Lest square estimation method

LS estimator

LS method - assumptions

Lets consider a linear regression

$$y_i = G(x_i; \theta) + u_i$$

Models:

- linear model: $G(x_i; \theta) = x_i\theta$
- nonlinear model: $G(x_i; \theta) = x_i\beta_1 + x_i\beta_2\Phi(z_i; \mu, \sigma^2)$ with $\Phi(z_i; \mu, \sigma^2)$ being a probit function,
- factor model: $G = \lambda' F_i$ with both λ and F_i unknown.

Linear regression - LS estimator

Lets consider a linear regression

$$y_i = x_i\theta + u_i$$

then

$$y = x\theta + u$$

Notation:

- $y = [y_1, ..., y_N]'$ - a $(N \times 1)$ vector of endogenous variables,
- $x = [x_1', ..., x_N']'$ - a $(N \times K)$ matrix of exogenous variables,
- $u = [u_1, ..., u_N]'$ - a $(N \times 1)$ vector of random noise,
- θ is a $(K \times 1)$ vector of parameters.

LS method

Least square estimation method aims at minimizing a sum of squares of residuals

$$\hat{\theta} = argmin_{\theta \in \Theta} \sum_{i=1}^{N} u_i^2 = argmin_{\theta \in \Theta} u' u$$

Remark: The LS estimation method is an optimization problem. It can be applied for linear and nonlinear functions, even when not all assumptions of CLS are valid (then the estimator may not have all desired properties).

LS method

Least square estimation method aims at minimizing a sum of squares of residuals

$$\hat{\theta} = argmin_{\theta\in\Theta} \sum_{i=1}^{N} u_i^2 = argmin_{\theta\in\Theta} u'u$$

Remark: The LS estimation method is an optimization problem. It can be applied for linear and nonlinear functions, even when not all assumptions of CLS are valid (then the estimator may not have all desired properties).

LS estimator

In a linear model

$$y = x\theta + u$$

hence

$$\hat{\theta} = argmin_{\theta\in\Theta}(y - x\theta)'(y - x\theta)$$

In a nonlinear model

$$y = G(x, \theta) + u$$

$$\hat{\theta} = argmin_{\theta\in\Theta}(y - G(x, \theta))'(y - G(x, \theta))$$

LS estimator

In a linear model

$$y = x\theta + u$$

hence

$$\hat{\theta} = argmin_{\theta \in \Theta}(y - x\theta)'(y - x\theta)$$

In a nonlinear model

$$y = G(x, \theta) + u$$

$$\hat{\theta} = argmin_{\theta \in \Theta}(y - G(x, \theta))'(y - G(x, \theta))$$

F.O.C

We want to minimize the loss function

$$L = u'u$$

What are the F.O.C. for

- linear model with $u = y - x\theta$?

$$-2(y - x\theta)'x = 0$$

- non-linear model with $u = y - G(\theta; x)$?

$$-2(y - G(\theta; x))' \frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

F.O.C

We want to minimize the loss function

$$L = u'u$$

What are the F.O.C. for

- linear model with $u = y - x\theta$?

$$-2(y - x\theta)'x = 0$$

- non-linear model with $u = y - G(\theta; x)$?

$$-2(y - G(\theta; x))' \frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

F.O.C

We want to minimize the loss function

$$L = u'u$$

What are the F.O.C. for

- linear model with $u = y - x\theta$?

$$-2(y - x\theta)'x = 0$$

- non-linear model with $u = y - G(\theta; x)$?

$$-2(y - G(\theta; x))'\frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

F.O.C

We want to minimize the loss function

$$L = u'u$$

What are the F.O.C. for

- linear model with $u = y - x\theta$?

$$-2(y - x\theta)'x = 0$$

- non-linear model with $u = y - G(\theta; x)$?

$$-2(y - G(\theta; x))'\frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

F.O.C

We want to minimize the loss function

$$L = u'u$$

What are the F.O.C. for

- linear model with $u = y - x\theta$?

$$-2(y - x\theta)'x = 0$$

- non-linear model with $u = y - G(\theta; x)$?

$$-2(y - G(\theta; x))'\frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

Examples

Examples of models

1. $y_i = \mu + u_i$
2. $y_i = \mu + \beta_1 x_{1i} + \ldots + \beta_{K-1} x_{K-1i} + u_i$
3. $y_i = \alpha x_{1i}^{\beta} x_{2i}^{\gamma} + u_i$

Model 1

In the Model 1 $y_i = \mu + u_i$ Model residuals are

$$u_i = y_i - \mu = y_i - x_i\mu$$

with $x_i = 1$

$$u = y - x\mu$$

$$\hat{\mu} = (x'x)^{-1}x'y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Model 1

In the Model 1 $y_i = \mu + u_i$ Model residuals are

$$u_i = y_i - \mu = y_i - x_i\mu$$

with $x_i = 1$

$$u = y - x\mu$$

$$\hat{\mu} = (x'x)^{-1}x'y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Model 1

In the Model 1 $y_i = \mu + u_i$ Model residuals are

$$u_i = y_i - \mu = y_i - x_i\mu$$

with $x_i = 1$

$$u = y - x\mu$$

$$\hat{\mu} = (x'x)^{-1}x'y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Model 1

In the Model 1 $y_i = \mu + u_i$ Model residuals are

$$u_i = y_i - \mu = y_i - x_i\mu$$

with $x_i = 1$

$$u = y - x\mu$$

$$\hat{\mu} = (x'x)^{-1}x'y = \frac{1}{N}\sum_{i=1}^{N} y_i$$

Model 2

In the Model 2

$$y_i = x_i'\theta + u_i$$

with $\theta = [\mu, \beta_1, ..., \beta_{K-1}]'$ and $x_i = [1, x_{i1}, ..., x_{iK-1}]$ the model is

$$y - x\theta$$

LS estimator

$$\hat{\theta} = (x'x)^{-1}x'y$$

Model 2

In the Model 2

$$y_i = x_i'\theta + u_i$$

with $\theta = [\mu, \beta_1, ..., \beta_{K-1}]'$ and $x_i = [1, x_{i1}, ..., x_{iK-1}]$ the model is

$$y - x\theta$$

LS estimator

$$\hat{\theta} = (x'x)^{-1}x'y$$

Model 2

In the Model 2

$$y_i = x_i'\theta + u_i$$

with $\theta = [\mu, \beta_1, ..., \beta_{K-1}]'$ and $x_i = [1, x_{i1}, ..., x_{iK-1}]$ the model is

$$y - x\theta$$

LS estimator

$$\hat{\theta} = (x'x)^{-1}x'y$$

Model 3

The parameter vector is

$$\theta = [\alpha, \beta, \gamma]'$$

The function $G(\theta; x_i) = \alpha x_{1i}^{\beta} x_{2i}^{\gamma}$ and the residuals are

$$u_i = y_i - G(\theta; x_i)$$

Hence F.O.C. are

$$u' \frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

Model 3

The parameter vector is

$$\theta = [\alpha, \beta, \gamma]'$$

The function $G(\theta; x_i) = \alpha x_{1i}^{\beta} x_{2i}^{\gamma}$ and the residuals are

$$u_i = y_i - G(\theta; x_i)$$

Hence F.O.C. are

$$u' \frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

Model 3

The parameter vector is

$$\theta = [\alpha, \beta, \gamma]'$$

The function $G(\theta; x_i) = \alpha x_{1i}^{\beta} x_{2i}^{\gamma}$ and the residuals are

$$u_i = y_i - G(\theta; x_i)$$

Hence F.O.C. are

$$u' \frac{\partial G(\theta; x)}{\partial \theta'} = 0$$

Model 3

The derivative is

$$\frac{\partial G(\theta, x_i)}{\partial \theta'} = x_{1i}^{\beta} x_{2i}^{\gamma} \left[\begin{array}{ccc} 1, & \alpha \ln x_{1i}, & \alpha \ln x_{2i} \end{array} \right]$$

F.O.C. don't have a closed form solution and...

... needs to be approximated numerically (NEWTON algorithm...) or estimated with a nonlinear LS algorithm

Model 3

The derivative is

$$\frac{\partial G(\theta, x_i)}{\partial \theta'} = x_{1i}^{\beta} x_{2i}^{\gamma} \left[\begin{array}{ccc} 1, & \alpha \ln x_{1i}, & \alpha \ln x_{2i} \end{array} \right]$$

F.O.C. don't have a closed form solution and...
... needs to be approximated numerically (NEWTON algorithm...) or estimated with a nonlinear LS algorithm

LS estimation

To sum up

- For a linear model, the LS estimator has a closed form solution
$$\hat{\theta} = \left(X'X\right)^{-1} X'Y$$
- For a nonlinear model, there may not be a closed form solution for the LS problem. The minimum can be approximated numerically or with a nonlinear LS algorithm.

LS estimation

To sum up

- For a linear model, the LS estimator has a closed form solution
$$\hat{\theta} = \left(X'X\right)^{-1} X'Y$$
- For a nonlinear model, there may not be a closed form solution for the LS problem. The minimum can be approximated numerically or with a nonlinear LS algorithm.

CLS estimation

Sometimes, we need to estimate the constraint model, for example:

- identity constraint: $\theta_i = r$
- linear constraint: $R\theta = r$
- nonlinear constraint: $R(\theta) = 0$

CLS - identity constraint

Suppose, we work with a linear regression model

$$y = x\theta + u_i$$

and want to estimate the model with a constraint $\theta_1 = 1$. In the constraint model

$$y = x_1 + \tilde{x}\tilde{\theta} + u_i$$

$$\tilde{y} = \tilde{x}\tilde{\theta} + u_i$$

with

- $\tilde{y} = y - x_1$
- $\tilde{x} = [x_2, ..., x_K]$
- $\tilde{\theta} = [\theta_2, ..., \theta_K]'$

CLS - identity constraint

The model

$$\tilde{y} = \tilde{x}\tilde{\theta} + u$$

can be simple estimated with LS

$$\hat{\tilde{\theta}} = (\tilde{x}'\tilde{x})\tilde{x}'\tilde{y}$$

and

$$\hat{\theta} = [1, \hat{\tilde{\theta}}']'$$

CLS - linear restriction

Lets consider a linear model with restriction:

$$\theta_1 - \theta_2 = 1$$

then

$$\theta_1 = \theta_2 + 1$$

and the model becomes

$$y = x_1(\theta_2 + 1) + x_2\theta_2 + \ldots x_K\theta_K + u = x_1 + (x_1 + x_2)\theta_2 + \ldots + u$$

CLS - linear restriction

We can rewrite the model

$$\tilde{y} = \tilde{x}\tilde{\theta} + u$$

with

- $\tilde{y} = y - x_1$
- $\tilde{x} = [x_1 + x_2, x_3, ..., x_K]$
- $\tilde{\theta} = [\theta_2, ..., \theta_K]'$

The estimator

$$\hat{\tilde{\theta}} = (\tilde{x}'\tilde{x})\tilde{x}'\tilde{y}$$

and

$$\hat{\theta} = [\hat{\tilde{\theta}}_2 + 1, \hat{\tilde{\theta}}']'$$

CLS - linear restriction

Suppose, we want to estimate the model with M restrictions

$$R\theta = r$$

Lets assume that $rk(R) = M$. Then the parameters can be ordered in such a way that there exist a $(M \times K - M)$ matrix $\tilde{R}$ and a $(M \times 1)$ vector $\tilde{r}$ such that

$$\theta_{1:M} = \tilde{R}\theta_{M+1:K} + \tilde{r}$$

Question: how to use the restriction to reformulate the model

and estimate the model parameters with LS?

Partial LS

Lets consider a linear regression model with two groups of regressors

$$y = x_1\theta_1 + x_2\theta_2 + u.$$

Suppose, we are interested only in the first vector of parameters: θ_1.

Question: do we need to estimate the whole model?

Partial LS

Suppose, we regress y on x then the estimator is

$$\hat{\theta} = (x'x)^{-1}x'y$$

and residuals are

$$\hat{u} = y - x(x'x)^{-1}x'y = (I - x(x'x)^{-1}x')y = My$$

Notice: M is a symmetric, idempotent matrix and $M = MM$.

Partial LS

Lets look at $(x'x)^{-1}$:

$$(x'x)^{-1} = \begin{pmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{pmatrix}^{-1} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$$

Then

$$\hat{\theta}_1 = A_{11}x_1'y + A_{12}x_2'y$$

Partial LS

What is the inverse of a partition matrix $(x'x)^{-1}$:

$$(x'x)^{-1} = \begin{pmatrix} x_1'x_1 & x_1'x_2 \\ x_2'x_1 & x_2'x_2 \end{pmatrix}^{-1} = \begin{pmatrix} D & -D(x_1'x_2)(x_2'x_2)^{-1} \\ * & * \end{pmatrix}$$

with

$$D = \left((x_1'x_1) - (x_1'x_2)(x_2'x_2)^{-1}(x_2'x_1)\right)^{-1}$$

Partial LS

The partial estimator is

$$\hat{\theta}_1 = Dx_1'y - D(x_1'x_2)(x_2'x_2)^{-1}x_2'y$$

$$\hat{\theta}_1 = Dx_1'(I - x_2(x_2'x_2)^{-1}x_2')y$$

Lets denote by $M_2 = I - x_2(x_2'x_2)^{-1}x_2'$. It is similar to M a symmetric, idempotent matrix. Hence

$$\hat{\theta}_1 = Dx_1'M_2'y$$

Partial LS

It can be shown that

$$D = (x_1' M_2' M_2 x_1)^{-1}$$

Hence

$$\hat{\theta}_1 = (\tilde{x}_1{}'\tilde{x}_1)\tilde{x}_1{}' y$$

where:

- $\tilde{x}_1 = M_2 x_1$ are the residuals from regressing x_1 on x_2!

Partial LS

Remark: when x_1 and x_2 are orthogonal then:

$$\tilde{x}_1 = x_1,$$

and

$$\hat{\theta}_1 = (x_1' x_1) x_1' y$$

In all other cases:

$$\hat{\theta}_1 \neq (x_1' x_1) x_1' y$$

Assumptions and properties of LS estimation method

LS estimator

We know, that for a linear regression model, the LS estimator is

$$\hat{\theta} = \left(x'x\right)^{-1} x'y$$

- What are the necessary conditions needed to estimate the model?
- What are the properties of the estimator?

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i|x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i | x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i | x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i|x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i | x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i | x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

In the literature, the following assumptions are stated:

1. The model is **linear**
2. **Full rank**: the $N \times K$ matrix X, has a full column rank
3. **Exogeneity of the independent variables**: $E(\varepsilon_i | x_j) = 0$ for any i and j. There is no correlation between the disturbances and the independent variables
4. **Homoscedasticity and nonautocorrelation**: each disturbance ε_i has the same variance σ^2 and is uncorrelated with any other disturbance ε_j
5. X are **non stochastic**
6. **Normal distribution**: residuals are normally distributed

Assumptions

Questions

- Which of these assumptions are necessary?
- Why are these assumptions imposed?

Assumptions

Lets go back to the parameter estimator

$$\hat{\theta} = \left(x'x\right)^{-1} x'y$$

It can be computed, only when $(x'x)^{-1}$ exists.

A1: The columns of X are linearly independent (the a2 in CLS)

Assumptions

Lets go back to the parameter estimator

$$\hat{\theta} = \left(x'x\right)^{-1} x'y$$

It can be computed, only when $(x'x)^{-1}$ exists.

A1: The columns of X are linearly independent (the a2 in CLS)

Assumptions

When is the A1 assumption a problem?

- When we use some artificial indicators, which can be a linear function of variables
- When dummy variables are added in the model

Consistency

What conditions are necessary to ensure consistency?

$$\hat{\theta} = \theta + \left(x'x\right)^{-1} x'u$$

Hence

$$\hat{\theta} - \theta = \left(\frac{x'x}{N}\right)^{-1} \frac{x'u}{N}$$

We want

$$\hat{\theta} - \theta \to_p 0_{K\times 1}$$

Consistency

$$\hat{\theta} - \theta = \left(\frac{x'x}{N} \right)^{-1} \frac{x'u}{N}$$

A2: there exists a second, noncentral moment of X

$$\frac{x'x}{N} \rightarrow \Omega$$

A3: there is no correlation between the disturbances and the independent variables

$$\frac{x'u}{N} \rightarrow 0_{K \times 1}$$

Then

$$\hat{\theta} - \theta \rightarrow_p 0_{K \times 1}$$

Unbiasedness

Under conditions A1-A3, the LS estimator is unbiased

$$E(\hat{\theta} - \theta|x) = \left(\frac{x'x}{N}\right)^{-1} \frac{x'u}{N}|X) = 0$$

Hence the conditions A2-A3 are crucial for consistency and unbiasedness.

Unbiasedness

Under conditions A1-A3, the LS estimator is unbiased

$$E(\hat{\theta} - \theta | x) = \left(\frac{x'x}{N} \right)^{-1} \frac{x'u}{N} | X) = 0$$

Hence the conditions A2-A3 are crucial for consistency and unbiasedness.

Assumptions

When the assumptions are not valid:

- When X is nonstationary - some other normalization is required
- Omitted variables - IV estimation method

Assumptions

When the assumptions are not valid:

- When X is nonstationary - some other normalization is required
- Omitted variables - IV estimation method

Assumptions

When the assumptions are not valid:

- When X is nonstationary - some other normalization is required
- Omitted variables - IV estimation method

Example

Suppose, we want to estimate the model

$$Income_i = \mu + \beta_1 Age_i + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Sex_i + \varepsilon_i$$

What variables are not included in the model (left in residuals)?

- Social skills
- Intelligence
- Personal feature

Some of them may be correlated with education and experience.

Example

Suppose, we want to estimate the model

$$Income_i = \mu + \beta_1 Age_i + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Sex_i + \varepsilon_i$$

What variables are not included in the model (left in residuals)?

- Social skills
- Intelligence
- Personal feature

Some of them may be correlated with education and experience.

Example

Suppose, we want to estimate the model

$$Income_i = \mu + \beta_1 Age_i + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Sex_i + \varepsilon_i$$

What variables are not included in the model (left in residuals)?

- Social skills
- Intelligence
- Personal feature

Some of them may be correlated with education and experience.

Example

Suppose, we want to estimate the model

$$Income_i = \mu + \beta_1 Age_i + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Sex_i + \varepsilon_i$$

What variables are not included in the model (left in residuals)?

- Social skills
- Intelligence
- Personal feature

Some of them may be correlated with education and experience.

Example

Suppose, we want to estimate the model

$$Income_i = \mu + \beta_1 Age_i + \beta_2 Edu_i + \beta_3 Exp_i + \beta_4 Sex_i + \varepsilon_i$$

What variables are not included in the model (left in residuals)?

- Social skills
- Intelligence
- Personal feature

Some of them may be correlated with education and experience.

Efficiency

Remember: there is a third feature of an estimator - efficiency.

Question: what is the variance of the estimator?

It can be shown that

$$\Sigma(\hat{\theta} - \theta | x) = \left(x'x\right)^{-1} x' E\left(uu' | x\right) x \left(x'x\right)^{-1}$$

Efficiency

Remember: there is a third feature of an estimator - efficiency.

Question: what is the variance of the estimator?

It can be shown that

$$\Sigma(\hat{\theta} - \theta|x) = \left(x'x\right)^{-1} x'E\left(uu'|x\right) x \left(x'x\right)^{-1}$$

Efficiency

Remember: there is a third feature of an estimator - efficiency.

Question: what is the variance of the estimator?

It can be shown that

$$\Sigma(\hat{\theta} - \theta | x) = \left(x'x\right)^{-1} x' E\left(uu' | x\right) x \left(x'x\right)^{-1}$$

Efficiency

Under the assumption
A4: each disturbance u_i has the same variance σ^2 and is uncorrelated with any other disturbance u_j (homoscedasticity and no autocorrelation)

$$E\left(uu'\right) = \sigma^2 I_N$$

Then

$$\Sigma(\hat{\theta} - \theta | x) = \sigma^2 \left(x'x\right)^{-1}$$

Efficiency

Under the assumption
A4: each disturbance u_i has the same variance σ^2 and is uncorrelated with any other disturbance u_j (homoscedasticity and no autocorrelation)

$$E\left(uu'\right) = \sigma^2 I_N$$

Then

$$\Sigma(\hat{\theta} - \theta | x) = \sigma^2 \left(x'x\right)^{-1}$$

Efficiency

The assumption A4 is needed to compute the variance of the estimator ....

...but is not needed to ensure consistency and unbiasedness.

Efficiency

I order to estimate the variance of the residuals σ^2, two type of estimators are typically used

1. Unbiased, consistent

$$\hat{\sigma}^2 = \frac{e'e}{N-K}$$

2. Biased, consistent

$$\hat{\sigma}^2 = \frac{e'e}{N}$$

Where e is a vector of residuals computed for $\hat{\theta}$.

Efficiency

- Then the variance-covariance matrix of parameters is

$$\Sigma_{\hat{\theta}} = \hat{\sigma}^2 \left(x'x\right)^{-1}$$

- Parameter variances are

$$Var(\hat{\theta}) = diag\left(\Sigma_{\hat{\theta}}\right)$$

- Standard deviation

$$St.Dev(\hat{\theta}) = \sqrt{diag\left(\Sigma_{\hat{\theta}}\right)}$$

Asymptotic distribution

What is the asymptotic distribution of the LS estimator?

Under assumptions A1-A3 we have shown that

$$\hat{\theta} - \theta = \left(\frac{x'x}{N} \right)^{-1} \frac{x'u}{N} \rightarrow_p 0$$

Hence, $\hat{\theta}$ has a degenerated distribution

Asymptotic distribution

What is the asymptotic distribution of the LS estimator?

Under assumptions A1-A3 we have shown that

$$\hat{\theta} - \theta = \left(\frac{x'x}{N} \right)^{-1} \frac{x'u}{N} \rightarrow_p 0$$

Hence, $\hat{\theta}$ has a degenerated distribution

Asymptotic distribution

In order to use the CLT, we need to multiply $\hat{\theta}$ by $\sqrt{N}$. Then

$$\sqrt{N}(\hat{\theta} - \theta) = \left(\frac{x'x}{N}\right)^{-1} \frac{x'u}{\sqrt{N}}$$

From A2, it follows

$$\left(\frac{x'x}{N}\right)^{-1} \to_p \Omega^{-1}$$

Under Grenader conditions (weak conditions, which ensure that X is well behaved)

$$\frac{x'u}{\sqrt{N}} \to N(0, \Phi)$$

Asymptotic distribution

In order to use the CLT, we need to multiply $\hat{\theta}$ by $\sqrt{N}$. Then

$$\sqrt{N}(\hat{\theta} - \theta) = \left(\frac{x'x}{N} \right)^{-1} \frac{x'u}{\sqrt{N}}$$

From A2, it follows

$$\left(\frac{x'x}{N} \right)^{-1} \to_p \Omega^{-1}$$

Under Grenader conditions (weak conditions, which ensure that X is well behaved)

$$\frac{x'u}{\sqrt{N}} \to N(0, \Phi)$$

Asymptotic distribution

In order to use the CLT, we need to multiply $\hat{\theta}$ by $\sqrt{N}$. Then

$$\sqrt{N}(\hat{\theta} - \theta) = \left(\frac{x'x}{N}\right)^{-1} \frac{x'u}{\sqrt{N}}$$

From A2, it follows

$$\left(\frac{x'x}{N}\right)^{-1} \to_p \Omega^{-1}$$

Under Grenader conditions (weak conditions, which ensure that X is well behaved)

$$\frac{x'u}{\sqrt{N}} \to N(0, \Phi)$$

Asymptotic distribution

$$\frac{x'u}{\sqrt{N}} \to N(0, \Phi)$$

What is Φ?

$$\Phi_N = E(\frac{1}{N} x'uu'x)$$

Under A4, it converges to Φ

$$\Phi = \sigma^2 \Omega$$

Asymptotic distribution

$$\frac{x'u}{\sqrt{N}} \to N(0, \Phi)$$

What is Φ?

$$\Phi_N = E(\frac{1}{N}x'uu'x)$$

Under A4, it converges to Φ

$$\Phi = \sigma^2 \Omega$$

Asymptotic distribution

Finally

$$\sqrt{N}(\hat{\theta} - \theta) \to \Omega^{-1} N\left(0, \sigma^2 \Omega\right) = N\left(0, \sigma^2 \Omega^{-1}\right)$$

Remark: We obtain asymptotic normality, without assuming the normality of disturbances.

Asymptotic distribution

Finally

$$\sqrt{N}(\hat{\theta} - \theta) \to \Omega^{-1} N\left(0, \sigma^2 \Omega\right) = N\left(0, \sigma^2 \Omega^{-1}\right)$$

Remark: We obtain asymptotic normality, without assuming the normality of disturbances.

Assumptions

1. The model is **linear**- :)
2. **Full rank**: needed to estimate the model
3. **Exogeneity of the independent variables**: needed for consistency and unbiasedness
4. **Homoscedasticity and nonautocorrelation**: needed to compute the variance-covariance matrix of the estimator and ensure that it is the most efficient estimator
5. X are **non stochastic**: not needed
6. **Normal distribution**: needed only when small sample properties are considered

Assumptions

1. The model is **linear**- :)
2. **Full rank**: needed to estimate the model
3. **Exogeneity of the independent variables**: needed for consistency and unbiasedness
4. **Homoscedasticity and nonautocorrelation**: needed to compute the variance-covariance matrix of the estimator and ensure that it is the most efficient estimator
5. X are **non stochastic**: not needed
6. **Normal distribution**: needed only when small sample properties are considered

Assumptions

1. The model is **linear**- :)
2. **Full rank**: needed to estimate the model
3. **Exogeneity of the independent variables**: needed for consistency and unbiasedness
4. **Homoscedasticity and nonautocorrelation**: needed to compute the variance-covariance matrix of the estimator and ensure that it is the most efficient estimator
5. X are **non stochastic**: not needed
6. **Normal distribution**: needed only when small sample properties are considered

Assumptions

1. The model is **linear**- :)
2. **Full rank**: needed to estimate the model
3. **Exogeneity of the independent variables**: needed for consistency and unbiasedness
4. **Homoscedasticity and nonautocorrelation**: needed to compute the variance-covariance matrix of the estimator and ensure that it is the most efficient estimator
5. X are **non stochastic**: not needed
6. **Normal distribution**: needed only when small sample properties are considered

Assumptions

1. The model is **linear**- :)
2. **Full rank**: needed to estimate the model
3. **Exogeneity of the independent variables**: needed for consistency and unbiasedness
4. **Homoscedasticity and nonautocorrelation**: needed to compute the variance-covariance matrix of the estimator and ensure that it is the most efficient estimator
5. X are **non stochastic**: not needed
6. **Normal distribution**: needed only when small sample properties are considered

Assumptions

1. The model is **linear**- :)
2. **Full rank**: needed to estimate the model
3. **Exogeneity of the independent variables**: needed for consistency and unbiasedness
4. **Homoscedasticity and nonautocorrelation**: needed to compute the variance-covariance matrix of the estimator and ensure that it is the most efficient estimator
5. X are **non stochastic**: not needed
6. **Normal distribution**: needed only when small sample properties are considered