Erasmus
School of
Economics

Econometrics I

Lecture 4: Diagnostic tests I

Annika Camehl

November - December, 2021

Today

1. Heteroskedasticity

2. Serial correlation

3. Generalized Least Squares

Heteroskedasticity

Heteroskedasticity (A3)

We start with the linear model

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where the assumptions 1, 2, 4, 5, and 6 hold.

A3: homoskedasticity ($E(\varepsilon_i^2) = \sigma^2$) does not hold!

Instead we have heteroskedasticity:

$$E(\varepsilon_i^2) = \sigma_i^2.$$

$\Rightarrow$ Different observations have different amount of randomness.

Matrix notation

In matrix notation we have:

$$y = X\beta + \varepsilon,$$

with

$$E(\varepsilon\varepsilon') = \Omega = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}.$$

Example 4.1: Heteroskedasticity

Consequences of heteroskedasticity I

Heteroskedasticity has some consequences for estimating β.

- It holds that the OLS estimator

$$b_{OLS} = (X'X)^{-1}X'y$$

  is still unbiased (and consistent)

$$E[b_{OLS}] = \beta + E[(X'X)^{-1}X'\varepsilon] = \beta + (X'X)^{-1}X'E[\varepsilon] = \beta.$$

Consequences of heteroskedasticity II

- OLS is not efficient
- Usual OLS standard errors are not correct. The covariance matrix of b_{OLS}, $\text{Var}[b_{OLS}]$ equals $E[(b_{OLS} - \beta)(b_{OLS} - \beta)']$, such that

$$\begin{aligned}
\text{Var}[b_{OLS}] &= E[(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}] \\
&= (X'X)^{-1}X'E[\varepsilon\varepsilon']X(X'X)^{-1} \\
&= (X'X)^{-1}X'\Omega X(X'X)^{-1}
\end{aligned}$$

  with $E[\varepsilon\varepsilon'] = \Omega$.
- Note if $\Omega = \sigma^2 I$ $\text{Var}[b_{OLS}]$ simplifies to the usual $\sigma^2(X'X)^{-1}$.

White standard errors

As $X'X = \sum_{i=1}^{n} x_i x_i'$ and (thus) $X'\Omega X = \sum_{i=1}^{n} \sigma_i^2 x_i x_i'$ we have

$$\mathrm{Var}[b_{OLS}] = \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1} \left(\sum_{i=1}^{n} \sigma_i^2 x_i x_i' \right) \left(\sum_{i=1}^{n} x_i x_i' \right)^{-1}$$

White standard errors

As $X'X = \sum_{i=1}^n x_i x_i'$ and (thus) $X'\Omega X = \sum_{i=1}^n \sigma_i^2 x_i x_i'$ we have

$$\text{Var}[b_{OLS}] = \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \left(\sum_{i=1}^n \sigma_i^2 x_i x_i'\right) \left(\sum_{i=1}^n x_i x_i'\right)^{-1}$$

- σ_i^2 is unknown!
- White standard errors: use e_i^2 (with $e_i = y_i - x_i' b_{OLS}$) as estimator for σ_i^2

$\Rightarrow$ Advantage: We do not need a model for the heteroskedasticity

$\Rightarrow$ Disadvantage: OLS is no longer BLUE and if we know the model for heteroskedasticity, OLS is not efficient

Weighted Least Squares

Suppose we know that $\sigma_i^2 = \sigma^2 v_i$, where v_i is known. Estimate β by minimizing the weighted sum of squared residuals

$$\sum_{i=1}^{n} \frac{1}{v_i}(y_i - x_i'\beta)^2.$$

Weighted least squares estimator

$$b_{WLS} = \left(\sum_{i=1}^{n} \frac{1}{v_i} x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} \frac{1}{v_i} x_i y_i\right).$$

This can be written as

$$b_{WLS} = \left(\sum_{i=1}^{n} x_i^* x_i^{*\prime}\right)^{-1} \left(\sum_{i=1}^{n} x_i^* y_i^*\right).$$

where $y_i^* = y_i/\sqrt{v}_i$ and $x_i^* = x_i/\sqrt{v}_i$. That is, we can interpret WLS as OLS in

$$y_i^* = x_i^{*\prime}\beta + \varepsilon_i^*,$$

WLS

It holds that

- $E[\varepsilon_i^{*\,2}] = \sigma^2$ for all $i = 1, \ldots, n$: ε_i^* is homoskedastic.

$\rightarrow$ all assumptions are satisfied in the standardized model, that is, OLS is the optimal estimator (Gauss-Markov).

WLS

It holds that

- $E[\varepsilon_i^{*\,2}] = \sigma^2$ for all $i = 1, \ldots, n$: ε_i^* is homoskedastic.

$\rightarrow$ all assumptions are satisfied in the standardized model, that is, OLS is the optimal estimator (Gauss-Markov).

Properties WLS:

1. $b_{WLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ is unbiased (and consistent): $E[b_{WLS}] = \beta$
2. $\text{Var}[b_{WLS}] = (X'\Omega^{-1}X)^{-1} = \sigma^2 \left(X_*'X_*\right)^{-1} = \sigma^2 \left(\sum_{i=1}^n \frac{1}{v_i} x_i x_i'\right)^{-1}$
3. b_{WLS} is BLUE.

This implies that WLS is more efficient than OLS!

Two-step Feasible WLS

In practice a problem is that v_i is unknown (or unobserved). We can only use WLS if we have an estimate of the variances.

Two-step Feasible WLS

1. Estimate the variances (usually using some function of the OLS residuals)
2. Estimate β using WLS given the variances

Estimation of the variances for two important cases: $\sigma_i^2 = z_i'\gamma$ or $\sigma_i^2 = \exp(z_i'\gamma)$

1. Use e_i^2 of an OLS regression as an estimate for σ_i^2
2. Regress $e_i^2 = z_i'\gamma + \eta_i$ or $\log e_i^2 = z_i'\gamma + \eta_i$

Properties:

- If γ is estimated consistently, then b_{FWLS} is consistent
- In that case: b_{FWLS} is asymptotically efficient

Example 4.1: WLS

Testing for heteroskedasticity

Before applying WLS we should test for heteroskedasticity

H_0 : homoskedasticity and H_a : heteroskedasticity (either general or specific form)

There are many tests, for example

- Breusch-Pagan
- White

Breusch-Pagan test (LM test)

- Idea: regress squared OLS residuals on variables that may relate to the variance
- for models of the type $\sigma_i^2 = h(z_i'\gamma)$ with variables z_i
- $H_0 : \gamma_2 = \cdots = \gamma_p = 0$
- three steps:
  1. estimate $y = X\beta + \varepsilon$ with OLS, compute residuals e
  2. perform auxiliary regression $e_i^2 = \gamma_1 + \gamma_2 z_{2i} + \cdot + \gamma_p z_{pi} + \eta_i$
  3. $LM = nR^2$ of the auxiliary regression, asymptotically distributed as $\chi^2(p-1)$
- The Breusch-Pagan test tests homoskedasticity versus a specific model for heteroskedasticity under the alternative: $\sigma_i^2 = h(z_i'\gamma)$. The function $h()$ does not need to be specified.

White test (LM test)

- Idea: Regress squared residuals on all regressors (and cross terms)
- if variables z_i are unknown, replace the variables by functions of the explanatory variables
- H_0 : homoskedasticity
- White test: chooses squared explanatory terms $x_{2i}^2, ..., x_{ki}^2$ $(p - 1 = 2k - 2)$
- White test with cross terms: includes also all cross products $x_{ji}x_{hi}$ with $j \neq h$
- White-test does not give information on a suitable model for the variances σ_i^2 when homoskedasticity is rejected

Example 4.1: Testing for heteroskedasticity

Auto- or Serial Correlation

Auto- or Serial Correlation

We again start with the linear model

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where now Assumptions 1, 2, (3,) 5, and 6 hold.

Assumption 4: $E(\varepsilon_i \varepsilon_j) = 0$ for all $i \neq j$ does *not* hold.

Auto- or Serial Correlation

We again start with the linear model

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where now Assumptions 1, 2, (3,) 5, and 6 hold.

Assumption 4: $E(\varepsilon_i \varepsilon_j) = 0$ for all $i \neq j$ does *not* hold. Instead we have (auto)correlation:

$$E[\varepsilon_i \varepsilon_j] = \sigma_{ij}, \qquad i \neq j,$$

and (possibly) heteroskedasticity: $E[\varepsilon_i^2] = \sigma_i^2$.
Autocorrelation can be caused by

1. Neglected dynamics
2. Omitted variables
3. Functional misspecification

Model in matrix notation

The model becomes:

$$y = X\beta + \varepsilon,$$

with

$$E[\varepsilon\varepsilon'] = \Omega = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{pmatrix}$$

Usually we assume that σ_{ij} depends on the 'distance' between observations i and j

- For time series data, there is a logical chronological order.
- For cross-sectional data it is crucial to choose a sensible ordering
  $\rightarrow$ For example, choose one of the explanatory variables for the ordering

Example 4.2: Autocorrelation

Consequences of (Auto)correlation

Correlation has consequences for estimating β.

- it holds that the OLS estimator is unbiased (and consistent)

$$\begin{aligned} E[b_{OLS}] &= E[(X'X)^{-1}X'y] = E[(X'X)^{-1}X'(X\beta + \varepsilon)] \\ &= \beta + (X'X)^{-1}X'E[\varepsilon] = \beta. \end{aligned}$$

  Note: assumption is that X is non-stochastic (lagged y variables are not included).

- the usual OLS-standard errors are not correct. The covariance matrix of b_{OLS}, $V[b_{OLS}]$, is given by

$$V[b_{OLS}] = (X'X)^{-1}X'\Omega X(X'X)^{-1}.$$

- only if there is no correlation and homoskedasticity ($\Omega = \sigma^2 \mathrm{I}$), $V[b_{OLS}]$ simplifies to $\sigma^2(X'X)^{-1}$.

Standard errors

Because $X'X = \sum_{i=1}^n x_i x_i'$ and $X'\Omega X = \sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j'$ it holds that

$$V[b_{OLS}] = \left(\sum_{i=1}^n x_i x_i'\right)^{-1} \left(\sum_{i=1}^n \sum_{j=1}^n \sigma_{ij} x_i x_j'\right) \left(\sum_{i=1}^n x_i x_i'\right)^{-1}$$

- A direct estimator of the unknown covariance σ_{ij} is $e_i e_j$ (the cross product of the OLS residuals $e_i = y_i - x_i' b_{OLS}$).
- However, it holds that

$$\sum_{i=1}^n \sum_{j=1}^n e_i e_j x_i x_j' = X'ee'X = 0,$$

so this estimator is not useful to help estimate $V[b_{\mathrm{OLS}}]$.

Newey-West estimator for variance OLS estimator

The solution is to *weigh* the terms $e_i e_j$
$\Rightarrow$ This gives the "Newey-West" estimator of $V[b_{OLS}]$, and the corresponding "Newey-West standard errors".

$$\widehat{V[b_{\mathrm{OLS}}]} = \frac{1}{n}(\frac{1}{n}X'X)^{-1}(\widehat{\frac{1}{n}X'\Omega X})(\frac{1}{n}X'X)^{-1}$$

with

$$\widehat{\frac{1}{n}X'\Omega X} = \frac{1}{n}\sum_{i=1}^{n} e_i^2 x_i x_i' + \frac{1}{n}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_{j-i} e_i e_j (x_i x_j' + x_j x_i')$$

where w_h is the so-called kernel.
Example: Bartlett kernel

$$w_h = \begin{cases} 1 - \frac{h}{B} & h < B \\ 0 & h \geq B \end{cases}$$

$\Rightarrow$Newey-West standard errors are HAC
[Heteroskedasticity and Autocorrelation Consistent]

Correction book

Page 360 gives the following formula

$$\widehat{\text{var}}(b) = \frac{1}{n}(X'X)^{-1}\hat{V}(X'X)^{-1}$$

This should be

$$\widehat{\text{var}}(b) = \frac{1}{n}(\frac{1}{n}X'X)^{-1}\hat{V}(\frac{1}{n}X'X)^{-1}$$

or

$$\widehat{\text{var}}(b) = n(X'X)^{-1}\hat{V}(X'X)^{-1}$$

Estimation under serial correlation

Serial correlation in the linear regression model

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i, \qquad i = 1, \dots, n,$$

can be accounted for in different ways when estimating β_1 and β_2.

1. Include lagged variables $\Rightarrow$ Correlation between $\varepsilon_i = y_i - \beta_1 - \beta_2 x_i$ and $\varepsilon_{i-1} = y_{i-1} - \beta_1 - \beta_2 x_{i-1}$ can be caused by correlation between y_i and y_{i-1} or x_{i-1}. Therefore disturbances η_i in the model

$$y_i = \rho y_{i-1} + \beta_1 + \beta_2 x_i + \beta_3 x_{i-1} + \eta_i, \qquad i = 1, \dots, n,$$

could be uncorrelated.

Estimating under serial correlation

2. Iterative Cochrane-Orcutt procedure: Model the serial correlation with an autoregressive model for disturbances

   Suppose we can capture the serial correlation in the errors using an autoregressive (AR) model of order 1:

$$y_i = \beta_1 + \beta_2 x_i + \varepsilon_i,$$
$$\varepsilon_i = \gamma \varepsilon_{i-1} + \eta_i.$$

   Multiply the equation for y_{i-1} with γ and subtract the result :

$$y_i - \gamma y_{i-1} = \beta_1(1-\gamma) + \beta_2(x_i - \gamma x_{i-1}) + \eta_i.$$

   ⇒Given a value for γ we can estimate β_1 and β_2 using OLS.
   ⇒Given estimates of ε we can estimate γ using OLS in $\hat{\varepsilon}_i = \gamma \hat{\varepsilon}_{i-1} + \eta_i$.

Tests for autocorrelation

Before applying such methods one should test whether there indeed is serial correlation.
Many tests are available, for example

- Durbin-Watson (DW)
- Box-Pierce (BP)
- Ljung-Box (LB)
- Breusch-Godfrey (BG)

All four tests (indirectly) use autocorrelations of the OLS-residuals:

$$r_k = \frac{\sum_{i=k+1}^{n} e_i e_{i-k}}{\sum_{i=1}^{n} e_i^2}, \qquad k = 1, 2, \ldots$$

$H_0 : r_k = 0,\ k = 1, 2, \ldots,$ absence of serial correlation

Tests for autocorrelation

The DW statistic is defined as

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \approx 2(1 - r_1).$$

Tests for autocorrelation

The DW statistic is defined as

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \approx 2(1 - r_1).$$

The BP statistic is defined as

$$BP = n\sum_{k=1}^{p} r_k^2 \approx \chi^2(p) \text{ under } H_0.$$

Tests for autocorrelation

The DW statistic is defined as

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \approx 2(1 - r_1).$$

The BP statistic is defined as

$$BP = n\sum_{k=1}^{p} r_k^2 \approx \chi^2(p) \text{ under } H_0.$$

The LB statistic is defined as

$$LB = n\sum_{k=1}^{p} \frac{n+2}{n-k} r_k^2 \approx \chi^2(p) \text{ under } H_0.$$

Tests for autocorrelation

The DW statistic is defined as

$$DW = \frac{\sum_{i=2}^{n}(e_i - e_{i-1})^2}{\sum_{i=1}^{n} e_i^2} \approx 2(1 - r_1).$$

The BP statistic is defined as

$$BP = n\sum_{k=1}^{p} r_k^2 \approx \chi^2(p) \text{ under } H_0.$$

The LB statistic is defined as

$$LB = n\sum_{k=1}^{p} \frac{n+2}{n-k} r_k^2 \approx \chi^2(p) \text{ under } H_0.$$

Disadvantages of DW, BP & LB

- The test cannot be used if lagged dependent variables are used (needs fixed regressors)

Additional disadvantages for DW

- The distribution of DW under H_0 depends on the properties of the regressors X.

Breusch-Godfrey test

The BG test is an LM-test for $H_0 : \gamma_1 = \ldots = \gamma_p = 0$ in the model

$$y_i = x_i'\beta + \varepsilon_i,$$
$$\varepsilon_i = \gamma_1\varepsilon_{i-1} + \ldots + \gamma_p\varepsilon_{i-p} + \eta_i.$$

This test can be applied to most models and is therefore most suited to test for serial correlation.

Procedure:

1. Estimate parameters under H_0: $y_i = x_i'\beta + \eta_i$
2. Calculate residuals: $e_i = y_i - x_i'b$
3. Regress residuals on x_i *and* p lags of e_i
4. Under H_0: $nR^2 \approx \chi^2(p)$

Note: all x_i belong in the test regression!

Example 4.2: testing for autocorrelation

Generalized Least Squares

Generalized Least Squares

- WLS is a specific example of the method "generalized least squares" (GLS).
- General idea: transform data such that conditions for efficiency of OLS are satisfied
- Consider the linear model in matrix notation

$$y = X\beta + \varepsilon,$$

with $E[\varepsilon\varepsilon'] = \Omega = \sigma^2 V$, and assume that V is known.

$\Rightarrow$ Ω is a covariance matrix, thus it is symmetric and positive definite.

- Therefore there exists an invertible lower triangular matrix P such that

$$PP' = \Omega$$

(A decomposition that satisfies this is called the Choleski decomposition)

GLS

- Now define the transformed data $y^* = P^{-1}y$, $X^* = P^{-1}X$, such that the model becomes

$$y^* = X^*\beta + \varepsilon^*,$$

where $\varepsilon^* = P^{-1}\varepsilon$ with

$$\text{var}(\varepsilon^*) = \text{var}(P^{-1}\varepsilon) = P^{-1}\text{var}(\varepsilon)P^{-1\prime} = P^{-1}\Omega P^{-1\prime} = P^{-1}PP'P^{-1\prime} = \mathrm{I}$$

- In other words, the disturbances ε^* are homoskedastic and free of serial correlation.

$$\begin{aligned} b_{GLS} &= \left(X^{*\prime}X^*\right)^{-1}X^{*\prime}y^* \\ &= \left(X'P^{-1\prime}P^{-1}X\right)^{-1}X'P^{-1\prime}P^{-1}y \\ &= \left(X'(PP')^{-1}X\right)^{-1}X'(PP')^{-1}y \\ &= \left(X'\Omega^{-1}X\right)^{-1}X'\Omega^{-1}y. \end{aligned}$$

Properties of GLS

- Unbiased: $E[b_{GLS}] = \beta$
- The variance is

$$\begin{aligned} \text{var}(b_{GLS}) &= \left(X^{*\prime}X^{*}\right)^{-1} \qquad \text{note that } \text{Var}(\varepsilon^{*}) = 1 \times \text{I} \\ &= \left(X' P^{-1\prime} P^{-1} X\right)^{-1} \\ &= \left(X'\Omega^{-1}X\right)^{-1} \end{aligned}$$

- b_{GLS} is BLUE

Generalized Least Squares – specific examples I

In case of heteroskedasticity of the form $E[\varepsilon_i^2] = \sigma^2 z_i^2$ we have

$$\Omega = \sigma^2 \begin{pmatrix} z_1^2 & 0 & \dots & 0 \\ 0 & z_2^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_n^2 \end{pmatrix}.$$

such that

$$P = \sigma \begin{pmatrix} z_1 & 0 & \dots & 0 \\ 0 & z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_n \end{pmatrix}, P^{-1} = \frac{1}{\sigma} \begin{pmatrix} z_1^{-1} & 0 & \dots & 0 \\ 0 & z_2^{-1} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_n^{-1} \end{pmatrix}.$$

Generalized Least Squares – specific examples II

In case of serial correlation of the form $\varepsilon_i = \rho\varepsilon_{i-1} + \eta_i$ it holds that

$$\Omega = \frac{\sigma_\eta^2}{1-\rho^2}\begin{pmatrix} 1 & \rho & \dots & \rho^{n-1} \\ \rho & 1 & \dots & \rho^{n-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \dots & 1 \end{pmatrix}$$

such that

$$P = \sigma_\eta \begin{pmatrix} \frac{1}{a} & 0 & 0 & \dots & 0 & 0 & 0 \\ \frac{\rho}{a} & 1 & 0 & \dots & 0 & 0 & 0 \\ \frac{\rho^2}{a} & \rho & 1 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots & \vdots & \vdots \\ \frac{\rho^{n-3}}{a} & \rho^{n-4} & \rho^{n-5} & \ddots & 1 & 0 & 0 \\ \frac{\rho^{n-2}}{a} & \rho^{n-3} & \rho^{n-4} & \dots & \rho & 1 & 0 \\ \frac{\rho^{n-1}}{a} & \rho^{n-2} & \rho^{n-3} & \dots & \rho^2 & \rho & 1 \end{pmatrix}, \text{ with } a = \sqrt{1-\rho^2}$$

Feasible GLS

In practice Ω is unknown - GLS cannot be used unless we first get an estimate of Ω. This leads to the “feasible” GLS (FGLS) estimator:

1. Estimate β in $y_i = x_i'\beta + \varepsilon_i$ using OLS (OLS is consistent)
2. Estimate Ω using OLS residuals $e_i = y_i - x_i' b_{OLS}$.
3. Use $\widehat{\Omega}$ to calculate $\widehat{P}$.
4. Transform the data with $\widehat{P}^{-1}$: $y^* = \widehat{P}^{-1} y$ and $X^* = \widehat{P}^{-1} X$.
5. Estimate β with OLS in the model for the transformed data: $y_i^* = x_i^{*\prime}\beta + \varepsilon_i^*$.
6. (One can iterate this procedure: Iterated Feasible GLS)

Let's do a small quiz