Homoskedasticity

How big is the difference between the OLS estimator and the true parameter? To answer this question, we make an additional assumption called homoskedasticity:

$$\mathrm{Var}\,(u|X) = \sigma^2. \tag{23}$$

This means that the variance of the error term u is the same, regardless of the predictor variable X.

If assumption (23) is violated, e.g. if $\mathrm{Var}\,(u|X) = \sigma^2 h(X)$, then we say the error term is heteroskedastic.

Homoskedasticity

- Assumption (23) certainly holds, if u and X are assumed to be independent. However, (23) is a weaker assumption.

- Assumption (23) implies that σ^2 is also the unconditional variance of u, referred to as error variance:

$$\mathrm{Var}\,(u) = \mathrm{E}(u^2) - (\mathrm{E}(u))^2 = \sigma^2.$$

  Its square root σ is the standard deviation of the error.

- It follows that $\mathrm{Var}\,(Y|X) = \sigma^2$.

Variance of the OLS estimator

How large is the variation of the OLS estimator around the true parameter?

- Difference $\hat{\beta}_1 - \beta_1$ is 0 on average

- Measure the variation of the OLS estimator around the true parameter through the expected squared difference, i.e. the variance:

$$\mathrm{Var}\left(\hat{\beta}_1\right) = \mathrm{E}((\hat{\beta}_1 - \beta_1)^2) \tag{24}$$

- Similarly for $\hat{\beta}_0$: $\mathrm{Var}\left(\hat{\beta}_0\right) = \mathrm{E}((\hat{\beta}_0 - \beta_0)^2)$.

Variance of the OLS estimator

Variance of the slope estimator $\hat{\beta}_1$ follows from (22):

$$\begin{aligned}
\mathrm{Var}\left(\hat{\beta}_1\right) &= \frac{1}{N^2(s_x^2)^2}\sum_{i=1}^{N}(x_i-\overline{x})^2\mathrm{Var}\left(u_i\right) \\
&= \frac{\sigma^2}{N^2(s_x^2)^2}\sum_{i=1}^{N}(x_i-\overline{x})^2 = \frac{\sigma^2}{N s_x^2}. \qquad (25)
\end{aligned}$$

- The variance of the slope estimator is the larger, the smaller the number of observations N (or the smaller, the larger N).
- Increasing N by a factor of 4 reduces the variance by a factor of 1/4.

Variance of the OLS estimator

Dependence on the error variance σ^2:

- The variance of the slope estimator is the larger, the larger the error variance σ^2.

Dependence on the design, i.e. the predictor variable X:

- The variance of the slope estimator is the larger, the smaller the variation in X, measured by s_x^2.

Variance of the OLS estimator

The variance is in general different for the two parameters of the simple regression model. $\mathrm{Var}\left(\hat{\beta}_0\right)$ is given by (without proof):

$$\mathrm{Var}\left(\hat{\beta}_0\right) = \frac{\sigma^2}{N s_x^2} \sum_{i=1}^{N} x_i^2. \tag{26}$$

The standard deviations $\mathrm{sd}(\hat{\beta}_0)$ and $\mathrm{sd}(\hat{\beta}_1)$ of the OLS estimators are defined as:

$$\mathrm{sd}(\hat{\beta}_0) = \sqrt{\mathrm{Var}\left(\hat{\beta}_0\right)}, \quad \mathrm{sd}(\hat{\beta}_1) = \sqrt{\mathrm{Var}\left(\hat{\beta}_1\right)}.$$

Mile stone II

The Multiple Regression Model

- Step 1: Model Definition
- Step 2: OLS Estimation
- Step 3: Econometric Inference
- Step 4: OLS Residuals
- Step 5: Testing Hypothesis

Mile stone II

- Step 6: Model Evaluation and Model Comparison
- Step 7: Residual Diagnostics

Cross-sectional data

- We are interested in a dependent (left-hand side, explained, response) variable Y, which is supposed to depend on K explanatory (right-hand sided, independent, control, predictor) variables $X_1, \ldots, X_K$

- Examples: wage is a response and education, gender, and experience are predictor variables

- we are observing these variables for N subjects drawn randomly from a population (e.g. for various supermarkets, for various individuals):

  $(y_i, x_{1,i}, \ldots, x_{K,i}), i = 1, \ldots, N$

II.1 Model formulation

The multiple regression model describes the relation between the response variable Y and the predictor variables $X_1, \ldots, X_K$ as:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K + u, \tag{27}$$

$\beta_0, \beta_1, \ldots, \beta_K$ are unknown parameters.

Key assumption:

$$\mathrm{E}(u|X_1, \ldots, X_K) = \mathrm{E}(u) = 0. \tag{28}$$

Model formulation

Assumption (28) implies:

$$\mathrm{E}(Y|X_1,\ldots,X_K) = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K. \qquad (29)$$

$\mathrm{E}(Y|X_1,\ldots,X_K)$ is a linear function

- in the parameters $\beta_0, \beta_1, \ldots, \beta_K$ (important for „easy" OLS estimation),
- and in the predictor variables $X_1, \ldots, X_K$ (important for the correct interpretation of the parameters).

Understanding the parameters

The parameter β_k is the expected absolute change of the response variable Y, if the predictor variable X_k is increased by 1, and all other predictor variables remain the same (ceteris paribus):

$$
\begin{aligned}
&E(\Delta Y|\Delta X_k) = \mathrm{E}(Y|X_k = x + \Delta X_k) - \mathrm{E}(Y|X_k = x) = \\
&\beta_0 + \beta_1 X_1 + \ldots + \beta_k (x + \Delta X_k) + \ldots + \beta_K X_K \\
&-(\beta_0 + \beta_1 X_1 + \ldots + \beta_k x + \ldots + \beta_K X_K) = \\
&\beta_k \Delta X_k.
\end{aligned}
$$

Understanding the parameters

The sign shows the direction of the expected change:

- If $\beta_k > 0$, then the change of X_k and Y goes into the same direction.
- If $\beta_k < 0$, then the change of X_k and Y goes into different directions.
- If $\beta_k = 0$, then a change in X_k has no influence on Y.

The multiple log-linear model

The multiple log-linear model reads:

$$Y = e^{\beta_0} \cdot X_1^{\beta_1} \cdots X_K^{\beta_K} e^{u}. \tag{30}$$

The log transformation yields a model that is linear in the parameters $\beta_0, \beta_1, \ldots, \beta_K$,

$$\log Y = \beta_0 + \beta_1 \log X_1 + \ldots + \beta_K \log X_K + u, \tag{31}$$

but is nonlinear in the predictor variables $X_1, \ldots, X_K$. Important for the correct interpretation of the parameters.

The multiple log-linear model

- The coefficient β_k is the elasticity of the response variable Y with respect to the variable X_k, i.e. the expected relative change of Y, if the predictor variable X_k is increased by 1% and all other predictor variables remain the same (ceteris paribus).

- If X_k is increased by $p\%$, then (ceteris paribus) the expected relative change of Y is equal to $\beta_k p\%$. On average, Y increases by $\beta_k p\%$, if $\beta_k > 0$, and decreases by $|\beta_k| p\%$, if $\beta_k < 0$.

- If X_k is decreased by $p\%$, then (ceteris paribus) the expected relative change of Y is equal to $-\beta_k p\%$. On average, Y decreases by $\beta_k p\%$, if $\beta_k > 0$, and increases by $|\beta_k| p\%$, if $\beta_k < 0$.

EVIEWS Exercise II.1.2

Show in EViews, how to define a multiple regression model and discuss the meaning of the estimated parameters:

- Case Study Chicken, work file `chicken`;
- Case Study Marketing, work file `marketing`;
- Case Study profit, work file `profit`;

II.2 OLS-Estimation

Let $(y_i, x_{1,i}, \ldots, x_{K,i}), i = 1, \ldots, N$ denote a random sample of size N from the population. Hence, for each i:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i} + \ldots + \beta_K x_{K,i} + u_i. \tag{32}$$

The population parameters β_0, β_1, and β_K are estimated from a sample. The parameters estimates (coefficients) are typically denoted by $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K$. We will use the following vector notation:

$$\boldsymbol{\beta} = (\beta_0, \ldots, \beta_K)', \quad \hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K)'. \tag{33}$$

II.2 OLS-Estimation

The commonly used method to estimate the parameters in a multiple regression model is, again, OLS estimation:

- For each observation y_i, the prediction $\hat{y}_i(\boldsymbol{\beta})$ of y_i depends on $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_K)$.

- For each y_i, define the regression residuals (prediction error) $u_i(\boldsymbol{\beta})$ as:

$$u_i(\boldsymbol{\beta}) = y_i - \hat{y}_i(\boldsymbol{\beta}) = y_i - (\beta_0 + \beta_1 x_{1,i} + \ldots + \beta_K x_{K,i}). \quad (34)$$

OLS-Estimation for the Multiple Regression Model

- For each parameter value $\boldsymbol{\beta}$, an overall measure of fit is obtained by aggregating these prediction errors.

- The sum of squared residuals (SSR):

$$\mathrm{SSR} = \sum_{i=1}^{N} u_i(\boldsymbol{\beta})^2 = \sum_{i=1}^{N} (y_i - \beta_0 - \beta_1 x_{1,i} - \ldots - \beta_K x_{K,i})^2. \quad (35)$$

- The OLS-estimator $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_K)$ is the parameter that minimizes the sum of squared residuals.

How to compute the OLS Estimator?

For a multiple regression model, the estimation problem is solved by software packages like EViews.

Some mathematical details:

- Take the first partial derivative of (35) with respect to each parameter β_k, $k = 0, \ldots, K$.

- This yields a system $K + 1$ linear equations in $\beta_0, \ldots, \beta_K$, which has a unique solution under certain conditions on the matrix $\mathbf{X}$, having N rows and $K + 1$ columns, containing in each row i the predictor values $(1\, x_{1,i}\, \ldots\, x_{K,i})$.

Matrix notation of the multiple regression model

Matrix notation for the observed data:

$$\mathbf{X} = \begin{pmatrix} 1 & x_{1,1} & \vdots & x_{K,1} \\ 1 & x_{1,2} & \vdots & x_{K,2} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_{1,N-1} & \vdots & x_{K,N-1} \\ 1 & x_{1,N} & \vdots & x_{K,N} \end{pmatrix}, \qquad \mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_{N-1} \\ y_N \end{pmatrix}.$$

$\mathbf{X}$ is $N \times (K+1)$-matrix, $\mathbf{y}$ is $N \times 1$-vector.

The $\mathbf{X}'\mathbf{X}$ is a quadratic matrix with $(K+1)$ rows and columns.

$(\mathbf{X}'\mathbf{X})^{-1}$ is the inverse of $\mathbf{X}'\mathbf{X}$.

Matrix notation of the multiple regression model

In matrix notation, the N equations given in (32) for $i = 1, \ldots, N$, may be written as:

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

where

$$\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_N \end{pmatrix}, \qquad \boldsymbol{\beta} = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_K \end{pmatrix}.$$

The OLS Estimator

The OLS estimator $\hat{\boldsymbol{\beta}}$ has an explicit form, depending on $\mathbf{X}$ and the vector $\mathbf{y}$, containing all observed values $y_1, \ldots, y_N$.

The OLS estimator is given by:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{y}. \tag{36}$$

The matrix $\mathbf{X}^{'}\mathbf{X}$ has to be invertible, in order to obtain a unique estimator $\boldsymbol{\beta}$.

The OLS Estimator

Necessary conditions for $\mathbf{X}'\mathbf{X}$ being invertible:

- We have to observe sample variation for each predictor X_k; i.e. the sample variances of $x_{k,1}, \ldots, x_{k,N}$ is positive for all $k = 1, \ldots, K$.

- Furthermore, no exact linear relation between any predictors X_k and X_l should be present; i.e. the empirical correlation coefficient of all pairwise data sets $(x_{k,i}, x_{l,i})$, $i = 1, \ldots, N$ is different from 1 and -1.

EViews produces an error, if $\mathbf{X}'\mathbf{X}$ is not invertible.

Perfect Multicollinearity

A sufficient assumptions about the predictors $X_1, \ldots, X_K$ in a multiple regression model is the following:

- The predictors $X_1, \ldots, X_K$ are not linearly dependent, i.e. no predictor X_j may be expressed as a linear function of the remaining predictors $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_K$.

If this assumption is violated, then the OLS estimator does not exist, as the matrix $\mathbf{X}'\mathbf{X}$ is not invertible.

There are infinitely many parameters values $\boldsymbol{\beta}$ having the same minimal sum of squared residuals, defined in (35). The parameters in the regression model are not identified.

Case Study Yields

Demonstration in EVIEWS, workfile `yieldus`

$$
\begin{array}{rcl}
y_i & = & \beta_1 + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + u_i, \\
y_i & \ldots & \text{yield with maturity 3 months} \\
x_{2,i} & \ldots & \text{yield with maturity 1 month} \\
x_{3,i} & \ldots & \text{yield with maturity 60 months} \\
x_{4,i} & \ldots & \text{spread between these yields} \\
 & & x_{4,i} = x_{3,i} - x_{2,i}
\end{array}
$$

$x_{4,i}$ is a linear combination of $x_{2,i}$ and $x_{3,i}$

Case Study Yields

Let $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3, \beta_4)$ be a certain parameter

Any parameter $\boldsymbol{\beta}^\star = (\beta_1, \beta_2^\star, \beta_3^\star, \beta_4^\star)$, where $\beta_4^\star$ may be arbitrarily chosen and

$$\beta_3^\star = \beta_3 + \beta_4 - \beta_4^\star$$

$$\beta_2^\star = \beta_2 - \beta_4 + \beta_4^\star$$

will lead to the same sum of mean squared errors as $\boldsymbol{\beta}$. The OLS estimator is not unique.

II.3 Understanding Econometric Inference

Econometric inference: learning from the data about the unknown parameter $\boldsymbol{\beta}$ in the regression model.

- Use the OLS estimator $\hat{\boldsymbol{\beta}}$ to learn about the regression parameter.
- Is this estimator equal to the true value?
- How large is the difference between the OLS estimator and the true parameter?
- Is there a better estimator than the OLS estimator?

Unbiasedness

Under the assumptions (28), the OLS estimator (if it exists) is unbiased, i.e. the estimated values are on average equal to the true values:

$$\mathrm{E}(\hat{\beta}_j) = \beta_j, \qquad j = 0, \ldots, K.$$

In matrix notation:

$$\mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}, \qquad \mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = 0. \tag{37}$$

Unbiasedness of the OLS estimator

If the data are generated by the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{u}$, then the OLS estimator may be expressed as:

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

Therefore the estimation error may be expressed as:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{u}. \tag{38}$$

Result (37) follows immediately:

$$\mathrm{E}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\boldsymbol{u}) = \mathbf{0}.$$

Covariance Matrix of the OLS Estimator

Due to unbiasedness, the expected value $\mathrm{E}(\hat{\beta}_j)$ of the OLS estimator is equal to β_j for $j = 0, \ldots, K$.

Hence, the variance $\mathrm{Var}\left(\hat{\beta}_j\right)$ measures the variation of the OLS estimator $\hat{\beta}_j$ around the true value β_j:

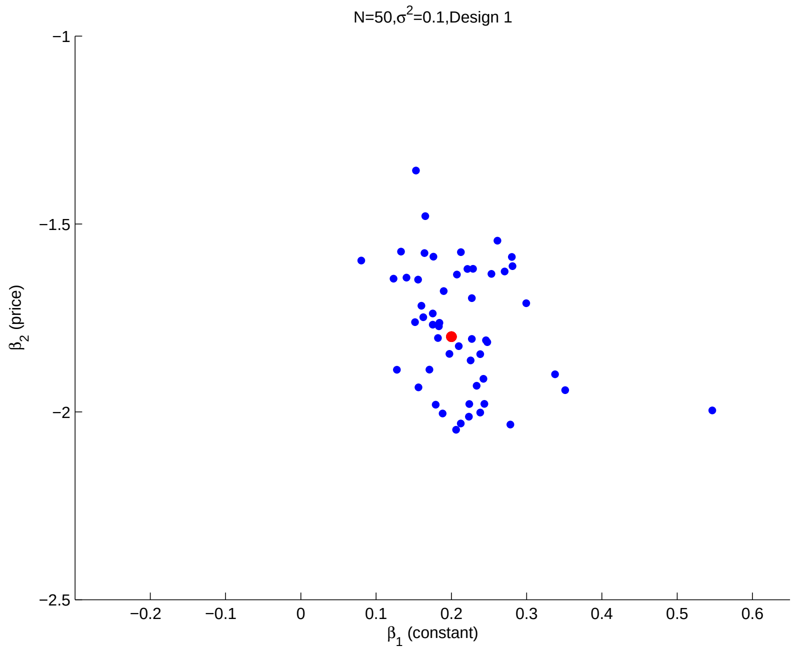
$$\mathrm{Var}\left(\hat{\beta}_j\right) = \mathrm{E}\left((\hat{\beta}_j - \mathrm{E}(\hat{\beta}_j))^2\right) = \mathrm{E}\left((\hat{\beta}_j - \beta_j)^2\right).$$

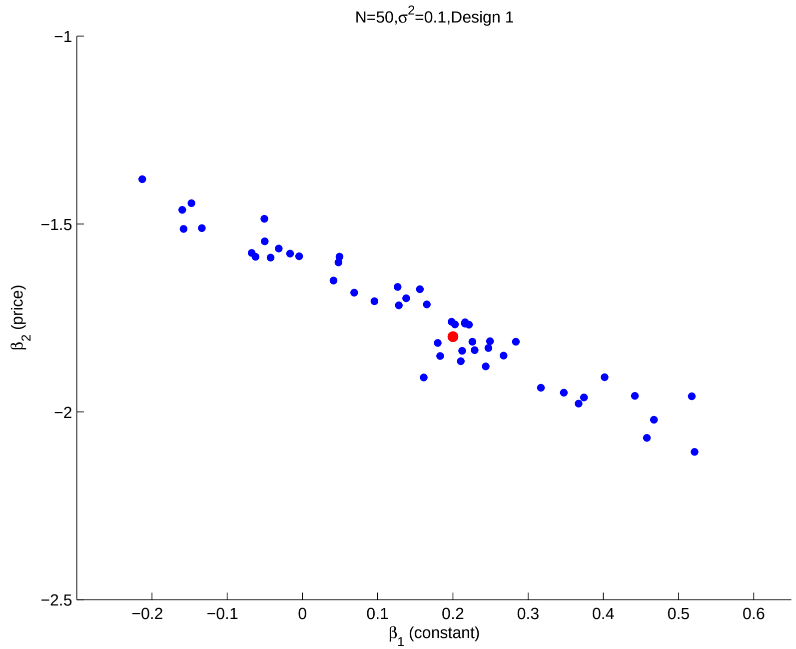
Are the deviation of the estimator from the true value correlated for different coefficients of the OLS estimators?

Covariance Matrix of the OLS Estimator

MATLAB Code: `regestall.m`

Design 1: $x_i \sim -.5 + \mathrm{Uniform}\,[0,1]$ (left hand side) versus Design 2: $x_i \sim 1 + \mathrm{Uniform}\,[0,1]$ $(N = 50,\ \sigma^2 = 0.1)$ (right hand side)





Covariance Matrix of the OLS Estimator

The covariance $\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_k)$ of different coefficients of the OLS estimators measures, if deviations between the estimator and the true value are correlated.

$$\mathrm{Cov}(\hat{\beta}_j, \hat{\beta}_k) = \mathrm{E}\left((\hat{\beta}_j - \beta_j)(\hat{\beta}_k - \beta_k)\right).$$

This information is summarized for all possible pairs of coefficients in the covariance matrix of the OLS estimator. Note that

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \mathrm{E}((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})').$$

Covariance Matrix of the OLS Estimator

The covariance matrix of a random vector is a square matrix, containing in the diagonal the variances of the various elements of the random vector and the covariances in the off-diagonal elements.

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \begin{pmatrix} \mathrm{Var}\left(\hat{\beta}_0\right) & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \cdots & \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_K) \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{Var}\left(\hat{\beta}_1\right) & \cdots & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_K) \\ \vdots & \cdots & \ddots & \vdots \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_K) & \cdots & \mathrm{Cov}(\hat{\beta}_{K-1}, \hat{\beta}_K) & \mathrm{Var}\left(\hat{\beta}_K\right) \end{pmatrix}.$$

Homoskedasticity

To derive $\mathrm{Cov}(\hat{\boldsymbol{\beta}})$, we make an additional assumption, namely homoskedasticity:

$$\mathrm{Var}\left(u|X_1,\ldots,X_K\right)=\sigma^2. \tag{39}$$

This means that the variance of the error term u is the same, regardless of the predictor variables $X_1,\ldots,X_K$.

It follows that

$$\mathrm{Var}\left(Y|X_1,\ldots,X_K\right)=\sigma^2.$$

Error Covariance Matrix

- Because the observations are a random sample from the population, any two observations y_i and y_l are uncorrelated. Hence also the errors u_i and u_l are uncorrelated.

- Together with (39) we obtain the following covariance matrix of the error vector $\boldsymbol{u}$:

$$\mathrm{Cov}(\boldsymbol{u}) = \sigma^2 \mathbf{I},$$

with $\mathbf{I}$ being the identity matrix.

Covariance Matrix of the OLS Estimator

Under assumption (28) and (39), the covariance matrix of the OLS estimator $\hat{\boldsymbol{\beta}}$ is given by:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2(\mathbf{X}^{'}\mathbf{X})^{-1}. \tag{40}$$

Covariance Matrix of the OLS Estimator

Proof. Using (38), we obtain:

$$\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} = \mathbf{A}\boldsymbol{u}, \qquad \mathbf{A} = (\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}.$$

The following holds:

$$\mathrm{E}((\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})') = \mathrm{E}(\mathbf{A}\boldsymbol{u}\boldsymbol{u}'\mathbf{A}') = \mathbf{A}\mathrm{E}(\boldsymbol{u}\boldsymbol{u}')\mathbf{A}' = \mathbf{A}\mathrm{Cov}(\boldsymbol{u})\mathbf{A}'.$$

Therefore:

$$\mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \sigma^2\mathbf{A}\mathbf{A}' = \sigma^2(\mathbf{X}^{'}\mathbf{X})^{-1}\mathbf{X}^{'}\mathbf{X}(\mathbf{X}^{'}\mathbf{X})^{-1} = \sigma^2(\mathbf{X}^{'}\mathbf{X})^{-1}$$

Covariance Matrix of the OLS Estimator

The diagonal elements of the matrix $\sigma^2(\mathbf{X}'\mathbf{X})^{-1}$ define the variance $\mathrm{Var}\left(\hat{\beta}_j\right)$ of the OLS estimator for each component.

The standard deviation $\mathrm{sd}(\hat{\beta}_j)$ of each OLS estimator is defined as:

$$\mathrm{sd}(\hat{\beta}_j) = \sqrt{\mathrm{Var}\left(\hat{\beta}_j\right)} = \sigma\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{j+1,j+1}}. \tag{41}$$

It measures the estimation error on the same unit as β_j.

Evidently, the standard deviation is the larger, the larger the variance of the error. What other factors influence the standard deviation?

Multicollinearity

In practical regression analysis very often high (but not perfect) multicollinearity is present.

How well may X_j be explained by the other regressors?

Consider X_j as left-hand variable in the following regression model, whereas all the remaining predictors remain on the right hand side:

$$X_j = \tilde{\beta}_0 + \tilde{\beta}_1 X_1 + \ldots + \tilde{\beta}_{j-1} X_{j-1} + \tilde{\beta}_{j+1} X_{j+1} + \ldots + \tilde{\beta}_K X_K + \tilde{u}.$$

Use OLS estimation to estimate the parameters and let $\hat{x}_{j,i}$ be the values predicted from this (OLS) regression.

Multicollinearity

- Define R_j as the correlation between the observed values $x_{j,i}$ and the predicted values $\hat{x}_{j,i}$ in this regression.

- If R_j^2 is close to 0, then X_j cannot be predicted from the other regressors. X_j contains additional, "independent" information.

- The closer R_j^2 is to 1, the better X_j is predicted from the other regressors and multicollinearity is present. X_j does not contain much ,,independent" information.

The variance of the OLS estimator

Using R_j, the variance $\mathrm{Var}\left(\hat{\beta}_j\right)$ of the OLS estimators of the coefficient β_j corresponding to X_j may be expressed in the following way for $j = 1, \ldots, K$:

$$\mathrm{Var}\left(\hat{\beta}_j\right) = \frac{\sigma^2}{N s_{x_j}^2 (1 - R_j^2)}.$$

Hence, the variance $\mathrm{Var}\left(\hat{\beta}_j\right)$ of the estimate $\hat{\beta}_j$ is large, if the regressors X_j is highly redundant, given the other regressors (R_j^2 close to 1, multicollinearity).

The variance of the OLS estimator

All other factors same as for the simple regression model, i.e. the variance $\mathrm{Var}\left(\hat{\beta}_j\right)$ of the estimate $\hat{\beta}_j$ is large, if

- the variance σ^2 of the error term u is large;
- the sampling variation in the regressor X_j, i.e. the variance $s^2_{x_j}$, is small;
- the sample size N is small.

II.4 OLS Residuals

Consider the estimated regression model under OLS estimation:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \ldots + \hat{\beta}_K x_{K,i} + \hat{u}_i = \hat{y}_i + \hat{u}_i,$$

where $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \ldots + \hat{\beta}_K x_{K,i}$ is called the fitted value.

$\hat{u}_i$ is called the OLS residual. OLS residuals are useful:

- to estimate the variance σ^2 of the error term;
- to quantify the quality of the fitted regression model;
- for residual diagnostics

EVIEWS Exercise II.4.1

Discuss in EVIEWS how to obtain the OLS residuals and the fitted regression:

- Case Study profit, workfile `profit`;
- Case Study Chicken, workfile `chicken`;
- Case Study Marketing, workfile `marketing`;

OLS residuals as proxies for the error

Compare the underlying regression model

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_K X_K + u, \tag{42}$$

with the estimated model for $i = 1, \ldots, N$:

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_{1,i} + \ldots + \hat{\beta}_K x_{K,i} + \hat{u}_i.$$

- The OLS residuals $\hat{u}_1, \ldots, \hat{u}_N$ may be considered as a "sample" of the unobservable error u.

- Use the OLS residuals $\hat{u}_1, \ldots, \hat{u}_N$ to estimate $\sigma^2 = \mathrm{Var}\,(u)$.

Algebraic properties of the OLS estimator

The OLS residuals $\hat{u}_1, \ldots, \hat{u}_N$ obey $K+1$ linear equations and have the following algebraic properties:

- The sum (average) of the OLS residuals $\hat{u}_i$ is equal to zero:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i = 0. \tag{43}$$

- The sample covariance between $x_{k,i}$ and $\hat{u}_i$ is zero:

$$\frac{1}{N}\sum_{i=1}^{N}x_{k,i}\hat{u}_i = 0, \qquad \forall k = 1, \ldots, K. \tag{44}$$

Estimating σ^2

A naive estimator of σ^2 would be the sample variance of the OLS residuals $\hat{u}_1, \ldots, \hat{u}_N$:

$$\tilde{\hat{\sigma}}^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{u}_i^2 - \frac{1}{N} \sum_{i=1}^{N} \hat{u}_i \right)^2 = \frac{1}{N} \sum_{i=1}^{N} \hat{u}_i^2 = \frac{\mathrm{SSR}}{N},$$

where we used (43) and $\mathrm{SSR} = \sum_{i=1}^{N} \hat{u}_i^2$ is the sum of squared OLS residuals.

However, due to the linear dependence between the OLS residuals, $\hat{u}_1, \ldots, \hat{u}_N$ is not an independent sample. Hence, $\tilde{\hat{\sigma}}^2$ is a biased estimator of σ^2.

Estimating σ^2

Due to the linear dependence between the OLS residuals, only $\mathrm{df} = (N - K - 1)$ residuals can be chosen independently.

df is also called the degrees of freedom.

An unbiased estimator of the error variance σ^2 in a homoscedastic multiple regression model is given by:

$$\hat{\sigma}^2 = \frac{\mathrm{SSR}}{\mathrm{df}}, \tag{45}$$

where $\mathrm{df} = (N - K - 1)$, N is the number of observations, and K is the number of predictors $X_1, \ldots, X_K$

.

The standard errors of the OLS estimator

The standard deviation $\mathrm{sd}(\hat{\beta}_j)$ of the OLS estimator given in (46) depends on $\sigma = \sqrt{\sigma^2}$.

To evaluate the estimation error for a given data set in practical regression analysis, σ^2 is substituted by the estimator (45). This yields the so-called standard error $\mathrm{se}(\hat{\beta}_j)$ of the OLS estimator:

$$\mathrm{se}(\hat{\beta}_j) = \sqrt{\hat{\sigma}^2}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{j+1,j+1}}. \tag{46}$$

EVIEWS Exercise II.4.1

EViews (and other packages) report for each predictor the OLS estimator together with the standard errors:

- Case Study profit, work file `profit`;
- Case Study Chicken, work file `chicken`;
- Case Study Marketing, work file `marketing`;

Note: the standard errors computed by EViews (and other packages) are valid only under the assumption made above, in particular, homoscedasticity.

Quantifying the model fit

How well does the multiple regression model (42) explain the variation in Y? Compare it with the following simple model without any predictors:

$$Y = \beta_0 + \tilde{u}. \qquad (47)$$

The OLS estimator of β_0 minimizes the following sum of squared residuals:

$$\sum_{i=1}^{N} (y_i - \beta_0)^2$$

and is given by $\hat{\beta}_0 = \overline{y}$.

Coefficient of Determination

The minimal sum is equal to the total variation

$$\mathrm{SST} = \sum_{i=1}^{N} (y_i - \overline{y})^2.$$

Is it possible to reduce the sum of squared residuals SST of the simple model (47) by including the predictor variables $X_1, \ldots, X_N$ as in (42)?

Coefficient of Determination

The minimal sum of squared residuals SSR of the multiple regression model (42) is always smaller than the minimal sum of squared residuals SST of the simple model (47):

$$\mathrm{SSR} \leq \mathrm{SST}. \tag{48}$$

The coefficient of determination R^2 of the multiple regression model (42) is defined as:

$$\mathrm{R}^2 = \frac{\mathrm{SST} - \mathrm{SSR}}{\mathrm{SST}} = 1 - \frac{\mathrm{SSR}}{\mathrm{SST}}. \tag{49}$$

Coefficient of Determination

Proof of (48). The following variance decomposition holds:

$$\mathrm{SST} = \sum_{i=1}^{N} (y_i - \hat{y}_i + \hat{y}_i - \overline{y})^2 = \sum_{i=1}^{N} \hat{u}_i^2 + 2 \sum_{i=1}^{N} \hat{u}_i (\hat{y}_i - \overline{y}) + \sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2.$$

Using the algebraic properties (43) and (44) of the OLS residuals, we obtain:

$$\sum_{i=1}^{N} \hat{u}_i (\hat{y}_i - \overline{y}) = \hat{\beta}_0 \sum_{i=1}^{N} \hat{u}_i + \hat{\beta}_1 \sum_{i=1}^{N} \hat{u}_i x_{1,i} + \ldots + \hat{\beta}_K \sum_{i=1}^{N} \hat{u}_i x_{K,i} - \overline{y} \sum_{i=1}^{N} \hat{u}_i = 0.$$

Therefore:

$$\mathrm{SST} = \mathrm{SSR} + \sum_{i=1}^{N} (\hat{y}_i - \overline{y})^2 \geq \mathrm{SSR}.$$

Coefficient of Determination

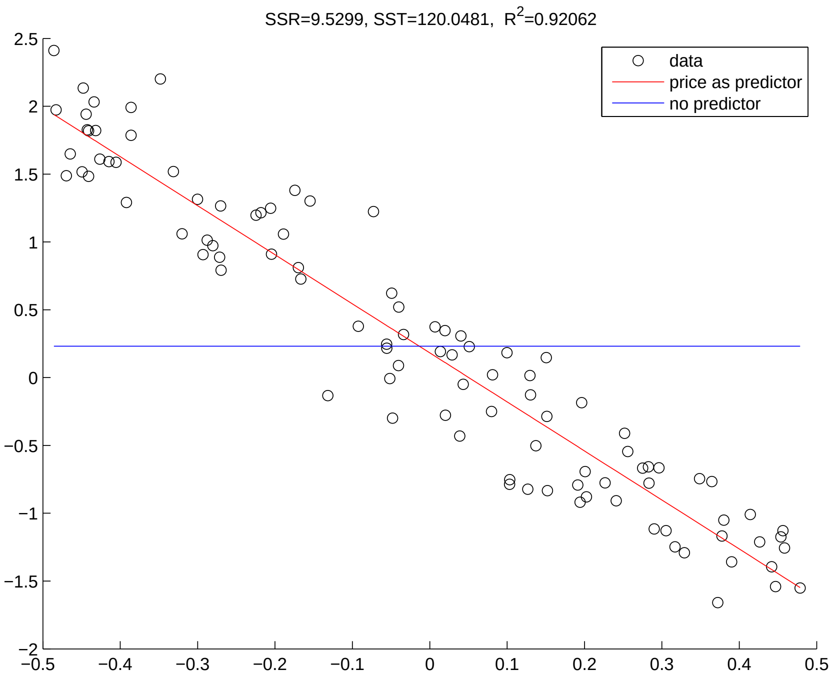
The coefficient of determination R^2 is a measure of goodness-of-fit:

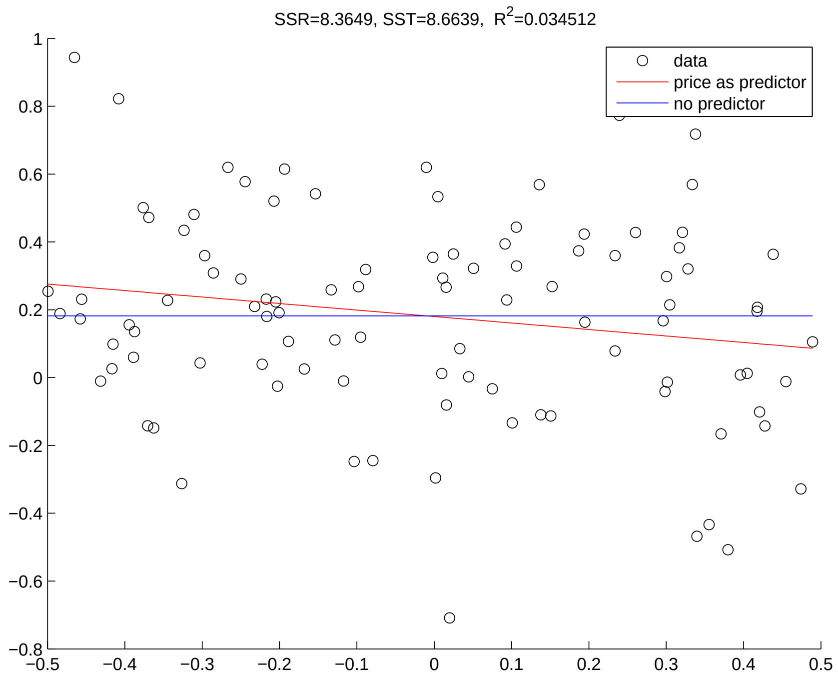
- If $\mathrm{SSR} \approx \mathrm{SST}$, then there is little gained by including the predictors. R^2 is close to 0. The multiple regression model explains the variation in Y hardly better than the simple model (47).

- If $\mathrm{SSR} << \mathrm{SST}$, then much is gained by including all predictors. R^2 is close to 1. The multiple regression model explains the variation in Y much better than the simple model (47).

Programm packages like EViews report SSR and R^2.

Coefficient of Determination





MATLAB Code: `reg-est-r2.m`

The Gauss Markov Theorem

The Gauss Markov Theorem. Under the assumptions (28) and (39), the OLS estimator is BLUE, i.e. the

- **B**est
- **L**inear
- **U**nbiased
- **E**stimator

Here "best" means that any other linear unbiased estimator has larger standard errors than the OLS estimator.

II.5 Testing Hypothesis

Multiple regression model:

$$Y = \beta_0 + \beta_1 X_1 + \ldots + \beta_j X_j + \ldots + \beta_K X_K + u, \tag{50}$$

Does the predictor variable X_j exercise an influence on the expected mean $\mathrm{E}(Y)$ of the response variable Y, if we control for all other variables $X_1, \ldots, X_{j-1}, X_{j+1}, \ldots, X_K$? Formally,

$$\boxed{\beta_j = 0?}$$

Understanding the testing problem

- Simulate data from a multiple regression model with $\beta_0 = 0.2$, $\beta_1 = -1.8$, and $\beta_2 = 0$:

$$Y = 0.2 - 1.8X_1 + 0 \cdot X_2 + u, \quad u \sim \text{Normal}\left(0, \sigma^2\right).$$

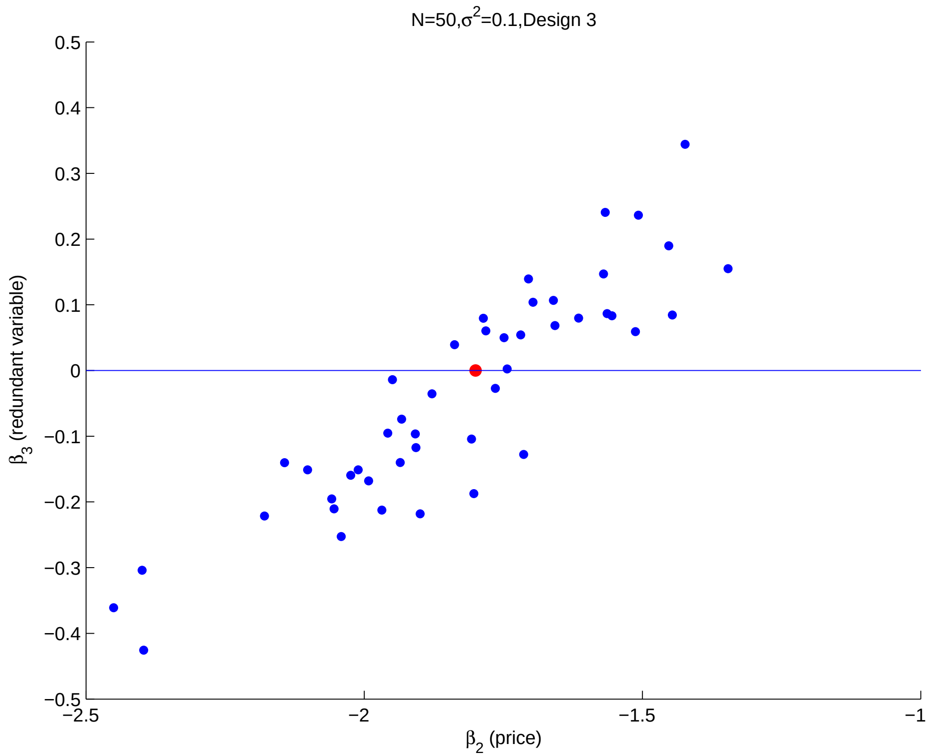
- Run OLS estimation for a model where β_2 is unknown:

$$Y = \beta_0 + \beta_1 X_2 + \beta_2 X_3 + u, \quad u \sim \text{Normal}\left(0, \sigma^2\right),$$

to obtain $(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)$. Is $\hat{\beta}_2$ different from 0?

MATLAB Code: `regtest.m`

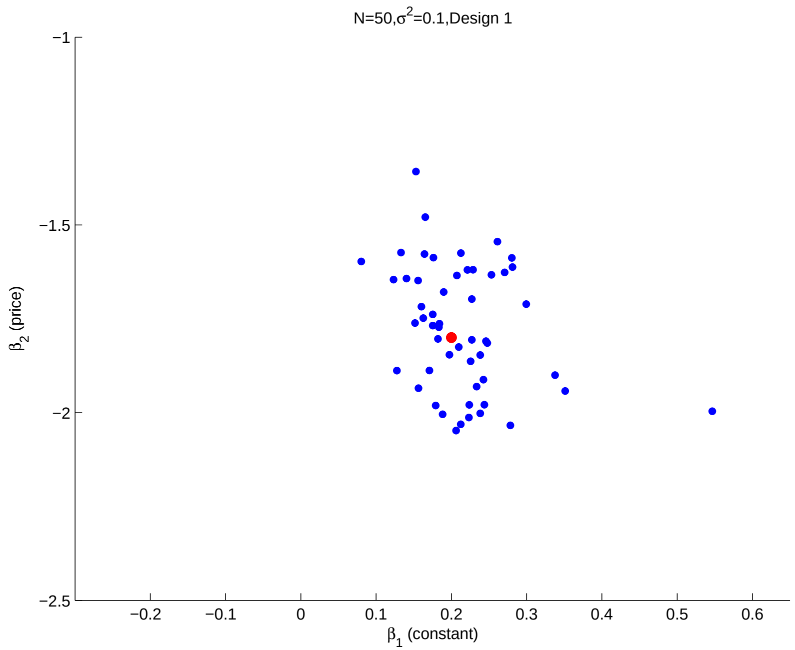
Understanding the testing problem

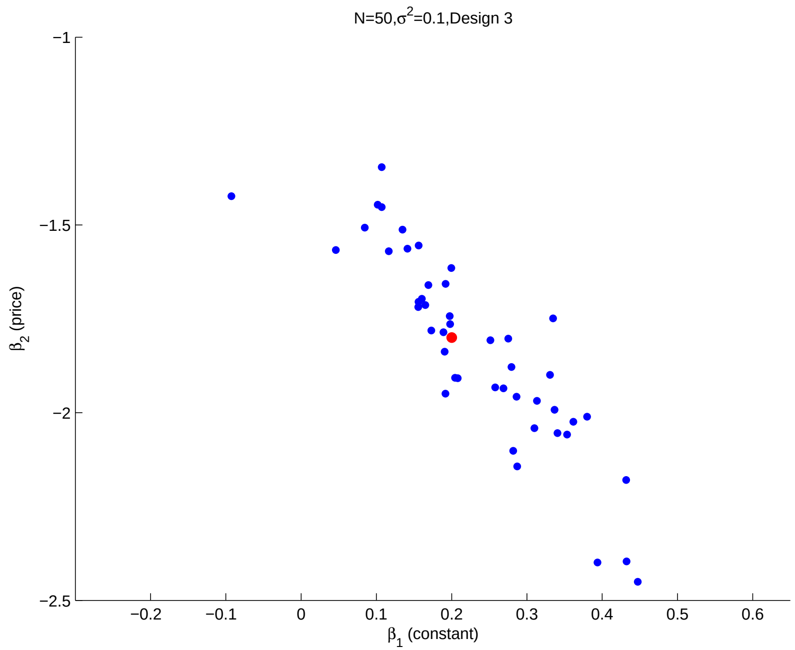


The OLS estimator $\hat{\beta}_2$ of $\beta_2 = 0$ differs from 0 for a single data set, but is 0 on average.

Understanding the testing problem

OLS estimation for the true model in comparison to estimating a model with a redundant predictor variable: including the redundant predictor X_2 increases the estimation error for the other parameters β_0 and β_1.





Testing of hypotheses

- What may we learn from the data about hypothesis concerning the unknown parameters in the regression model, especially about the hypothesis that $\beta_j = 0$?

- May we reject the hypothesis $\beta_j = 0$ given data?

- Testing, if $\beta_j = 0$ is not only of importance for the substantive scientist, but also from an econometric point of view, to increase efficiency of estimation of non-zero parameters.

It is possible to answer these questions, if we make additional assumptions about the error term u in a multiple regression model.