Linear combinations of parameters

Suppose we want to test the hypothesis that two regression coefficients are equal, e.g. $\beta_1 = \beta_2$. This is equivalent to testing the following linear constraint (null hypothesis):

$$\beta_1 - \beta_2 = 0. \tag{58}$$

Test statistic based on the difference of the OLS estimators $\hat{\beta}_1 - \hat{\beta}_2$:

- If $\hat{\beta}_1 - \hat{\beta}_2$ is small, then the hypothesis (58) is not rejected.
- If $|\hat{\beta}_1 - \hat{\beta}_2|$ is large, then the hypothesis (58) is rejected.

What is the distribution of $\hat{\beta}_1 - \hat{\beta}_2$ under the null hypothesis?

Testing linear combinations of parameters

The distribution of $\hat{\beta}_1 - \hat{\beta}_2$ under the null hypothesis (58) is equal to following univariate normal distribution:

$$\hat{\beta}_1 - \hat{\beta}_2 \sim \text{Normal}\left(0, \text{Var}\left(\hat{\beta}_1 - \hat{\beta}_2\right)\right),$$

$$\text{Var}\left(\hat{\beta}_1 - \hat{\beta}_2\right) = \text{Var}\left(\hat{\beta}_1\right) + \text{Var}\left(\hat{\beta}_2\right) - 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \quad (59)$$

The test statistic

$$t = \frac{\hat{\beta}_1 - \hat{\beta}_2}{\text{se}(\hat{\beta}_1 - \hat{\beta}_2)} \sim t_{\text{df}}, \quad (60)$$

where σ^2 is substituted by $\hat{\sigma}^2$, follows the t_{df} distribution with $\text{df} = (N - K - 1)$.

Testing linear combinations of parameters

Testing the linear constraint $\beta_1 - \beta_2 = 0$ for $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_K)'$ is equivalent with testing:

$$\mathbf{L}\boldsymbol{\beta} = 0, \quad \mathbf{L} = \begin{pmatrix} 0 & 1 & 1 & 0 & \cdots & 0 \end{pmatrix}. \tag{61}$$

What is the distribution of $\mathbf{L}\hat{\boldsymbol{\beta}}$?

Using $\hat{\boldsymbol{\beta}} - \boldsymbol{\beta} \sim \mathrm{Normal}_{K+1}\left(\mathbf{0}, \mathrm{Cov}(\hat{\boldsymbol{\beta}})\right)$, we obtain the following:

$$\mathbf{L}\hat{\boldsymbol{\beta}} - \mathbf{L}\boldsymbol{\beta} \sim \mathrm{Normal}_q\left(\mathbf{0}, \mathrm{Cov}(\tilde{\hat{\boldsymbol{\beta}}})\right),$$

$$\mathrm{Cov}(\tilde{\hat{\boldsymbol{\beta}}}) = \mathbf{L}\mathrm{Cov}(\hat{\boldsymbol{\beta}})\mathbf{L}'.$$

Example

E.g. for (61) with $K = 3$ regressors - compute $\mathbf{L}\text{Cov}(\hat{\boldsymbol{\beta}})\mathbf{L}'$:

$$\begin{pmatrix} 0 & 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \text{Var}\left(\hat{\beta}_0\right) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_2) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_3) \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Var}\left(\hat{\beta}_1\right) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_3) \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_2) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) & \text{Var}\left(\hat{\beta}_2\right) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) \\ \text{Cov}(\hat{\beta}_0, \hat{\beta}_3) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_3) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_3) & \text{Var}\left(\hat{\beta}_3\right) \end{pmatrix} \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}$$

$$= \text{Var}\left(\hat{\beta}_1\right) + \text{Var}\left(\hat{\beta}_2\right) - 2\text{Cov}(\hat{\beta}_1, \hat{\beta}_2).$$

This yields (59).

Testing linear constraints

The F-statistic may also be used to test more than one linear constraint on the coefficients, i.e. $\mathbf{L}\boldsymbol{\beta} = 0$, where $\mathbf{L}$ is $q \times (K+1)$-matrix, with $q > 1$.

According to (62), the OLS estimator $\tilde{\hat{\boldsymbol{\beta}}} = \mathbf{L}\hat{\boldsymbol{\beta}}$ follows the multivariate normal distribution $\mathrm{Normal}_q\left(\mathbf{0}, \mathrm{Cov}(\tilde{\hat{\boldsymbol{\beta}}})\right)$ under the null hypothesis.

The F-statistic is constructed as above and follows an $F_{q,\mathrm{df}}$-distribution, where q is the number of linear constraints.

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 a larger sum of squared residuals (SSR) than the OLS estimator.

Efficiency of OLS estimation

Under assumption (51) about the error term, the OLS estimator is not only BLUE. A stronger optimality result holds.

> Efficiency of OLS estimation. Under assumption (51), the OLS estimator $\hat{\boldsymbol{\beta}}$ is the minimum variance unbiased estimator.

Any other unbiased estimator $\tilde{\boldsymbol{\beta}}$ (which need not be a linear estimator) has larger standard deviations than the OLS estimator:

- $\mathrm{sd}(\tilde{\beta}_j) \geq \mathrm{sd}(\hat{\beta}_j)$;
- $\mathrm{Cov}(\tilde{\boldsymbol{\beta}}) - \mathrm{Cov}(\hat{\boldsymbol{\beta}})$ is positive semi-definite.

Consistency of OLS estimation

Let $\hat{\beta}_N$ be an estimator for β, based on sample size N.

Then, $\hat{\beta}_N$ is a consistent estimator for β, if for every $\epsilon > 0$, the following holds:

$$\Pr\{|\hat{\beta}_N - \beta| \geq \epsilon\} \to 0 \qquad \text{as } N \to \infty,$$

or, equivalently,

$$\Pr\{|\hat{\beta}_N - \beta| < \epsilon\} \to 1 \qquad \text{as } N \to \infty.$$

Note that ϵ may be arbitrarily small.

Consistency of OLS estimation

- Consistency means that the OLS estimator converges ,,in probability" to the true value with increasing number of observations N.

- A sufficient condition for this convergence in probability is that $\mathrm{E}(\hat{\beta}_N) \rightarrow \beta$ and $\mathrm{sd}(\hat{\beta}_N) \rightarrow 0$ as $N \rightarrow \infty$.

- Under the Gauss Markov assumptions, the OLS estimator is a consistent estimator of $\boldsymbol{\beta}$.

- Note that consistency also holds, if the normality assumption (51) is violated.

Consistency of OLS estimation

"Proof". For each $j = 1, \ldots, K$:

- The OLS estimator is unbiased, i.e. $\mathrm{E}(\hat{\beta}_j) = \beta_j$
- The standard deviation $\mathrm{sd}(\hat{\beta}_j)$ goes to 0 for $N \rightarrow \infty$:

$$\mathrm{sd}(\hat{\beta}_j) \;=\; \frac{\sigma}{\sqrt{N s_{x_j}^2 (1 - R_j^2)}} \rightarrow 0, \qquad \text{as } N \rightarrow \infty$$

II.7 Residual diagnostics

Hypothetical model:

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

Estimated model:

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

where $\hat{u}_i$ is the OLS residual.

- Due to consistency, the OLS residual $\hat{u}_i$ approaches the unobservable error u_i, as N increases.
- Use OLS residuals $\hat{u}_i$ to test assumptions about u_i.

EVIEWS Exercise II.6.1

Discuss in EVIEWS how to obtain the OLS residuals

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

Testing Normality

The error follows a normal distribution:

$$u_i \sim \text{Normal}\left(0, \sigma^2\right).$$

- Roughly 95% of the OLS residuals lie between $[-2 \cdot \hat{\sigma}^2, 2 \cdot \hat{\sigma}^2]$;
- Assumption often violated, if outliers are present.
- Normality often improved through transformations.

To test normality of the true errors u_i, check normality of the OLS residuals $\hat{u}_i$:

Testing Normality

- Histogram

- Skewness close to 0?

$$m_3 = \frac{1}{\hat{\sigma}^3} \left(\frac{1}{N} \sum_{i=1}^{N} \hat{u}_i^3 \right) \tag{62}$$

- Kurtosis close to 3?

$$m_4 = \frac{1}{\hat{\sigma}^4} \left(\frac{1}{N} \sum_{i=1}^{N} \hat{u}_i^4 \right) \tag{63}$$

$\hat{\sigma}^2$ is the estimated variance of the OLS residuals.

Testing Normality

Jarque-Bera-Statistics:

$$J = \frac{N-K}{6}\left(m_3^2 + \frac{1}{4}(m_4 - 3)^2\right). \qquad (64)$$

- Null hypothesis: the errors follow a normal distribution
- Under the null hypothesis, J follows asymptotically (i.e. for large N) a χ^2-distribution with 2 degrees of freedom.
- Reject the null hypothesis, if the p-value of J is small.

Case Study Profit

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i, \tag{65}$$

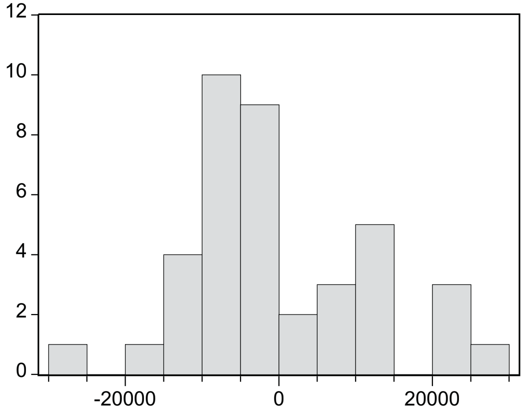
where

$$\begin{aligned} y_i & \quad \ldots \quad \text{profit 1994} \\ x_{1,i} & \quad \ldots \quad \text{profit 1993} \\ x_{2,i} & \quad \ldots \quad \text{turnover 1994} \end{aligned}$$

Consider only large firms ($i = 1, \ldots, 39$)

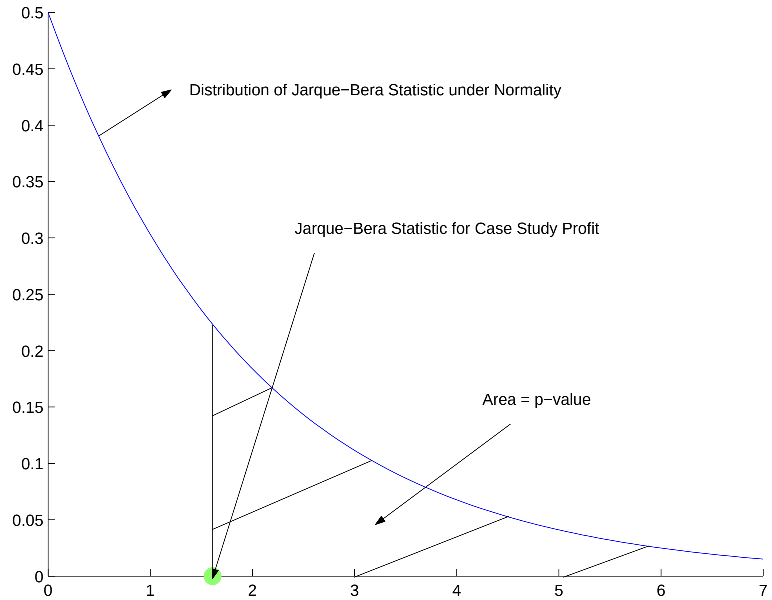
Demonstration in EViews, work file `profit`

Case Study Profit



Series: Residuals
Sample 1 39
Observations 39

Mean	-5.36E-13
Median	-2081.893
Maximum	29116.01
Minimum	-26654.90
Std. Dev.	11933.24
Skewness	0.497860
Kurtosis	3.010465
Jarque-Bera	1.611298
Probability	0.446798



Case Study Yield

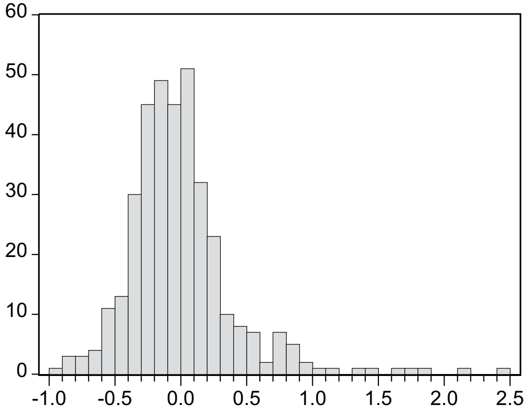
$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i, \qquad (66)$$

where

$$\begin{aligned} y_i & \quad \ldots \quad \text{yield with maturity 3 months} \\ x_{2,i} & \quad \ldots \quad \text{yield with maturity 1 month} \\ x_{3,i} & \quad \ldots \quad \text{yield with maturity 60 months} \end{aligned}$$

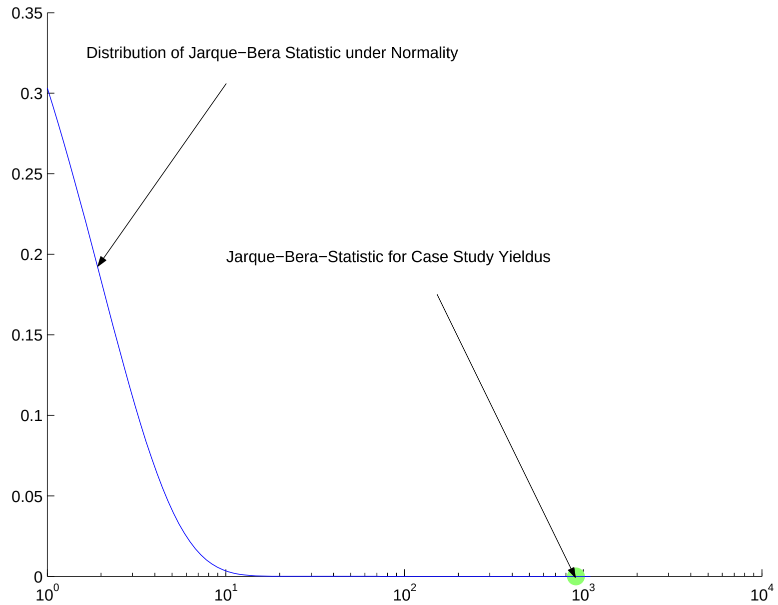
Demonstration in EVIEWS, workfile `yield`

Case Study Yield



Series: Residuals
Sample 1964:01 1993:12
Observations 360

Mean	-4.93E-18
Median	-0.047621
Maximum	2.454587
Minimum	-0.909939
Std. Dev.	0.419164
Skewness	1.872579
Kurtosis	9.829840
Jarque-Bera	910.0937
Probability	0.000000



Checking assumption (39)

The variance of u_i is homoscedastic, i.e.

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

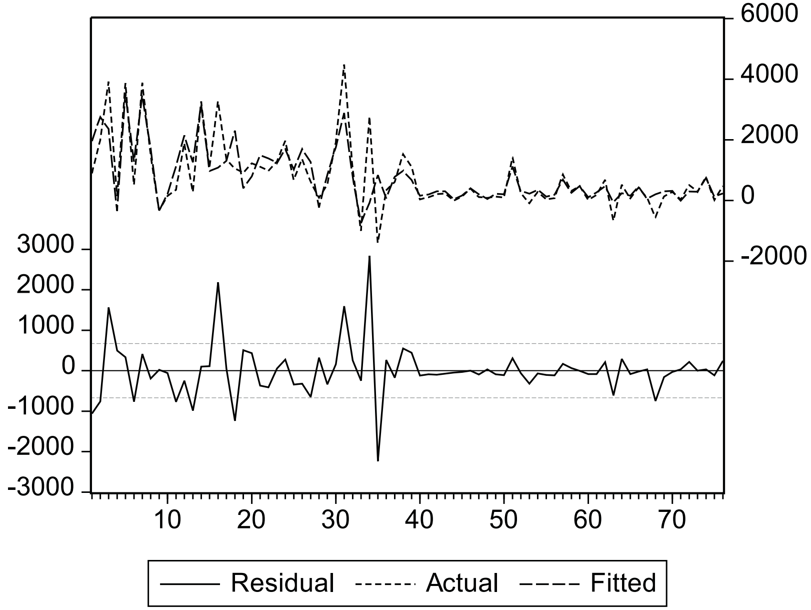
- If this assumption is violated, the model is said to have heteroscedastic errors.

- This assumption is often violated if the variance of $\mathrm{Var}\left(u\right)$ depends on a predictor variable.

First informal check: residual plot; more about formal tests later

Case Study Profit

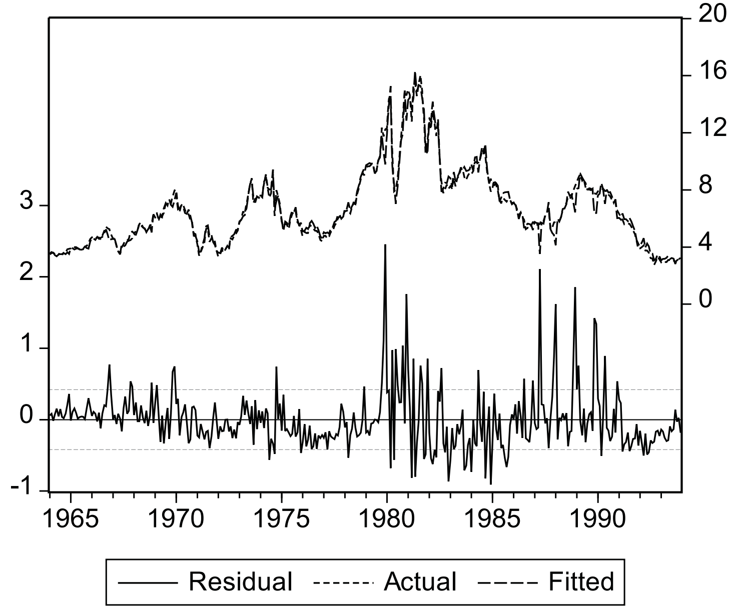
Demonstration in EViews, work file `profit` (full sample)



$\hat{u}_i^2$ depends on the size of the firm (heterogeneity)

Case Study Yield

Demonstration in EViews, work file `yieldus`



$\hat{u}_i^2$ exhibits volatility clusters

Checking assumption (28)

The model does not contain any systematic error, i.e.

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

- If assumption A1 is violated, the model is said to have a specification error:
  - the true value of y_i will be underrated, if $\mathrm{E}(u_i|\cdot)>0$;
  - the true value of y_i will be overrated, if $\mathrm{E}(u_i|\cdot)<0$.
- This assumption is often violated, when an important predictor variable has been omitted.

Checking assumption (28)

Demonstration: $\Rightarrow$

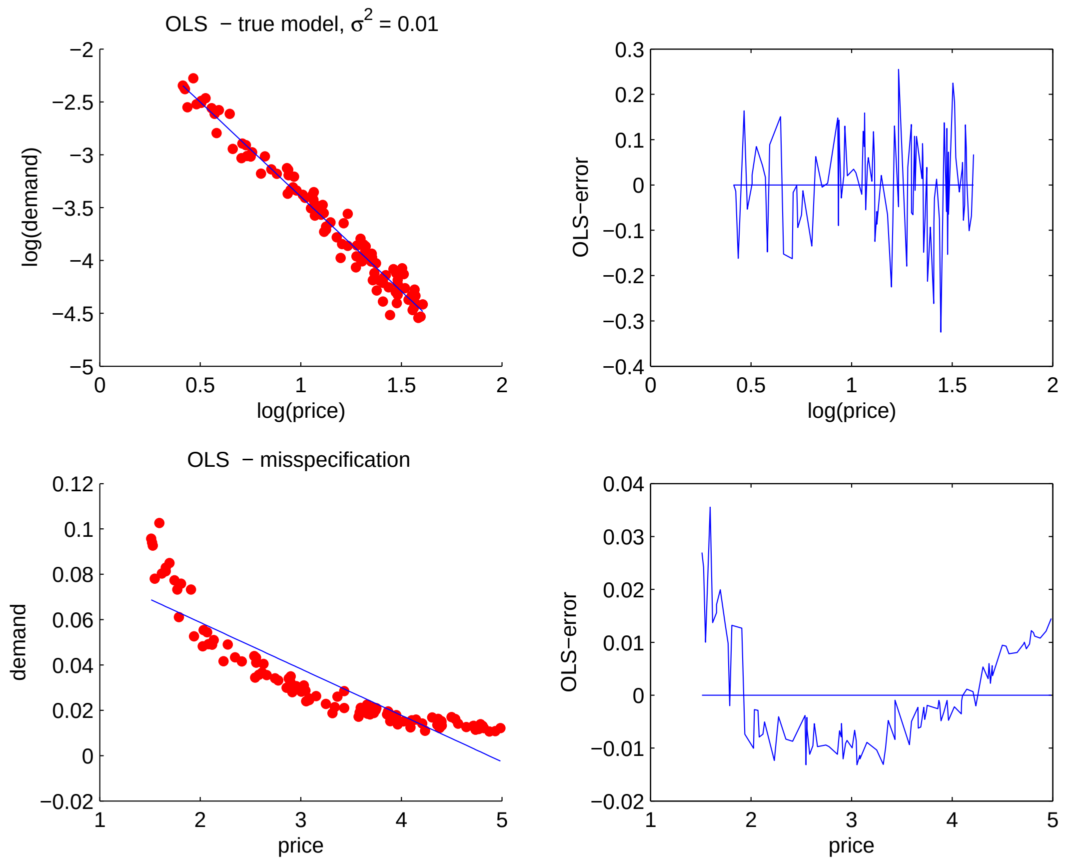
MATLAB Code: `regresa1.m`

Simulate data from a simple log-linear regression model with $\tilde{\beta}_1 = 0.2$ and $\beta_2 = -1.8$:

$$y_i = 0.2 \cdot x_i^{-1.8} e^{u_i}, \tag{67}$$

- Residual plot for the log-linear regression model
- Residual plot for the linear regression model: specification error

Checking assumption (28)



Case Study Profit

Demonstration in EViews, work file `profit`

Model profit 1994 only as a function of profit 1993