Multiple Regression Model: Violations of the G-M assumptions

2025/2026, Semester 1

Ralf A. Wilke
Copenhagen Business School

Heteroskedasticity

Consequences of Heteroskedasticity for OLS

- The homoscedasticity assumption 5 is required for the derivation of the sampling and the asymptotic distribution of OLS in the “classical” regression model.
- Not required for consistency and unbiasedness of OLS.
- If there is no homoskedasticity the estimated variances of $\hat{\beta}_j$ are biased and t- and F-statistics and CIs are not valid anymore.
- OLS no longer efficient (smallest variance).

- A generalised model is by relaxing assumption 5.

Assumption 5' (Heteroscedasticity)

(i) $\mathrm{Var}(u_i|\mathbf{X}) = \sigma_i^2 = \sigma^2\omega_i$ for *i=1,…,N*.

(ii) $\mathrm{Cov}(u_i, u_j|\mathbf{X}) = 0$ for all $i \neq j$.

In matrix form this is

$$\mathrm{Var}(\mathbf{u}|\mathbf{X}) = \sigma^2\mathbf{\Omega},$$

where $\sigma^2\mathbf{\Omega}$ is a (*N x N*) diagonal matrix with diagonal elements σ_i^2. $\mathbf{\Omega}$ is a diagonal matrix with some positive elements $\omega_i > 0$.

- $\mathbf{\Omega}$ is assumed to be positive definite.

- The exact sampling distribution of $\hat{\beta}$ becomes:

Theorem 1.8 (Variance Covariance Matrix of the OLS Estimator)

Under Assumptions 1 through 4 and 5’:

$$\mathrm{var}(\hat{\boldsymbol{\beta}}\,|\mathbf{X}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\boldsymbol{\Omega}\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}.$$

Proof: First, remark that $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}$.

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

and conditional on ***X*** :

$$\begin{aligned} var(\hat{\beta}|\mathbf{X}) &= var[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u}|\mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[var(\mathbf{u}|\mathbf{X})]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'(\sigma^2\boldsymbol{\Omega})\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\Omega}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

- Similarly, the asymptotic variance is:

$$\text{avar}(\hat{\beta}|\mathbf{X}) = (E(\mathbf{x}'\mathbf{x}))^{-1}(E(u^2\mathbf{x}'\mathbf{x}))(E(\mathbf{x}'\mathbf{x}))^{-1}$$

for which a valid estimator is:

$$\widehat{avar}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\left(\sum_i \hat{u}_i^2 \mathbf{x}_i'\mathbf{x}_i\right)(\mathbf{X}'\mathbf{X})^{-1}$$

- This is called the heteroscedasticity robust standard error for $\hat{\beta}_j$.
- Also called Huber or White standard error.
- One can show that the asymptotic distribution of $\hat{\beta}$ is:

Theorem 1.9 (Asymptotic Distribution of the OLS Estimator)

Under Assumptions 1 through 4 - 5':

$$\hat{\boldsymbol{\beta}} \overset{a}{\sim} N(\boldsymbol{\beta}, \sigma^2(E(\mathbf{x}'\mathbf{x}))^{-1}(E(u^2\mathbf{x}'\mathbf{x}))(E(\mathbf{x}'\mathbf{x}))^{-1}).$$

- Common test for individual and joint significance applicable under some modifications.
- For example:
  - A heteroskedastic-robust *t* statistic is:

$$t = \frac{estimate - hypothesized \quad value}{standard \quad error}$$

  - Heteroskedasticity robust versions of Wald test and F test.

- Example: log wage equation (wage1.dta)

$$
\begin{aligned}
\widehat{log(wage)} \quad = \quad & 0.321 + 0.213 marrmale - 0.198 marrfem - 0.110 singfem \\
& (0.100) \quad (0.055) \qquad\qquad (0.058) \qquad\qquad (0.056) \\
& [0.109] \quad [0.057] \qquad\qquad [0.058] \qquad\qquad [0.057] \\
& +0.079 educ + 0.027 exper - 0.00054 exper^2 + 0.029 tenure - 0.00053 tenure^2 \\
& (0.0067) \qquad (0.0055) \qquad (0.00011) \qquad (0.0068) \qquad (0.00023) \\
& [0.0074] \qquad [0.0051] \qquad [0.00011] \qquad [0.0069] \qquad [0.00024] \\
& n = 526, \quad R^2 = 0.461
\end{aligned}
$$

- Usual OLS standard errors are in parentheses, ().
- Heteroskedasticity-robust standard errors are in brackets, [].

- The two sets of standard errors are similar, no changes in significance.
- Heteroskedasticity-robust standard errors can be larger or smaller than the usual OLS standard errors.
- In applications, the heteroskedastic standard errors are often larger.

Testing for Heteroskedasticity

- The null is that Assumption 5 is true: $H_0 : \text{var}(u|\mathbf{x}) = \sigma^2$
  - The heteroskedasticity function is independent of ***x***.
- Because of Assumption 4 the null is equivalent to:

$$H_0 : E(u^2|\mathbf{x}) = E(u^2) = \sigma^2$$

- This means we want to test, whether u^2 is related to one or more of the explanatory variables. The null is false, if u^2 is any function of $\mathbf{x}$.
- The various tests differ in the specification of the heteroskedasticity function.

- A simple approach is to assume a linear function:

$$u^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_K x_K + \nu$$

- The null of homoskedasticity is: $H_0 : \delta_1 = ... = \delta_K = 0$
- However, the problem is that we do not observe u.
- Instead, use the OLS residuals as a consistent estimate for the error:

$$\hat{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_K x_K + \nu$$

- One can show that the OLS residuals asymptotically do not affect the *F* and *LM* statistics.
- For this reason compute the *F* or *LM* statistic for the significance of this regression.

Breusch Pagan Test for Heteroskedasticity

1. Estimate the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ by OLS and obtain $\tilde{u}^2$.
2. Run the regression

$$\tilde{u}^2 = \delta_0 + \delta_1 x_1 + \delta_2 x_2 + ... + \delta_k x_K + \nu$$

   and compute $R^2_{\tilde{u}^2}$.
3. Form either the F or the LM statistic and compare it to the critical value (or compute its P-value). This determines whether we can reject the hypothesis of homoscedasticity.

- Example: Housing price equation (hprice1.dta)

$$\widehat{price} = -21.77 + 0.00207 lotsize + 0.123 sqrft + 13.85 bdrms$$
$$(29.48) \quad (0.00064) \quad (0.013) \quad (9.01)$$
$$n = 88, \quad R^2 = 0.672$$

  - In this case $R^2_{\tilde{u}^2} = 0.1601$, with N=88 and K=3:

$$F = [0.1601/(1-0.1601)](84/3) \approx 5.34$$
$$LM = 88 * 0.1601 \approx 14.10$$

    This implies that the assumption of homoscedasticity is rejected and that OLS statistics are not reliable.

White Test for Heteroskedasticity

- It can be shown that the homoscedasticity assumption $\mathrm{var}(u|\mathbf{x}) = \sigma^2$ can be replaced by the weaker assumption that u^2 is uncorrelated with all x_j, x_j^2 and $x_j x_l \ (j \neq l)$.
- In this case we can regress $\hat{u}^2$ on all x_j, x_j^2 and $x_j x_l \ (j \neq l)$.

$$\begin{aligned} \hat{u}^2 &= \delta_0 + \delta_1 x_1 + \ldots + \delta_K x_K + \delta_{K+1} x_1^2 + \ldots + \delta_{K+K} x_K^2 \\ &\quad + \delta_{K+K+1} x_1 x_2 + \delta_{K+K+2} x_1 x_3 + \ldots + \nu \end{aligned}$$

- The White test for heteroskedasticity is the *LM* (or *F*) statistic for testing that all the δ_j in the equation are zero.
- A disadvantage of the White test is that it uses many degrees of freedom.

- An interesting alternative which preserves the spirit of the White test is to use the OLS fitted values $\hat{\mathbf{y}} = \mathbf{X}\hat{\beta}$, instead of the independent variables, their squares and cross products:

$$\hat{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \nu$$

- We use the fitted values because they are linear functions of the independent variables.
- We can use the F or LM statistic for the null hypothesis

$$H_0 : \delta_1 = 0, \delta_2 = 0$$

- This test conserves more degrees of freedom and it is easy to implement.
- Since $\hat{y}$ is an estimate of *y* given x_j, the test is in particular useful in situations where we suspect that the variance changes with the level of $E(y|\mathbf{x})$.

A special case of the White Test

1. Estimate the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$ by OLS and obtain $\tilde{u}^2$.
2. Run the regression $\tilde{u}^2 = \delta_0 + \delta_1 \hat{y} + \delta_2 \hat{y}^2 + \nu$ and compute $R^2_{\tilde{u}^2}$.
3. Form either the *F* or the *LM* statistic and compute the p-value or critical value using the $F_{2,N-3}$ or the χ^2_2 distribution.

- Note: results between the special form of the white test and Breusch-Pagan test can differ.
- Test for heteroskedasticity are sometimes also considered as test for specification (e.g. omitted variables) because in this case there is also no homoskedasticity. This is, however, not the best approach to test for this.

Generalised Least Squares Estimation

- If there is heteroskedasticity present, OLS is not efficient anymore.
  - Use **GLS/FGLS**.
- This will also lead to correct s.e. and new F and t statistics.

The heteroskedasticity function.

- We assume that $\mathrm{var}(u|x) = \sigma^2\omega(x)$, where *ω(x)>0* is some positive function.
- For a random drawing *i* from the population, we can write: $\sigma_i^2 = var(u_i|\mathbf{x}_i) = \sigma^2\omega(\mathbf{x}_i) = \sigma^2\omega_i$.
- For example, consider the simple savings function:

$$sav_i = \beta_1 + \beta_2 inc_i + u_i \qquad \mathrm{var}(u_i|inc_i) = \sigma^2 inc_i$$

  - Here: *ω(x)=ω(inc)=inc*. The error variance is proportional to the level of income.
- How, can we estimate such a model?
  - We transform it into a model which satisfies the Gauss-Markov assumptions and which has homoskedastic errors.

- Since $\omega_i = \omega(\mathbf{x}_i)$ is just a function of $\mathbf{x}_i$, $u_i/\sqrt{\omega}_i$ has zero expected value and its variance is σ^2 (both conditional on $\mathbf{x}_i$):

$$E\left((u_i/\sqrt{\omega}_i)^2|\mathbf{x}_i\right) = E(u_i^2|\mathbf{x}_i)/\omega_i = (\sigma^2\omega_i)/\omega_i = \sigma^2$$

- Thus, we divide our model by $\sqrt{\omega_i}$:

$$y_i/\sqrt{\omega_i} = \beta_1(x_1/\sqrt{\omega_i}) + ... + \beta_K(x_{iK}/\sqrt{\omega_i}) + u_i/\sqrt{\omega_i}$$

or

$$\begin{aligned} y_i^* &= \beta_1 x_{i1}^* + ... + \beta_K x_{iK}^* + u_i^* \\ &= \mathbf{x}_i^* \beta + u_i^* \end{aligned}$$

- We can estimate this model efficiently by OLS. Why?
  - Because if the original model satisfies Assumptions 1-4, then also this model does, plus Assumption 5 holds for the latter.

- Example: savings function (cont.)
  - $sav_i = \beta_0 + \beta_1 inc_i + u_i \qquad \mathrm{var}(u_i|inc_i) = \sigma^2 inc_i$
  - The transformed model is:

$$sav_i/\sqrt{inc_i} = \beta_0(1/\sqrt{inc_i}) + \beta_1\sqrt{inc_i} + u_i^*$$

  - The coefficient β_1 in this model is still the marginal propensity to save, the interpretation does not change.

- The estimators of the transformed model are examples of the generalized least squares estimators (**GLS**).
- All statistics (except the R^2) and estimators of the transformed model will be valid, the interpretation of the coefficients is as in the original model.

- The GLS estimator for correcting heteroskedasticity is:

$$\begin{aligned} \beta^* &= (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}(\mathbf{X}^{*'}\mathbf{y}^*) \\ &= (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{y}) \\ &= \beta + (\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u}) \end{aligned}$$

  - When we assume:

$E(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{X}) = A$ is nonsingular *KxK*.

$$E(\mathbf{X}'\mathbf{\Omega}^{-1}\mathbf{u}) = \mathbf{0}$$

Consistency follows from an application of a law of large numbers and Slutsky’s theorem:

$$plim\beta^* = \beta + \mathbf{A}^{-1}\mathbf{0} = \beta$$

- β^* is asymptotically normal with:

$$\sqrt{N}(\boldsymbol{\beta}^* - \boldsymbol{\beta}) \overset{a}{\sim} N(\mathbf{0}, \sigma^2(E(\mathbf{X}'\boldsymbol{\Omega}^{-1}\mathbf{X}))^{-1})$$

  - Asymptotically efficient because the OLS estimator on the transformed data has minimum variance (Gauss-Markov).

- β^* is unbiased under a zero conditional mean condition on the error.

- Requires known $\boldsymbol{\Omega}$.

The heteroskedasticity Function must be estimated: Feasible GLS (FGLS)

- If we do not have an idea about the function $\omega(\mathbf{x}_i)$, we need to estimate it from data.
- Using $\hat{\omega}_i$ instead of ω_i in GLS, yields an estimator called **FGLS** estimator.

- Example of a positive variance function:

$$\text{var}(u|\mathbf{x}) = \sigma^2 exp(\delta_0 + \delta_1 x_1 + ... + \delta_k x_k)$$

with $\delta_0, ...\delta_k$ as unknown parameters.

- We can write using $E(\nu|\mathbf{x}) = 1$:

$$u^2 = \sigma^2 exp(\delta_0 + \delta_1 x_1 + ... + \delta_k x_k)\nu$$

and we assume $\text{cov}(\mathbf{x}, \nu) = 0$. We obtain:

$$\log(u^2) = \alpha_0 + \delta_1 x_1 + ... + \delta_k x_k + e$$

with *e* has zero conditional mean and is independent of ***x***. In general $\delta_0 \neq \alpha_0$ but this is not important.
- Since the Gauss-Markov assumptions are satisfied in this regression, we obtain consistent and unbiased estimates of δ_j.
- In an application, we replace *u* by its estimate $\hat{u}$ and regress $log(\hat{u}^2)$ on $\mathbf{x}$.
- Denote the fitted values of this regression as $\hat{g}_i$ and estimate ω_i by $\hat{\omega}_i = exp(\hat{g}_i)$.

- A Feasible GLS procedure to correct for heteroskedasticity.
  1. Run the regression of y on ***x*** and obtain the residuals $\hat{u}$.
  2. Create $\log(\hat{u}^2)$.
  3. Run the regression $log(\hat{u}^2)$ on $\mathbf{x}$ and compute the fitted values $\hat{g}_i$
  4. Compute $\hat{\omega}_i = exp(\hat{g}_i)$.
  5. Estimate the equation

$$y = \beta_0 + \beta_1 x_1 + ... + \beta_k x_k + u$$

  by WLS, using weights $1/\hat{\omega}$. Each equation is divided by $\sqrt{\hat{w}}$.

- Since ω_i is now estimated, FGLS is not an unbiased estimator, but it is still consistent. Moreover, one can show that it is more efficient than OLS.
- FGLS is an attractive alternative to OLS if there is evidence for unknown heteroscedasticity.

- Example: Demand for cigarettes (smoke.dta)
  - We estimate the demand function for daily cigarette consumption.
  - The dependent variable is often zero, because there are non smokers in the data. A linear model is therefore not optimal because it predicts negative values but we can still learn something from this model.
  - We estimate by OLS:

$$
\begin{aligned}
\widehat{cigs} \quad = \quad & -3.64 + 0.880 log(income) - 0.751 log(cigpric) \\
& (24.08) \quad (0.728) \qquad\qquad\qquad (5.773) \\
& -0.501 educ + 0.771 age - 0.0090 age^2 - 2.83 restaurn \\
& (0.167) \qquad\quad (0.160) \qquad\quad (0.0017) \qquad (1.11) \\
& n = 807, \quad R^2 = 0.0526
\end{aligned}
$$

with *restaurn*: smoking restrictions in restaurant at state level.

  - 2% of the fitted values are below 0.
  - Is heteroskedasticity present? We compute the Breusch-Pagan test and obtain LM=32.26, which is strong evidence of heteroskedasticity.

- We estimate FGLS according to our specification of the conditional variance function:

$$
\begin{aligned}
\widehat{cigs} \quad = \quad & 5.64 + 1.30 log(income) - 2.94 log(cigpric) \\
& (17.80) \quad (0.44) \qquad\qquad\qquad (4.46) \\
& -0.463 educ + 0.482 age - 0.0056 age^2 - 3.46 restaurn \\
& (0.120) \qquad\quad (0.097) \qquad\quad (0.0009) \qquad (0.80) \\
& n = 807, \quad R^2 = 0.1134
\end{aligned}
$$

- Compared to the OLS regression:
- The income effect is now statistically significant and larger.
- The price effect is bigger but still insignificant.

- Remark: F statistic for testing multiple exclusion restrictions based on WLS estimates cannot directly be used. They have to be reweighed.

Summary

- We analysed the properties of OLS in presence of heteroskedasticity.
  - It does not cause bias or inconsistency, just inefficiency
  - Standard errors and t- and F-statistics are not valid anymore.
  - We have seen a heteroskedasticity robust estimator for the standard error. Many computer packages compute heteroskedasticity robust statistics.
- We have seen two tests for heteroskedasticity: the Breusch-Pagan and the White test.
- We have introduced GLS/FGLS as an efficient alternative to OLS.

Serial Correlation

Main text: parts of Chapter 12 in Wooldridge "Introductory Econometrics", 2025.

- Error terms are correlated across observations/periods.
  - Common in panel data where the same unit has repeated observations.
  - Common in time series data where the same unit is observed repeatedly.
  - Also called Autocorrelation.
- Given that serial correlation is normally due to the longitudinal dimension (time), we use the index $t=1,\ldots,N$.
- Serial correlation can be sometimes removed by using differenced data (between periods) instead of the levels.
- Similar to heteroscedasticity, OLS is still consistent but no longer BLUE.
  - Gauss-Markov Theorem requires homoskedasticity and serially uncorrelated errors.

- The variance of the OLS estimator may be smaller or larger than the usual OLS variance under homoskedasticity, depending on the direction of the serial correlation.

- Testing for Serial correlation:
  - Strictly exogenous regressors: $E(u_t|\mathbf{X}) = 0$
    - Error in period *t* not correlated with regressors from any period.
    - T-test for AR(1), F/LM test for higher order serial correlation

- Serial error correlation in models with lagged dependent variable as regressor require special care (also called AR-models).

$$y_t = \beta_0 + \beta_1 y_{t-1} + u_t$$

  - Dynamic structure.
  - Time series/financial econometrics.
  - **Dynamic panel models.**

  - Breusch-Godfrey Test (F-/LM test) for higher order serial correlation

- T-test for AR(1)
  - Large sample test
  - We consider the AR(1) error model

$$u_t = \rho u_{t-1} + e_t, t = 1, 2, ..., n$$

with $E(e_t|u_{t-1}, u_{t-2}, ...) = 0$ and $var(e_t) = \sigma_e^2$. $|\rho| \leq 1.$

  - Null Hypothesis: $H_0 : \rho = 0$
  - How can we test for this?
    - Under the assumptions above, the OLS estimator for the model above is asymptotically normal.
    - Run the regression of u_t on u_{t-1} (without intercept) and conduct a t-test!
    - Since u_t are not observed, use a consistent estimate, the OLS residuals of the main model: $\hat{u}_t$.
      - The large sample distribution of the t-statistic is not affected by this (when regressors are strictly exogenous).

- Step by step procedure:

1. Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all *t=1,…,n*.
2. Run the regression of $\hat{u}_t$ on $\hat{u}_{t-1}$ for all *t=2,…,n* and construct the t-statistic $t_{\hat{\rho}}$ for the coefficient $\hat{\rho}$ on $\hat{u}_{t-1}$. (the regression may contain an intercept or not. Asymptotically this does not affect $t_{\hat{\rho}}$).
3. Use $t_{\hat{\rho}}$ to test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$ in the usual way. Normally, conclude that there is evidence for serial correlation when H_0 is rejected at the 5% level.

- Variation: To make this test robust to heteroskedasticity in e_t, use the usual heteroskedasticity robust version for $t_{\hat{\rho}}$.

<u>Example</u>: Testing for AR(1) serial correlation in the Phillips curve.

- Macro relation between inflation and unemployment rate.
- Data: phillips.dta , sample code: ar1_ttest.R

- Testing for higher order serial correlation AR(q)
  - An extension of the t-test approach for AR(1) is straightforward
  - Suppose there is serial correlation of order *q* in errors :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + ... + \rho_q u_{t-q} + e_t, t = 1, 2, ..., N$$

  - The null hypothesis is: $H_0 : \rho_1 = 0, \rho_2 = 0, ..., \rho_q = 0$
  - The idea is to apply a test for joint significance of $\hat{u}_{t-1}, \hat{u}_{t-2}, ..., \hat{u}_{t-q}$.
  - This can be either the F-test or LM-test.
  - Heteroskedasticity robust version as for *q=1*.
  - As for *q=1*, it is an asymptotic test.

- [For finite sample test: Durbin-Watson test]

- Step by step procedure

1. Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all *t=1,…,n*.
2. Run the regression of $\hat{u}_t$ on $\hat{u}_{t-1}, \ldots, \hat{u}_{t-q}$ for all *t=q+1,…,n* .
3. Compute the F-test or LM test for joint significance of $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-q}.$

Reminder: The LM statistic in 3. is $LM = (n-q)R^2_{\hat{u}} \sim \chi^2_q$, where is the R-squared of the model in 2.

- Example: arq_test.R

- An interesting variation is to test only for selected lagged errors to be serially correlated, for example u_t and u_{t-4} if the data are quarterly.
  - Then only use the relevant lagged residuals on the right hand side in the model in 2.

- The LM test version of this test is a special case of a more general model without strictly exogenous regressors.
  - This is for example the case if one uses a lagged dependent variable as regressor (quite common in time series analysis).
  - It can be shown that if one uses $x_{t1}, x_{t2}, \ldots, x_{tk}$ as additional regressors in model 2, removes the “endogeneity distortion” in the statistic in 3.
  - The LM version of this test is called **Breusch-Godfrey** test for AR(q) serial correlation.

- Step by step procedure of Breusch-Godfrey test

1. Run the OLS regression of y_t on $x_{t1}, \ldots, x_{tk}$ and obtain the OLS residuals, $\hat{u}_t$, for all *t=1,…,n*.
2. Run the regression of $\hat{u}_t$ on $x_{t1}, x_{t2}, \ldots, x_{tk}, \hat{u}_{t-1}, \ldots, \hat{u}_{t-q}$ for all *t=q+1,…,n* .
3. Compute the LM test for joint significance of $\hat{u}_{t-1}, \hat{u}_{t-2}, \ldots, \hat{u}_{t-q}$.

- Suppose there is evidence of serial correlation.
  - OLS estimates are still consistent but no longer BLUE and standard errors are incorrect.
  - What can we do about this?

1. Transforming data:
   - Similar to heteroskedasticity there exist GLS transformations. Application of OLS to transformed data produces estimates that are BLUE. FGLS requires pre-estimation of ρ and transformation of data in t=1 requires special care. Transformation depends on order of serial correlation.
   - If ρ is positive, large and even possibly =1 (random walk), first differencing of the data is appealing.

2. Serial correlation robust standard errors

- (First-)Differencing and serial correlation:
  - Consider the differenced version of the model

$$y_t = \beta_0 + \beta_1 x_t + u_t, \quad t = 1, ..., n$$

that is

$$\Delta y_t = \beta_1 \Delta x_t + \Delta u_t, \quad t = 2, ..., n$$

with $\Delta y_t = y_t - y_{t-1}$, $\Delta x_t = x_t - x_{t-1}$ and $\Delta u_t = u_t - u_{t-1}$.

  - It can be shown that the differencing removes strong positive serial correlation but increases the amount of serial correlation when it is negative.

  - The interpretation of β_1 is the same in the two models.
  - OLS of differenced data produces (more) correct standard errors when u_t possesses strong positive serial correlation.

Example of first differencing: FD.R

- Data: (intdef.dta) time series of US macroeconomic data on three-months T-bill rate (*i3* – a measure for short-term interest rates), annual inflation rate (*inf*) and federal budget deficit as % of GDP (*def*).
- Estimate level and first differenced version of the model:

$$i3_t = \beta_0 + \beta_1 inf_t + \beta_2 def_t + u_t$$

- Do test of first order serial correlation in the two models.
- It is found that $\hat{\rho} = 0.623$ and highly significant in the level model and $\hat{\rho} = 0.073$ and insignificant in the differenced model. Better to use FD model!
- [Another observation: If regressors are exogenous (uncorrelated with the error term), both estimates are consistent. The results, however, point to significant changes. This likely points to endogeneity in the level model. (Unit root or omitted heterogeneity.) Another reason for FD!]

- Serial correlation robust standard errors and statistics
  - This is appealing because it does not require knowledge of the form of serial correlation (it applies to fairly arbitrary forms of serial correlation).
  - A general approach to correct the standard errors for fairly arbitrary forms of heteroskedasticity and serial correlation has been suggested by Newey and West (1987).
  - No details here but example code: nw_se.R
    - Estimation of Phillips curve with serial correlation robust standard errors.
  - S.E. can both increase or decrease but likely to increase if $\rho > 0$.