2.3 Estimation properties of OLS method

Any estimation method has certain **quality requirements** which depend on certain conditions being met. As regards the **ordinary least squares method (OLS method)**, certain **estimation properties** are known, provided that the standard assumptions of the multiple regression model are met. If the disturbance variable on average is equal to 0 and its variance is constant, then the **OLS estimators** for the regression coefficients are **best linear unbiased estimators (BLUE)** in the absence of autocorrelation, with appeal to the Gauss-Markov theorem.

Unbiasedness implies that the mean values of the OLS-estimated regression coefficients are conform with the (unknown) population regression coefficients. Best estimator refers to the accuracy of the estimation and means efficiency. Under the standard assumptions, the OLS estimator in the linear regression model is thus **unbiased** and **efficient.** No other linear and unbiased estimator of the regression coefficients exists which leads to a smaller variance.

In addition, the OLS method is **consistent.** Consistency means that with increasing sample size the estimated regression coefficients better and better approach the unknown population regression coefficients.

- **Unbiasedness**

An estimator is unbiased if its expected value matches the parameter of the population.

To show the unbiasedness of the OLS estimator $\hat{\boldsymbol{\beta}}$ we first insert the regression model (2.5) in the estimator of $\hat{\boldsymbol{\beta}}$ (2.15):

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

After multiplication one obtains

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

from which

(2:33) $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$

follows.

Now the expected value of $\hat{\boldsymbol{\beta}}$ can be determined:

$$\begin{aligned}\mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left|\boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u})\right| = \mathrm{E}(\boldsymbol{\beta}) + \mathrm{E}\left|(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right| \\ &= \boldsymbol{\beta} + E\left[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}\right] = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\,\mathbf{X}'\mathbf{E}(\mathbf{u})\end{aligned}$$

With E ($\mathbf{u}$) = $\mathbf{0}$ one finally obtains

(2.34) $$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$

This proves the unbiasedness of the OLS estimator $\hat{\boldsymbol{\beta}}$. Making use of the extensive notation of (2.34)

$$\begin{pmatrix} E(\hat{\beta}_1) \\ E(\hat{\beta}_2) \\ \vdots \\ E(\hat{\beta}_k) \end{pmatrix} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix},$$

it can be seen that the mean values of the OLS estimators $\hat{\beta}_1, \hat{\beta}_2, ..., \hat{\beta}_k$ conform with the unknown regression coefficients of the econometric model.

• Efficiency

An unbiased estimator, which has the smallest variance compared to all alternative unbiased estimators, is called efficient. We call such an estimator a best unbiased estimator. The OLS estimator $\hat{\boldsymbol{\beta}}$ is efficient in the class of linear estimators.

Let $\hat{\beta}_j$ denote the OLS estimator of the unknown regression coefficient β_j and $\tilde{\beta}_j$ any other linear estimator. Efficiency means then that the variance of the OLS estimator $\hat{\beta}_j$ is the smallest:

(2.35) $\mathrm{Var}(\hat{\beta}_j) \leq \mathrm{Var}(\tilde{\beta}_j)$, $j = 1,2, ..., k$.

• Consistency

When an estimator is required to increasingly approach the values of a parameter vector with increasing sample sizes, then is in particular its limiting behavior for n -> ∞ of interest, i.e. a property relating to large samples. In looking for consistency we are concerned with the question as to whether an estimator will finally coincide with the parameter vector $\boldsymbol{\beta}$ if sample size increases beyond all bounds:

$$\lim_{n\to\infty} \mathrm{E}(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta} \quad \text{and} \quad \lim_{n\to\infty} \mathrm{Cov}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$$

2.4 Coefficient of determination and multiple correlation coefficient

- **Coefficient of determination**

- Evaluates global goodness of fit of the regression model

- Quantification of the explanatory power of the exogenous variables in an econometric model

Variance of the dependent variable Y:

(2.36) $$s_y^2 = \frac{1}{n}\sum(y_t - \bar{y})^2 = \frac{1}{n}\sum y_t^2 - \bar{y}^2 = \frac{1}{n}\mathbf{y}'\mathbf{y} - \bar{y}^2$$

Variance decomposition:

(2:37) $$s_y^2 = s_{\hat{y}}^2 + s_{\hat{u}}^2$$

with

(2:38) $$s_{\hat{y}}^2 = \frac{1}{n}\sum(\hat{y}_t - \bar{y})^2 = \frac{1}{n}\sum \hat{y}_t^2 - \bar{y}^2 = \frac{1}{n}\hat{\mathbf{y}}'\hat{\mathbf{y}} - \bar{y}^2$$ *because of* $\bar{\hat{y}} = \bar{y}$

(2:39) $$s_{\hat{u}}^2 = \frac{1}{n}\sum \hat{u}_t^2 = \frac{1}{n}\hat{\mathbf{u}}'\hat{\mathbf{u}}$$ *because of* $\bar{\hat{u}} = 0$

Coefficient of determination:

(2.40a) $$R^2 = \frac{s_{\hat{y}}^2}{s_y^2}$$ or (2.40b) $$R^2 = 1 - \frac{s_{\hat{u}}^2}{s_y^2}$$

Value range: $0 \le R^2 \le 1$

Formulas:

(2.41a) $$R^2 = \frac{\hat{\mathbf{y}}'\hat{\mathbf{y}} - n \cdot \overline{y}^2}{\mathbf{y}'\mathbf{y} - n \cdot \overline{y}^2} = \frac{n\hat{\mathbf{y}}'\hat{\mathbf{y}} - \left(\sum y_t\right)^2}{n\mathbf{y}'\mathbf{y} - \left(\sum y_t\right)^2}$$

or

(2.41b) $$R^2 = \frac{\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - n \cdot \overline{y}^2}{\mathbf{y}'\mathbf{y} - n \cdot \overline{y}^2} = \frac{n\hat{\boldsymbol{\beta}}'\mathbf{X}'\mathbf{y} - \left(\sum y_t\right)^2}{n\mathbf{y}'\mathbf{y} - \left(\sum y_t\right)^2}$$

Example:

Wanted is the proportion of the variance of private consumption attributable to the dispersion of disposable income. In order to be able to make a statement on this question, the coefficient of determination for the macroeconomic consumption function applied to the period 1994 by 2012 will be calculated using formula (2.41b)
Using

$$\hat{\beta}' = \begin{bmatrix} -38,5222 & 0,93354 \end{bmatrix},$$

$$X'y = \begin{bmatrix} \sum y_t \\ \sum x_t y_t \end{bmatrix} = \begin{bmatrix} 24034,68 \\ 33960226,83 \end{bmatrix} \quad y'y = \sum y_t^2 = 30778325$$

we calculate

$$\hat{y}'\hat{y} = \hat{\beta}' X'y = 30777206$$

With n = 19 and $\bar{y} = 1264{,}98$ the coefficient of determination is then given by:

$$R^2 = \frac{19 \cdot 30777206 - 24034,68^2}{19 \cdot 30778325 - 24034,68^2} = \frac{583166740}{583187986} = 0,9968$$

- **Simple and multiple correlation coefficient**

- Simple Regression:

(2.42) $$R^2 = r_{yx}^2$$

Because of the linear transformation $\hat{y}_t = \hat{\beta}_1 + \hat{\beta}_2 \cdot x_t$ one gets

$$s_{\hat{y}}^2 = \hat{\beta}_2^2 \cdot s_x^2$$

following which (2.42) can be easily demonstrated:

$$R^2 = \frac{s_{\hat{y}}^2}{s_y^2} = \frac{\hat{\beta}_2^2 \cdot s_x^2}{s_y^2} = \frac{\left(\frac{s_{xy}}{s_x^2}\right)^2 \cdot s_x^2}{s_y^2} = \frac{s_{xy}^2}{s_x^2 \cdot s_y^2} = r_{xy}^2$$

In the simple regression model, the square root of the coefficient of determination therefore depicts the absolute value of the coefficient of correlation between the regressand y and the regressor x:

(2.43) $$|r_{yx}| = \sqrt{R^2}$$

- Multiple Regression:

The coefficient of determination is then the square root of the multiple correlation coefficient $r_{y\hat{y}}$, indicating the correlation between the dependent variable y and all independent variables $x_1, x_2, ..., x_k$ (=correlation between the dependent variable y_t and the regression values $\hat{y}_t$).

(2:44) $$R^2 = r^2_{y\hat{y}} = r^2_{yx_{1,2,...k}}$$

The sign of the multiple correlation coefficient $r_{yx_{1,2,...k}}$ remains undefined, since positive and negative effects of the regressors can occur simultaneously:

(2.45) $$\left|r_{y\hat{y}}\right| = \left|r_{yx_{1,2,...k}}\right| = \sqrt{R^2}$$

- **Adjusted coefficient of determination**

Although the coefficient of determination can be easily interpreted, it is of only limited use when comparing different regression models with different numbers of independent variables. If additional regresors are included in a regression model, R^2 increases even when it has no economic relevance.

One can avoid this by taking account of the number of degrees of freedom in the calculation of the coefficient of determination. In contrast to R^2 the **adjusted coefficient of determination**

(2.46) $$\bar{R}^2 = 1 - \frac{\frac{1}{n-k}\hat{\mathbf{u}}'\hat{\mathbf{u}}}{\frac{1}{n-1}\left(\mathbf{y}'\mathbf{y} - n\bar{y}^2\right)}$$

can decline in the case of the inclusion of irrelevant regressors.

A calculation of the adjusted coefficient of determination may be done via R^2:

(2:47) $$\bar{R}^2 = 1 - \frac{n-1}{n-k}\left(1 - R^2\right)$$