrightarrow{P} W
\end{aligned}$$

Then the asymptotic distribution is given as:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, V)$$

where:

$$V = (C'WC)^{-1}C'W\Lambda WC(C'WC)^{-1}$$

10.3.3 Optimal Weighting Matrix

If we set $W = \Lambda^{-1}$, then we can simplify of asymptotic variance of β to:

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow{d} N(\mathbf{0}, (C'\Lambda^{-1}C)^{-1})$$

Then, for any other W, we can see that:

$$(C'WC)^{-1}C'W\Lambda WC(C'WC)^{-1} - (C'\Lambda^{-1}C)^{-1}$$

$\star$ We can prove that this is positive semi-definite. Thus, we set $W = \Lambda^{-1}$ is the most efficient GMM estimator. And this is actually the inverse of the variance of $m_i(\beta)$.

Now, the problem is that how can we get the asymptotic estimation of the variance $(C'\Lambda^{-1}C)^{-1}$.

For C, we can estimate it by:

$$\hat{C} = \frac{Z'X}{n} \xrightarrow{p} C$$

10.3.3.1 Homoskedastic Case

For Λ, in the homoskedastic case, we have that:

$$\Lambda = Var(z_i \varepsilon_i) = \sigma^2 \mathbb{E}[z_i z_i']$$

and it can be estimated by:

$$\hat{W}_{opt} = \left(\hat{\sigma}^2 \frac{Z'Z}{n} \right)^{-1} \xrightarrow{p} \Lambda^{-1}$$

Thus, the efficient GMM estimator is actually:

$$
\begin{aligned}
\hat{\beta} &= (X'Z\hat{W}Z'X)^{-1}X'Z\hat{W}Z'Y \\
&= (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'Y \\
&= (X'P_z X)^{-1}X'P_z Y
\end{aligned}
$$

This is actually the 2 steps least squares estimator.

10.3.3.2 Hetrokedastic Case

For the hetroskedastic case, where:

$$Var[\varepsilon_i|z_i] \neq \sigma^2$$

We can first find a random matrix satisfying **A.GMM4**, for example:

$$\hat{W} = \left(\frac{Z'Z}{n}\right)^{-1}$$

Then, we use this to get the consistent GMM estimator $\tilde{\beta}$, and compute the residuals:

$$\tilde{\varepsilon}_i = y_i - x_i'\tilde{\beta}$$

Then, we can estimate the $\hat{W}_{opt}$ by:

$$\hat{W}_{opt} = \left(\frac{1}{n}\sum_{i=1}^{n}\tilde{\varepsilon}_i z_i z_i'\right)^{-1}$$

and we can see that this estimator is consistent.(See 9.2)
This is called 2-step GMM estimator.

10.4   Test of Overidentifying Restrictions

If $L > k$, the model is overidentified.
Then, we can use k moment conditions to estimate β consistently, and use the remaining the $L - k$ conditions to test the exogeneity of the instruments.

Now we let $\hat{\beta}$ be the GMM estimator with the optimal weighting matrix $\hat{W}_{opt}$, where $\hat{W}_{opt} \xrightarrow{p} \Lambda^{-1} = Var[m_i(\beta)]^{-1}$. Then we have the theorem:
Theorem 13.1
Suppose the assumptions for the consistency and asymptotic normality of the GMM estimator hold. If $H_0 : \mathbb{E}[m_i(\beta)] = 0$ is true, the minimized value of the objective function with the optimal weighting matrix satisfies:

$$nQ_n(\hat{\beta}) = n\bar{m}_n(\hat{\beta})'\hat{W}_{opt}\bar{m}_n(\hat{\beta}) \xrightarrow{d} \chi^2(L - k)$$

If we use β to replace $\hat{\beta}$, we have that:

$$nQ_n(\beta) = n\bar{m}_n(\beta)'\hat{W}_{opt}\bar{m}_n(\beta) \xrightarrow{d} \chi^2(L)$$

11 Models for Panel Data

11.1 Models for Panel Data

Panel data means that we observe n individuals for T times:

$$y_{it} = x_{it}'\beta + c_i + \varepsilon_{it}, \ i = 1, \ldots, n; \ t = 1, \ldots, T$$

Here c_i is the individual specific effect.

x_{it} excludes all the constant term to c_i.

ε_{it} is i.i.d. error term satisfying $\mathbb{E}[\varepsilon_{it}|x_{it}] = 0$

Depending on the property of c_i, there are 3 models:

1. Pooled Regression

We assume that $c_i = \alpha$ is constant.

2. Fixed Effect

We assume that c_i is unobservable and correlated with x_{it}.

3. Random Effect

We assume that c_i is random but $\mathbb{E}[c_i|x_{it}]$ is constant.

11.2 Pooled Regression Model

Let $X_i = (x_{i1}, \ldots, x_{iT})$ Since c_i is constant, the model is a standard regression model if we assume:

1. $(y_{i1}, \ldots, y_{iT}, X_i, \varepsilon_{i1}, \ldots, \varepsilon_{iT})$ is i.i.d. across i.

2. $(\varepsilon_{i1}, \ldots, \varepsilon_{iT})$ is independent of X_i (must be X_i, x_{it} may lead to endogeneity).

3. ε_{it} has 0 mean, $\mathbb{E}[\varepsilon_{it}|x_{it}] = 0$.

And under these assumptions, pooled OLS is consistent, asymptotically normal, and efficient.

11.3 Fixed Effects Model

The fixed effects model means there is a correlation between α_i and x_{it}:

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \ i = 1, \ldots, n; \ t = 1, \ldots, T$$

If we only regress y on x, the estimator is biased.

11.3.1 Fixed Effects Transformation

To eliminate the individual effects of α_i, we consider compute the average for each i:

$$\bar{y}_i = \bar{x}_i'\beta + \alpha_i + \bar{\varepsilon}_i$$

And we subtract it from the original model:

$$y_{it} - \bar{y}_i = (x_{it}' - \bar{x}_i')\beta + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

The pooled OLS estimator of this transformed model is called the fixed effects estimator, which is written as:

$$\begin{aligned}
b_{FE} &= \frac{S_{xy}^{within}}{S_{xx}^{within}} \\
&= \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)}{\sum_{i=1}^{n}\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)'}
\end{aligned}$$

11.3.2 The Matrix Form of the Fixed Effects Model

Here, we define:

$$y_i = \begin{bmatrix} y_{i1} \\ \vdots \\ y_{iT} \end{bmatrix}, X_i = \begin{bmatrix} x_{i1}' \\ \vdots \\ x_{iT}' \end{bmatrix}, \varepsilon_i = \begin{bmatrix} \varepsilon_{i1} \\ \vdots \\ \varepsilon_{iT} \end{bmatrix}$$

And we define the within transformation matrix:

$$M^0 = I_T - \iota(\iota'\iota)^{-1}\iota'$$

where ι is a $T \times 1$ vector:

$$\iota = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}$$

Then, we can see that:

$$\begin{bmatrix} y_{i1} - \bar{y}_i \\ \vdots \\ y_{iT} - \bar{y}_i \end{bmatrix} = M^0 y_i, \begin{bmatrix} x'_{i1} - \bar{x}'_i \\ \vdots \\ x'_{iT} - \bar{x}'_i \end{bmatrix} = M^0 X_i$$

Thus, the fixed effects estimator is actually:

$$b_{FE} = \left(\sum_{i=1}^{n} X'_i M^0 X_i \right)^{-1} \sum_{i=1}^{n} X'_i M^0 y_i$$

11.3.3 The Dummy Variable Method

Another method is to add a dummy variable for each individual:

$$y_i = X_i \beta + \iota_i \alpha_i + \varepsilon_i$$

Then, we define:

$$Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}, X = \begin{bmatrix} X_1 \\ \vdots \\ X_n \end{bmatrix}, D = \begin{bmatrix} \iota & & 0 \\ & \ddots & \\ 0 & & \iota \end{bmatrix}, \alpha = \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{bmatrix}, \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

Then, the model is:

$$Y = X\beta + D\alpha + \varepsilon$$

Then, we define:

$$
\begin{aligned}
M_D &= I - D(D'D)^{-1}D' \\
&= \begin{bmatrix} I_T - \iota(\iota'\iota)^{-1}\iota' & & 0 \\ & \ddots & \\ 0 & & I_T - \iota(\iota'\iota)^{-1}\iota' \end{bmatrix} \\
&= \begin{bmatrix} M^0 & & 0 \\ & \ddots & \\ 0 & & M^0 \end{bmatrix}
\end{aligned}
$$

Then, by FWL:

$$\boldsymbol{b}_{LSDV} = (X'M_DX)^{-1}X'M_DY$$

$$= \left(\sum_{i=1}^{n} X_i'M^0X_i\right)^{-1} \sum_{i=1}^{n} X_i'M^0y_i$$

$$= \boldsymbol{b}_{FE}$$

11.3.4   Estimation of the Individual Effect

From the OLS F.O.C, we know that:

$$D'(Y - Xb_{LSDV} - Da) = 0$$

We can solve out the estimated individual effect:

$$a = (D'D)^{-1}D'(Y - Xb_{LSDV})$$

For each row:

$$a_i = \bar{y}_i - \bar{x}_i'\boldsymbol{b}_{LSDV} = \alpha_i + \bar{x}_i'\beta + \bar{\varepsilon}_i - \bar{x}_i'\boldsymbol{b}_{LSDV}$$

Even $\boldsymbol{b}_{LSDV}$ is consistent, a_i still not consistent.

11.3.5   The Least Squares Estimation of the Fixed Effects Model

We gives the assumptions of the fixed effects model:
A.FE1 (Linearity)

$$y_{it} = x_{it}'\beta + \alpha_i + \varepsilon_{it}, \ i = 1, \ldots, n, \ t = 1, \ldots, T$$

A.FE2 (Random Sampling)
$\{y_{it}, x_{it}, \varepsilon_{it}\}_{t=1}^{T}, \ i = 1, \ldots, n$ is a sequence of i.i.d. observations.
A.FE3 (Strict Exogeneity)

$$\mathbb{E}[\varepsilon_{is}|\{x_{it}\}_{t=1}^{T}] = 0, \ \forall \ s = 1, \ldots, T$$

A.FE4 (Homoskedasticity and nonautocorrelation)

$$Var[\varepsilon_{it}|\{\boldsymbol{x}_{it}\}_{t=1}^{T}] = \sigma_{\varepsilon}^2$$

$$Cov[\varepsilon_{it}, \varepsilon_{is}|\{\boldsymbol{x}_{it}\}_{t=1}^{T}] = 0, \ \forall \ t \neq s$$

Then, from the LSDV estimator, we have that:

$$\begin{aligned}
\boldsymbol{b}_{FE} &= (X'M_DX)^{-1}X'M_DY \\
&= \boldsymbol{\beta} + (X'M_DX)^{-1}X'M_D\varepsilon
\end{aligned}$$

11.3.5.1 The Unbiasedness and Consistency

First, we know that:

$$\mathbb{E}[\boldsymbol{b}_{FE}|X] = \boldsymbol{\beta}$$

It is unbiased.

Second, for consistency:

$$\begin{aligned}
\boldsymbol{b}_{FE} &= \boldsymbol{\beta} + \left(\frac{X'M_DX}{n}\right)^{-1}\frac{X'(\varepsilon - \bar{\varepsilon})}{n} \\
&\xrightarrow{p} \boldsymbol{\beta} + \mathbb{E}[X'M_DX]^{-1}(\mathbb{E}[X'\varepsilon] - \mathbb{E}[X']\bar{\varepsilon}) = \boldsymbol{\beta}
\end{aligned}$$

Thus, it is consistent.

11.3.5.2 The Efficiency of the Estimator

We consider the variance of the fixed effects model estimator:

$$\begin{aligned}
Var[\boldsymbol{b}_{FE}|X] &= \mathbb{E}[(X'M_DX)^{-1}X'M_D\varepsilon\varepsilon'M_DX(X'M_DX)^{-1}|X] \\
&= \sigma_{\varepsilon}^2(X'M_DX)^{-1} \\
&= \sigma_{\varepsilon}^2\left(\sum_{i=1}^{n} X_i'M^0X_i\right)^{-1}
\end{aligned}$$

Consider the pooled OLS estimator:

$$\boldsymbol{b}_{pooled} = (X'X)^{-1}X'Y$$

its variance is:

$$Var[\boldsymbol{b}_{pooled}|X] = \sigma_{\varepsilon}^2 (\sum_{i=1}^{n} X'X)^{-1}$$

Because the demeaned variables $\boldsymbol{M^0 X_i}$ have less variation than the original regressors $\boldsymbol{X_i}$, thus we have:

$$\sum_{i=1}^{n} X'X \geq \sum_{i=1}^{n} X_i' \boldsymbol{M^0} X_i$$

which means that

$$Var[\boldsymbol{b}_{pooled}|X] \leq Var[\boldsymbol{b}_{FE}|X]$$

We sacrifices some of the efficiency for the unbiasedness of the estimator.

11.3.5.3 The Asymptotic Estimation of the Variance

We know that:

$$
\begin{aligned}
\boldsymbol{e} &= \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}_{FE} - \boldsymbol{D}\boldsymbol{a} \\
&= \boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}_{FE} - \boldsymbol{P_D}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{b}_{FE}) \\
&= \boldsymbol{M_D}\boldsymbol{Y} - \boldsymbol{M_D}\boldsymbol{X}\boldsymbol{b}_{FE} \\
&= \boldsymbol{M_D}\boldsymbol{Y} - \boldsymbol{M_D}\boldsymbol{X}(\boldsymbol{X'M_D X})^{-1}\boldsymbol{X'M_D Y} \\
&= (\boldsymbol{I} - \boldsymbol{M_D}\boldsymbol{X}(\boldsymbol{X'M_D X})^{-1}\boldsymbol{X'M_D})\boldsymbol{M_D Y} \\
&= (\boldsymbol{I} - \boldsymbol{M_D}\boldsymbol{X}(\boldsymbol{X'M_D X})^{-1}\boldsymbol{X'M_D})\boldsymbol{M_D \varepsilon}
\end{aligned}
$$

Then:

$$
\begin{aligned}
\mathbb{E}[\boldsymbol{e'e}|X] &= \mathbb{E}[\boldsymbol{\varepsilon' M_D}(\boldsymbol{I} - \boldsymbol{M_D}\boldsymbol{X}(\boldsymbol{X'M_D X})^{-1}\boldsymbol{X'M_D})\boldsymbol{M_D \varepsilon}|X] \\
&= \mathbb{E}[tr(\boldsymbol{M_D}(\boldsymbol{I} - \boldsymbol{M_D}\boldsymbol{X}(\boldsymbol{X'M_D X})^{-1}\boldsymbol{X'M_D})\boldsymbol{M_D \varepsilon' \varepsilon})|X] \\
&= \sigma_{\varepsilon}^2 \mathbb{E}[tr(\boldsymbol{M_D} - \boldsymbol{M_D}\boldsymbol{X}(\boldsymbol{X'M_D X})^{-1}\boldsymbol{X'M_D})|X] \\
&= \sigma_{\varepsilon}^2 \mathbb{E}[tr(\boldsymbol{M_D}) - tr((\boldsymbol{X'M_D X})^{-1}(\boldsymbol{X'M_D X}))|X] \\
&= \sigma_{\varepsilon}^2 (nT - n - K)
\end{aligned}
$$

Thus, the unbiased estimator for variance σ_ε^2 is:

$$s^2 = \frac{e'e}{nT - n - K}$$

11.3.5.4 Small T Asymptotics Under Homoskedasticity

From the i.i.d. LLN, we have that:

$$\sqrt{n}\frac{X'M_D\varepsilon}{n} \xrightarrow{d} N(0, \sigma_\varepsilon^2 \mathbb{E}[X_i'M_D X_i])$$

Then, we have that:

$$\sqrt{n}(b - \beta) = (\frac{X'M_D X}{n})^{-1}\sqrt{n}\frac{X'M_D\varepsilon}{n}$$

$$\xrightarrow{d} N(0, \sigma_\varepsilon^2 \mathbb{E}[X_i'M_D X_i]) = N(0, \sigma_\varepsilon^2 \mathbb{E}[\sum_{t=1}^{T}(x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)']^{-1})$$

11.3.6   First-Difference Estimator

Another way to deal with the individual effects is to do the first-differencing transformation of y_{it}:

$$\Delta y_{it} = y_{it} - y_{i,t-1}$$

The new model becomes:

$$\Delta y_{it} = \Delta x_{it}'\beta + \Delta \varepsilon_{it}$$

The estimator for this model is called the first-difference estimator.

For $T = 2$, the first-difference estimator is identical to b_{FE}.

The problem is that there is always a autocorrelation problem in $\Delta \varepsilon_{it}$.

11.4   Radom Effects Model

If the individual effects α_i are uncorrelated with the regressors, it can treated as a random disturbance.

For example, we decompose the random disturbance α_i into two parts α and u_i. Then the model becomes:

$$y_{it} = x_{it}'\beta + \alpha + u_i + \varepsilon_{it}$$

We make the following assumptions:

1. $(\boldsymbol{y}_i, \boldsymbol{X}_i, \boldsymbol{u}_i, \boldsymbol{\varepsilon}_i) \sim i.i.d..$
2. Strict exogeneity:

$$\mathbb{E}[\varepsilon_{it}|\boldsymbol{X}_i] = 0, \ \mathbb{E}[u_i|\boldsymbol{X}_i] = 0$$

3. Homoskedasticity:

$$\mathbb{E}[\varepsilon_{it}^2|\boldsymbol{X}_i] = \sigma_\varepsilon^2, \ \mathbb{E}[u_i^2|\boldsymbol{X}_i] = \sigma_u^2$$

4. No correlation between u_i and ε_{it}:

$$\mathbb{E}[u_i\varepsilon_{it}|\boldsymbol{X}_i] = 0$$

5. No serial correlation in ε_{it}:

$$\mathbb{E}[\varepsilon_{it}\varepsilon_{is}|\boldsymbol{X}_i] = 0, \ t \neq s$$

If we compose u_i and ε_{it} to η_{it}, and take α into $\boldsymbol{x}_i$ as a constant term, the model becomes:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \eta_{it}$$

Then, the assumption above gives that:

$$\mathbb{E}[\eta_{it}|\boldsymbol{X}_i] = 0$$
$$\mathbb{E}[\eta_{it}^2|\boldsymbol{X}_i] = \sigma_\varepsilon^2 + \sigma_u^2$$
$$\mathbb{E}[\eta_{it}\eta_{is}|\boldsymbol{X}_i] = \sigma_u^2$$

Then, the regression for ith individual is:

$$\boldsymbol{y}_i = \boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{\eta}_i$$

The mean and the variance of the error term is:

$$\mathbb{E}[\boldsymbol{\eta}_i | X_i] = \mathbf{0}$$

$$Var[\boldsymbol{\eta}_i | X_i] = \mathbb{E}[\boldsymbol{\eta}_i \boldsymbol{\eta}_i' | X_i]$$

$$= \boldsymbol{\Sigma}$$

$$= \begin{bmatrix} \sigma_\varepsilon^2 + \sigma_u^2 & \sigma_u^2 & \cdots & \sigma_u^2 \\ \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \sigma_u^2 \cdots & \sigma_u^2 & \sigma_\varepsilon^2 + \sigma_u^2 \end{bmatrix}$$

$$= \sigma_\varepsilon^2 \boldsymbol{I} + \sigma_u^2 \boldsymbol{u}'$$

11.4.1 Least Squares Estimation

For this model:

$$y_{it} = \boldsymbol{x}_{it}' \boldsymbol{\beta} + \alpha + u_i + \varepsilon_{it}$$

With assumption above, we use the pooled OLS and the estimator is consistent.
If we use the norm fixed effects estimator, it is also consistent.
We have that:

$$\frac{\boldsymbol{e}_{pooled}' \boldsymbol{e}_{pooled}}{nT} = \frac{\boldsymbol{\eta}' \boldsymbol{M} \boldsymbol{\eta}}{nT}$$

$$= \frac{\boldsymbol{\eta}' \boldsymbol{\eta}}{nT} - \frac{\boldsymbol{\eta}' \boldsymbol{P}_X \boldsymbol{\eta}}{nT}$$

$$\xrightarrow{p} \mathbb{E}[\eta_{it}^2] + 0 = \sigma_\varepsilon^2 + \sigma_u^2$$

$$\frac{\boldsymbol{e}_{FE}' \boldsymbol{e}_{FE}}{nT} = \frac{\boldsymbol{\varepsilon}' \boldsymbol{M}_D (\boldsymbol{I} - \boldsymbol{M}_D X (X' \boldsymbol{M}_D X) X' \boldsymbol{M}_D) \boldsymbol{M}_D \boldsymbol{\varepsilon}}{nT}$$

$$\xrightarrow{p} \mathbb{E}[\varepsilon_{it} - \bar{\varepsilon}_i]^2 = (1 - \frac{1}{T}) \sigma_\varepsilon^2$$

11.4.2   Generalized Least Squares

Here we define the covariance matrix:

$$\boldsymbol{\Omega} = \begin{bmatrix} \boldsymbol{\Sigma} & 0 & \dots & 0 \\ 0 & \boldsymbol{\Sigma} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \boldsymbol{\Sigma} \end{bmatrix}$$

Since it is block diagonal, we can get its inverse:

$$\boldsymbol{\Omega}^{-1} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1} & 0 & \dots & 0 \\ 0 & \boldsymbol{\Sigma}^{-1} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \boldsymbol{\Sigma}^{-1} \end{bmatrix}$$

Similar to the GLS in previous chapter, we define:

$$\boldsymbol{\Omega}^{-1/2} = \begin{bmatrix} \boldsymbol{\Sigma}^{-1/2} & 0 & \dots & 0 \\ 0 & \boldsymbol{\Sigma}^{-1/2} & \dots & 0 \\ & & \ddots & \\ 0 & 0 & \dots & \boldsymbol{\Sigma}^{-1/2} \end{bmatrix}$$

where:

$$\boldsymbol{\Sigma}^{-1/2} = (\sigma_\varepsilon^2 \boldsymbol{I} + \sigma_u^2 \boldsymbol{u}')^{-1/2}$$
$$= \frac{1}{\sigma_\varepsilon}(\boldsymbol{I} - \frac{\theta}{T}\boldsymbol{u}')$$
$$\theta = 1 - (\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2})^{1/2}$$

Then we times this matrix in both side:

$$\boldsymbol{\Omega}^{-1/2}\boldsymbol{Y} = \boldsymbol{\Omega}^{-1/2}\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\Omega}^{-1/2}\boldsymbol{\varepsilon}$$

The GLS estimator is:

$$
\begin{aligned}
\boldsymbol{b}_{GLS} &= (\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \\
&= (\sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i'\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i \\
&= (\sum_{i=1}^{n} \boldsymbol{X}_i'\sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{X}_i)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i'\sigma_\varepsilon^2\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}_i
\end{aligned}
$$

which means we can slightly change the model:

$$
\boldsymbol{Y}^\star = \boldsymbol{X}^\star \boldsymbol{\beta} + \boldsymbol{\varepsilon}^\star
$$

where:

$$
\begin{aligned}
\boldsymbol{Y}^\star &= \sigma_\varepsilon \boldsymbol{\Sigma}^{-1/2}\boldsymbol{Y} \\
&= \begin{bmatrix} y_{i1} - \theta\bar{y}_i \\ \vdots \\ y_{nT} - \theta\bar{y}_n \end{bmatrix}
\end{aligned}
$$

Similar for $\boldsymbol{X}^\star$, then the Random effects estimator is:

$$
\boldsymbol{b}_{RE} = (\sum_{i=1}^{n} (\boldsymbol{X}_i^\star)'\boldsymbol{X}_i^\star)^{-1} \sum_{i=1}^{n} (\boldsymbol{X}_i^\star)'\boldsymbol{Y}_i^\star
$$

and the RE transformation uses quasi-demeaned data:

$$
\begin{aligned}
y_{it}^\star &= y_{it} - \theta\bar{y}_i \\
\boldsymbol{x}_{it}^\star &= \boldsymbol{x}_{it} - \theta\bar{\boldsymbol{x}}_i
\end{aligned}
$$

where:

$$
\theta = 1 - \left(\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T\sigma_u^2} \right)^{1/2}
$$

Then, we can see that:

if $\theta \simeq 0 \Rightarrow$ RE estimator $\simeq$ pooled OLS

if $\theta \simeq 1 \Rightarrow$ RE estimator $\simeq$ FE estimtaor

11.4.3 Feasible GLS

To get the estimation of σ_u^2 and σ_ε^2, we can use the fact that:

$$\frac{\boldsymbol{e}'_{pooled}\boldsymbol{e}_{pooled}}{nT} \xrightarrow{p} \sigma_\varepsilon^2 + \sigma_u^2$$

$$\frac{\boldsymbol{e}'_{FE}\boldsymbol{e}_{FE}}{nT} \xrightarrow{p} (1 - \frac{1}{T})\sigma_\varepsilon^2$$

We can compute the estimator:

$$\hat{\sigma}_\varepsilon^2 = \frac{\boldsymbol{e}'_{FE}\boldsymbol{e}_{FE}}{nT - n} \xrightarrow{p} \sigma_\varepsilon^2$$

$$\hat{\sigma}_u^2 = \frac{\boldsymbol{e}'_{FE}\boldsymbol{e}_{FE}}{nT} - \hat{\sigma}_\varepsilon^2 \xrightarrow{p} \sigma_u^2$$

11.4.4 Hausman Specification Test for the Random Effects Model

We want to test that the individual effects is random or individual specific. If it is random, the RE estimator is consistent. If it is individual specific, the FE estimator is consistent.

Then, we can construct the H_0:

$$H_0 : \text{ the assumptions of the random effects model are correct}$$

Then, we can construct the Wald statistics:

$$W = n(\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE})'\hat{\boldsymbol{\Psi}}^{-1}(\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE})$$

where:

$$\hat{\boldsymbol{\Psi}} \xrightarrow{p} \boldsymbol{\Psi}$$

If H_0 is correct, we know that:

$$\mathbb{E}[\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE}] = \boldsymbol{0}$$

$$Var[\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE}] = \mathbb{E}[(\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE})(\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE})']$$

$$= Var[\boldsymbol{b}_{FE}] - Var[\boldsymbol{b}_{RE}] = \boldsymbol{\Psi}$$

By i.i.d. LLN, we have that:

$$\sqrt{n}(\boldsymbol{b}_{FE} - \boldsymbol{b}_{RE}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Psi})$$

Then, we can see that under H_0, we have that:

$$W \xrightarrow{d} \chi^2(dim(\boldsymbol{\beta}))$$

One thing very important is that the Hausman test relies on the assumptions **2,3,4,5**:

$$\mathbb{E}[\boldsymbol{\eta}_i \boldsymbol{\eta}_i' | X_i] = \boldsymbol{\Sigma}$$

11.5 Cluster-Robust Covariance Matrix Estimation

In the previous chapter, we often assume that ε_{it} is homoskedastic and non-autocorrelation. In this part, we relax this part.

Here, we give the general model for the panel data, depending on the property of α and w_i, this model can be transformed to the 3 models:

$$y_{it} = \boldsymbol{x}_{it}'\boldsymbol{\beta} + \alpha + w_{it}$$

For every individual, we give the regression matrix:

$$\boldsymbol{y}_i = X_i\boldsymbol{\beta} + \boldsymbol{w}_i$$

Here, we only assume:

1. $(\boldsymbol{y}_i, X_i, \boldsymbol{w}_i) \sim i.i.d.$ for i
2. $\mathbb{E}[\boldsymbol{w}_i | X_i] = \boldsymbol{0}$ (Strict exogeneity)

We assume that:

$$\mathbb{E}[\boldsymbol{w}_i \boldsymbol{w}_i' X_i] = \boldsymbol{\Omega}_i$$

here it may not be diagonal.

11.5.1 Pooled Model

For the pooled model:

$$
\begin{aligned}
\boldsymbol{b}_{pooled} &= (X'X)^{-1}X'Y \\
&= \boldsymbol{\beta} + (X'X)^{-1}X'W \\
&= \boldsymbol{\beta} + (\sum_{i=1}^{n} X_i'X_i)^{-1} \sum_{i=1}^{n} X_i'\boldsymbol{w}_i
\end{aligned}
$$

Then, for the asymptotic normality:

$$
\sqrt{n}(\boldsymbol{b}_{pooled} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{Q}_x^{-1}\boldsymbol{Q}_x^w\boldsymbol{Q}_x^{-1})
$$

where:

$$
\begin{aligned}
\boldsymbol{Q}_x &= \mathbb{E}[X_i'X_i] \\
\boldsymbol{Q}_x^w &= \mathbb{E}[X_i'\boldsymbol{w}_i\boldsymbol{w}_i'X_i] = \mathbb{E}[X_i'\boldsymbol{\Omega}_iX_i]
\end{aligned}
$$

For the feasible estimation:

$$
\begin{aligned}
\hat{\boldsymbol{Q}}_x &= \frac{1}{n}\sum_{i=1}^{n} X_i'X_i \xrightarrow{p} \boldsymbol{Q}_x \\
\hat{\boldsymbol{Q}}_x^w &= \frac{1}{n}\sum_{i=1}^{n} X_i'\hat{\boldsymbol{w}}_i\hat{\boldsymbol{w}}_i'X_i \xrightarrow{p} \boldsymbol{Q}_x^w
\end{aligned}
$$

11.5.2 Random Effects Model

Similar to the previous model:

$$
\begin{aligned}
\boldsymbol{b}_{RE} &= (\sum_{i=1}^{n} (X_i^\star)'X_i^\star)^{-1} \sum_{i=1}^{n} (X_i^\star)'Y_i^\star \\
&= \boldsymbol{\beta} + (\sum_{i=1}^{n} (X_i^\star)'X_i^\star)^{-1} \sum_{i=1}^{n} (X_i^\star)'\boldsymbol{w}_i^\star
\end{aligned}
$$

The asymptotic normality is:

$$
\sqrt{n}(\boldsymbol{b}_{RE} - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, (\boldsymbol{Q}_x^\star)^{-1}\boldsymbol{Q}_x^{w\star}(\boldsymbol{Q}_x^\star)^{-1})
$$

where:

$$Q_x^\star = \mathbb{E}[(X_i^\star)' X_i^\star]$$
$$Q_x^{w\star} = \mathbb{E}[(X_i^\star)' w_i^\star (w_i^\star)' X_i^\star]$$

For the feasible estimation:

$$\hat{Q}_x^\star = \frac{1}{n} \sum_{i=1}^n (X_i^\star)' X_i^\star$$
$$\hat{Q}_x^{w\star} = \frac{1}{n} \sum_{i=1}^n (X_i^\star)' \hat{w}_i^\star (\hat{w}_i^\star)' X_i^\star$$

where:

$$\hat{w}_i = Y_i^\star - X_i^\star b_{RE}$$

11.5.3 Fixed Effects Model

We know that:

$$b_{FE} = (X' M_D X)^{-1} X' M_D Y$$
$$= \beta + \left(\sum_{i=1}^n X_i' M^0 X_i \right)^{-1} \sum_{i=1}^n X_i' M^0 \varepsilon_i$$

The asymptotic normality is:

$$\sqrt{n}(b_{FE} - \beta) \xrightarrow{d} N(0, (\dot{Q}_x)^{-1} \dot{Q}_x^w (\dot{Q}_x)^{-1})$$

where:

$$\dot{Q}_x = \mathbb{E}[(X_i)' M^0 X_i]$$
$$\dot{Q}_x^w = \mathbb{E}[(X_i)' M^0 w_i (w_i)' M^0 X_i]$$

For the feasible estimation:

$$\hat{Q}_x = \frac{1}{n} \sum_{i=1}^n (X_i)' M^0 X_i$$
$$\hat{Q}_x^w = \frac{1}{n} \sum_{i=1}^n (X_i)' M^0 \hat{w}_i (\hat{w}_i)' M^0 X_i$$

where:

$$\hat{w}_i = M^0 Y_i - M^0 X_i b_{FE}$$