Heteroskedasticity and serial correlation. Generalized least squares estimator. Weighted least squares. Robust and clustered standard errors.

Jakub Mućk
SGH Warsaw School of Economics

Least squares estimator

- **Least squares estimator** :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \varepsilon \tag{1}$$

  where
  - y is the (outcome) dependent variable;
  - $x_1, x_2, \ldots, x_K$ is the set of independent variables;
  - ε is the error term.
- The dependent variable is explained with the components that vary with the **the dependent variable** and **the error term**.
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_K$ are the coefficients (slopes) on $x_1, x_2, \ldots, x_K$.

- **Least squares estimator** :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_K x_K + \varepsilon \quad (1)$$

  where
  - y is the (outcome) dependent variable;
  - $x_1, x_2, \ldots, x_K$ is the set of independent variables;
  - ε is the error term.
- The dependent variable is explained with the components that vary with the **the dependent variable** and **the error term**.
- β_0 is the intercept.
- $\beta_1, \beta_2, \ldots, \beta_K$ are the coefficients (slopes) on $x_1, x_2, \ldots, x_K$.

$\beta_1, \beta_2, \ldots, \beta_K$ measure the effect of change in $x_1, x_2, \ldots, x_K$ upon the expected value of y (*ceteris paribus*).

- **Assumption #1**: true DGP (data generating process):

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon. \tag{2}$$

- **Assumption #2**: the expected value of the error term is zero:

$$\mathbb{E}\left(\varepsilon\right) = 0, \tag{3}$$

and this implies that $\mathbb{E}\left(y\right) = \mathbf{X}\beta$.

- **Assumption #3**: Spherical variance-covariance error matrix.

$$var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon') = I\sigma^2 \tag{4}$$

. In particular:

  - the variance of the error term equals σ:

$$var\left(\varepsilon\right) = \sigma^2 = var\left(y\right). \tag{5}$$

  - the covariance between any pair of ε_i and ε_j is zero"

$$cov\left(\varepsilon_i, \varepsilon_j\right) = 0. \tag{6}$$

- **Assumption #4**: **Exogeneity.** The independent variable are **not random** and therefore they are not correlated with the error term.

$$\mathbb{E}(\mathbf{X}\varepsilon) = 0. \tag{7}$$

- **Assumption #5**: the full rank of matrix of explanatory variables (there is no so-called collinearity):

$$rank(\mathbf{X}) = K + 1 \leq N. \tag{8}$$

- **Assumption #6 (optional)**: the normally distributed error term:

$$\varepsilon \sim \mathcal{N}\left(0, \sigma^2\right). \tag{9}$$

Assumptions of the least squares estimators

Under the assumptions A#1-A#5 of the multiple linear regression model, the least squares estimator $\hat{\beta}^{OLS}$ has the smallest variance of all linear and unbiased estimators of β.

$\hat{\beta}^{OLS}$ is the **Best Linear Unbiased Estimators (BLUE)** of β.

- **The least squares estimator**

$$\hat{\beta}^{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}. \tag{10}$$

- **The variance of the least square estimator**

$$Var(\hat{\beta}^{OLS}) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1} \tag{11}$$

- **If the (optional) assumption about normal distribution of the error term is satisfied** then

$$\beta \sim \mathcal{N}\left(\hat{\beta}^{OLS}, Var(\hat{\beta}^{OLS})\right). \tag{12}$$

Consequences of non spherical errors

- General variance of the least square estimator ($\hat{\beta}^{OLS}$):

$$Var(\hat{\beta}^{OLS}) = \mathbb{E}\left[\left(\hat{\beta}^{OLS} - \beta\right)\left(\hat{\beta}^{OLS} - \beta\right)'\right]. \tag{13}$$

- Let rewrite the least square estimator:

$$\hat{\beta}^{OLS} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\left(\mathbf{X}\beta + \varepsilon\right) = \beta + \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon, \tag{14}$$

- then

$$\begin{aligned} Var(\hat{\beta}^{OLS}) &= \mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\left(\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\right)'\right] \\ &= \mathbb{E}\left[\left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\right] \\ &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbb{E}\left[\varepsilon\varepsilon'\right]\mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)^{-1} \end{aligned}$$

- If the assumption #3 about **spherical variance-covariance error matrix**, i.e.. $\mathbb{E}(\varepsilon\varepsilon') = \sigma^2 I$ is not satisfied, the above expression **cannot be simplified** and written as:

$$Var(\hat{\beta}^{OLS}) = \sigma^2\left(\mathbf{X}'\mathbf{X}\right)^{-1}. \tag{15}$$

SGH

- **Consequences**
  - The least squares estimator is still unbiased and consistent but it **no longer BLUE**.
  - **Inconsistency of variance**. The standard errors usually computed for the least squares estimator are unreliable.
    - Confidence intervals and hypothesis tests that use these standard errors may be misleading.
- **Detection**
  - **Visual inspection of residuals.**
  - **Formal tests.**
- **Dealing with non spherical errors**
  - **(Feasible) Generalized Least Squares**.
  - **Robust standard errors**.
- **Special cases**
  - **Heteroskedasticity of the error term.**
  - **Serial correlation.**

Heteroskedasticity

Homoskedasticity

- The simple linear model:

$$y_i = -\beta_0 + \beta_1 x_i + \varepsilon_i \qquad var(\varepsilon_i) = \sigma^2, \tag{16}$$

the variance of the least squares estimators for β_1:

$$var(\hat{\beta}_1^{LS}) = \frac{\sigma^2}{\sum_{i=1}^{N} (x_i - \bar{x})^2} \tag{17}$$

Heteroskedasticity

- The simple linear model (with heteroskedasticity):

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad var(\varepsilon_i) = \sigma_i^2, \tag{18}$$

the variance of the least squares estimators for β_1:

$$var(\hat{\beta}_1^{LS}) = \sum_{i=1}^{N} w_i \sigma_i^2 = \frac{\sum_{i=1}^{N} (x_i - \bar{x})^2 \sigma_i^2}{\left[\sum_{i=1}^{N} (x_i - \bar{x})^2\right]^2} \tag{19}$$

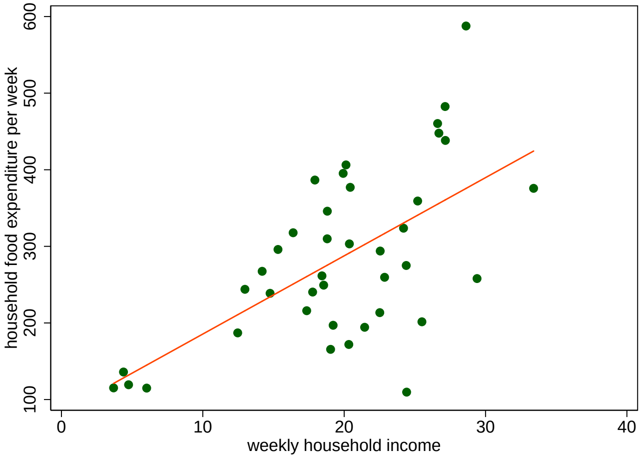
- Heteroskedasticity is often encountered when using **cross-sectional data**.
- Cross-section data invariably involve observation units of varying sizes, e.g., households, firms, workers.
- Intuition: as the size of the economic unit becomes larger, there is more uncertainty associated with the outcomes.
- Heteroskedasticity is sometimes present in time-series data.

Detecting Heteroskedasticity

Method that can be used to detect heteroskedasticity

1. An informal way using residual charts, i.e., **the squared** residuals versus explanatory variables.
2. A formal way using statistical tests:
   2.1 The Breusch-Pagan test;
   2.2 The White test;
   2.3 The Goldfeld-Quandt test.

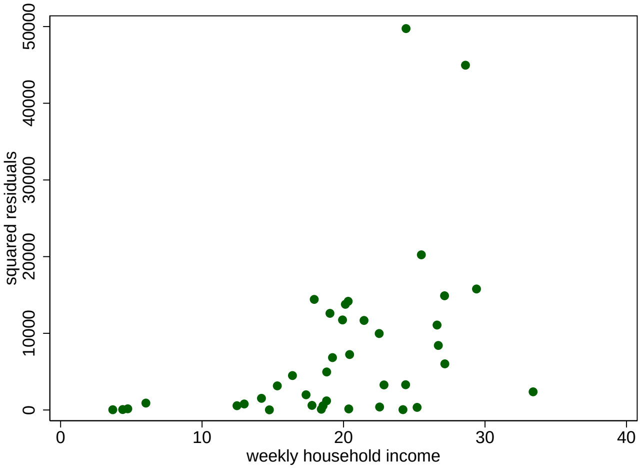
Example



household food expenditure per week
100
200
300
400
500
600
0
10
20
30
40
weekly household income

Example

SGH



squared residuals
0
10000
20000
30000
40000
50000
0
10
20
30
40
weekly household income

- **The Breusch-Pagan Lagrange Multiplier test** allows to test whether the variance of the error term depends on some explanatory variables z that are possibly different from x
- A **general form** for the variance function

$$var(y_i) = \sigma_i^2 = \mathbb{E}(\varepsilon_i^2) = h\left(\alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS}\right). \tag{20}$$

- Two possible functions for $h()$:
  - Exponential function:

$$h\left(\alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS}\right) = \exp\left(\alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS}\right). \tag{21}$$

  - Linear function:

$$h\left(\alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS}\right) = \alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS}, \tag{22}$$

    it should be noted that in the linear function one must be careful to ensure $h() > 0$.
- The null and alternative hypotheses are:

$$\begin{aligned} \mathcal{H}_0 : &\quad \alpha_1 = \alpha_2 = \ldots = \alpha_S = 0, \\ \mathcal{H}_1 : &\quad \text{not all} \quad \alpha_j = 0. \end{aligned}$$

The Breusch-Pagan test II

- The null is about homoskedasticity while the alternative is about heteroskedasticity.
- Note that for linear function we have:

$$\varepsilon_i^2 = \mathbb{E}(\varepsilon_i^2) + \nu_i = \alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS} + \nu_i, \tag{23}$$

  where ν_i is random.
- The test statistics based on the above regression (for linear function) obtained after substitution the least squares residuals$\hat{\varepsilon}_i^2$ for ε_i^2 :

$$\hat{\varepsilon}_i^2 = \alpha_0 + \alpha_1 z_{i1} + \ldots + \alpha_S z_{iS} + \nu_i. \tag{24}$$

- Finally, the test statistics based on the R^2 from the previous regression has a chi-square distribution with S degrees of freedom:

$$\chi^2 = NR^2 \sim \chi^2_{(S)}. \tag{25}$$

- The Breusch-Pagan/ Lagrange Multiplier test is a large sample test.
- In this test, the value of the statistic computed from the linear function is valid for testing an alternative hypothesis of heteroskedasticity where the variance function can be of any form given by $h()$.

- In the **White test** the explanatory variables x, their squares and cross-products are used instead of z.
- **Example.** In the linear

$$\mathbb{E}(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2. \tag{26}$$

the following variables will be used

$$z_1 = x_1, \quad z_2 = x_2, \quad z_3 = x_1^2, \quad z_4 = x_2^2, \quad \text{and} \quad z_5 = x_1 x_2.$$

- The White test is perform as $\mathcal{F}$ test or χ^2 test (as previously).
- The null is about homoskedasticity while the alternative is about heteroskedasticity.

- **The Goldfeld-Quandt test** is designed to test for this form of heteroskedasticity, where the sample can be partitioned into two groups and we suspect the variance could be different in the two groups.
- The sample can be partitioned with:
  - indicator variable,
  - qualitative variable.
- **Example:** wages for female and male workers:

$$\ln wage_i = \beta_0 + \beta_1 educ_i + \beta_2 female_i + \varepsilon_i, \quad i = 1, 2, \ldots, N. \tag{27}$$

- Splitting sample:

$$\ln wage_{Mi} = \beta_{M0} + \beta_{M1} educ_{Mi} + \varepsilon_{Mi}, \quad i = 1, 2, \ldots, N_M, \tag{28}$$

$$\ln wage_{Fi} = \beta_{F0} + \beta_{F1} educ_{Fi} + \varepsilon_{Fi}, \quad i = 1, 2, \ldots, N_F. \tag{29}$$

- The null hypothesis:

$$\mathcal{H}_0 \quad \sigma^2_M = \sigma^2_F. \tag{30}$$

- Test statistics:
$$\mathcal{F} = \frac{\hat{\sigma}^2_M / \sigma^2_M}{\hat{\sigma}^2_F / \sigma^2_F} \sim \mathcal{F}_{(N_M - K_M, F_M - K_F)}, \tag{31}$$
  when the null is true
$$\mathcal{F} = \frac{\hat{\sigma}^2_M}{\hat{\sigma}^2_F}, \tag{32}$$
  when $\hat{\sigma}^2_M > \hat{\sigma}^2_F$
- If the $\mathcal{F}$ is higher than its critical value we can reject the null.

Heteroskedasticity-Consistent Standard Errors

- In the presence of the heteroskedasticity the least squares estimator, although still being unbiased, is no longer blue.
- The typical least squares standard errors are incorrect.
- **Heteroskedasticity-Consistent Standard Errors** is a way of correcting the standard errors so that our interval estimates and hypothesis tests are valid since they take into consideration heteroskedasticity.
- **White's heteroskedasticity-consistent estimator** for the simple linear model:

$$\hat{var}\left(\hat{\beta}_1^{LS}\right) = \frac{N}{N-2}\frac{\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2\hat{\varepsilon}_i^2}{\left[\sum_{i=1}^{N}\left(x_i-\bar{x}\right)^2\right]^2}. \tag{33}$$

- The White's estimator for the variance helps avoid computing incorrect interval estimates or incorrect values for test statistics in the presence of heteroskedasticity but it does not address the other implication of heteroskedasticity.
    - But when sample size is large the variance of the least squares estimator may still be sufficiently small to get precise estimates.
    - Robust standard errors estimator does not require to specify a suitable variance function $h()$.

Clustered standard errors

- Clustered standard errors could be applied in the presence of the heteroskedasticity and when observation can be grouped/clustered.
- **Example:** student's result and classes.
- **Key assumption:** independence (of the error term) between clusters and dependence within clusters.
- The variance-covariance of the error term:

$$\mathbb{E}(\varepsilon_i \varepsilon_j) = \begin{cases} 0 & \text{if } i \text{ and } j \text{ belong to different clusters,} \\ \sigma_{ij} & \text{if } i \text{ and } j \text{ belong to the same group} \end{cases} \tag{34}$$

- The general variance-covariance of the error term matrix will be block diagonal.
- Denoting the group by $g = 1, 2, \ldots, G$, the variance can be estimated:

$$Var(\hat{\beta}^{LS}) = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \left(\sum_{g=1}^{G} \mathbf{x}' \hat{e}_g' \hat{e}_g \mathbf{x} \right) \left(\mathbf{X}'\mathbf{X}\right)^{-1}. \tag{35}$$

- **Key problem:** we should know our data to appropriately apply clustering.

Generalized Least Squares

GLS: known form of variance I

- Consider simply linear regression:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{36}$$

where the error term is heteroskedastic, i.e., $var(\varepsilon_i) = \sigma_i^2$.

- The **generalized least squares estimator (GLS)** depends on the **unknown variance of the error term σ_i^2**.
- However, one can assume some structure on σ_i^2. For instance,

$$var(\varepsilon_i) = \sigma_i^2 = \sigma^2 x_i. \tag{37}$$

- Under above assumption we can apply GLS transformation to our variables (dependent, explanatory and error term):

$$\frac{y_i}{\sqrt{x_i}} = \beta_0 \frac{1}{\sqrt{x_i}} + \beta_1 \frac{x_i}{\sqrt{x_i}} + \frac{\varepsilon_i}{\sqrt{x_i}}, \tag{38}$$

or more generally

$$z_i^* = \frac{z_i}{\sqrt{x_i}}. \tag{39}$$

GLS: known form of variance II

- The variance of the transformed error term is therefore constant:

$$var(\varepsilon_i^*) = var(\frac{\varepsilon_i}{\sqrt{x_i}}) = \frac{1}{x_i} var(\varepsilon_i) = \frac{1}{x_i}\sigma^2 x_i = \sigma^2. \quad (40)$$

- Therefore the least squares estimator can be applied to the regression that bases on transformed variables.

$$y_i^* = \beta_0 + \beta_1 x_i^* + \varepsilon_i^*. \quad (41)$$

- The GLS transformation/estimator can be viewed as a **weighted least squares** estimator:
  - Minimizing sum of ε_i^*, i.e., the transformed errors:

$$\sum_{i=1}^{N} \varepsilon_i^{*2} = \sum_{i=1}^{N} \frac{\varepsilon_i^2}{x_i} = \sum_{i=1}^{N} \left(\frac{\varepsilon_i}{x_i^{1/2}}\right)^2. \quad (42)$$

  - The error are weighted by $1/x_i^{1/2}$.
  - **Intuition:** observation with smaller error variance has a larger weights (importance).

- **Example:** wages for female and male workers in the divided samples"
- Splitting sample:

$$\ln wage_{Mi} = \beta_{M0} + \beta_{M1} educ_{Mi} + \varepsilon_{Mi}, \quad i = 1, 2, \ldots, N_M, \quad (43)$$
$$\ln wage_{Fi} = \beta_{F0} + \beta_{F1} educ_{Fi} + \varepsilon_{Fi}, \quad i = 1, 2, \ldots, N_F. \quad (44)$$

- The GLS estimator can be applied as follows:

$$\frac{\ln wage_{Mi}}{\sigma_M} = \beta_{M0}\frac{1}{\sigma_M} + \beta_{M1}\frac{educ_{Mi}}{\sigma_M} + \frac{\varepsilon_{Mi}}{\sigma_M}, \quad i = 1, 2, \ldots, N_M, (45)$$
$$\frac{\ln wage_{Fi}}{\sigma_F} = \beta_{F0}\frac{1}{\sigma_F} + \beta_{F1}\frac{educ_{Fi}}{\sigma_F} + \frac{\varepsilon_{Fi}}{\sigma_F}, \quad i = 1, 2, \ldots, N_F. \quad (46)$$

  where σ_M and σ_F is the standard deviation of the error term in the subsamples form male and female workers, respectively.
- How to get σ_M and σ_F estimates?
- We can use a **Feasible Generalized Least Squares (FGLS)** estimator. The steps are as follows:
  1. Obtained σ_M and σ_F estimates by applying the least squares separately to both subsamples (like in the Goldfeld-Quandt test).

2. Construct the general variance of the error term:

$$\hat{\sigma}_i = \begin{cases} \hat{\sigma}_M & \text{if} \quad FEMALE_i = 0, \\ \hat{\sigma}_F & \text{if} \quad FEMALE_i = 1. \end{cases} \tag{47}$$

3. Apply the least squares to the transformed initial model:

$$\frac{\ln wage_i}{\hat{\sigma}_i} = \beta_0 \frac{1}{\hat{\sigma}_i} + \beta_1 \frac{educ_i}{\hat{\sigma}_i} + \beta_2 \frac{female_i}{\hat{\sigma}_i} + \frac{\varepsilon_i}{\hat{\sigma}_i}. \tag{48}$$

General steps in applying the GLS when the form of variance is unknown

1. Estimate equation by least squares and compute the squares of the least squares residual $\hat{\varepsilon}_i$
2. Take squared residuals ($\hat{\varepsilon}_i^2$) and apply the least squares to the equation describing the variance. One of the possible form is :
$$\ln \hat{\varepsilon}_i^2 = \alpha_1 + \alpha_2 z_1 + \ldots + \alpha_S z_S + \nu_i \tag{49}$$
where ν_i is the random and z_j is explanatory variable or its transformation (e.g. logarithmic).
3. Compute the estimated variance $\hat{\sigma}_i^2$. For the example from previous point:
$$\hat{\sigma}_i^2 = \exp\left(\hat{\alpha}_1 + \hat{\alpha}_2 z_1 + \ldots + \hat{\alpha}_S z_S\right). \tag{50}$$
4. Transform variables (dependent and explanatory):
$$y_i^* = y_i/\hat{\sigma}_i \quad \text{and} \quad x_{ji}^* = x_{ji}/\hat{\sigma}_i. \tag{51}$$
5. Apply the least squares to transformed variables.

Serial correlation

- Serial correlation of error term is usually present in **time series**.
- In general, serial correlation is a measure of **persistence/inertia**. This is a common feature of many economic variables.
- Serial correlation of the error term suggests that **dynamic relationship** between variables is **misspecified.**

Detecting serial correlation of the error term

SGH

- **An informal way using residual charts**:
  - plotting residuals $\hat{e}_t$ versus time,
  - plotting residuals $\hat{e}_t$ versus lagged residuals $\hat{e}_{t-1}$,

- **An informal way using residual charts**:
  - plotting residuals $\hat{e}_t$ versus time,
  - plotting residuals $\hat{e}_t$ versus lagged residuals $\hat{e}_{t-1}$,
- **Formal ways using statistical tests**:
  - Testing autocorrelation of order one as well as of higher orders.
  - The Lagrange multiplier test.
  - The Durbin-Watson test.

- The population correlation x and y:

$$\rho_{xy} = \frac{cov\left(x, y\right)}{\sqrt{var\left(x\right) var\left(y\right)}}. \tag{52}$$

- **he population autocorrelation of order one:**

$$\rho_1 = \frac{cov\left(y_t, y_{t-1}\right)}{\sqrt{var\left(y_t\right) var\left(y_{t-1}\right)}}. \tag{53}$$

- **The sample autocorrelation (ACF) of order one:**

$$r_1 = \frac{\sum_{t=2}^{T}\left(y_t - \bar{y}\right)\left(y_{t-1} - \bar{y}\right)}{\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2}. \tag{54}$$

- **The k-th order sample autocorrelation (ACF) :**

$$r_k = \frac{\frac{1}{T-k}\sum_{t=k+1}^{T}\left(y_t - \bar{y}\right)\left(y_{t-k} - \bar{y}\right)}{\frac{1}{T}\sum_{t=1}^{T}\left(y_t - \bar{y}\right)^2}. \tag{55}$$

- Testing significance of autocorrelation.
  - The null is about no serial correlation, i.e.,

$$\mathcal{H}_0: \quad \rho_k = 0. \tag{56}$$

  - The test statistic:

$$Z = \sqrt{T} r_k \sim \mathcal{N}(0,1). \tag{57}$$

- **The Lagrange multiplier test** allows to test jointly correlations at more than one lag.
- The AR(1) model for error term:

$$e_t = \rho_1 e_{t-1} + \nu_t \tag{58}$$

where $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$.

- Substitution into a simple regression we get:

$$y_t = \beta_0 + \beta_1 x_t + \rho_1 e_{t-1} + \nu_t. \tag{59}$$

- Substituting by the residuals we get:

$$y_t = \beta_0 + \beta_1 x_t + \rho_1 \hat{e}_{t-1} + \nu_t, \tag{60}$$

and using a fact that $y_t = \hat{\beta}_0 + \hat{\beta}_1 + \hat{e}_t$:

$$\hat{\beta}_0 + \hat{\beta}_1 + \hat{e}_t = \beta_0 + \beta_1 x_t + \rho_1 \hat{e}_{t-1} + \nu_t, \tag{61}$$

which after manipulation leads to the following **auxiliary regression in the Lagrange Multiplier (LM) test**:

$$\hat{e}_t = \gamma_0 + \gamma_1 x_t + \rho \hat{e}_{t-1} + \nu_t. \tag{62}$$

- The null is about no autocorrelation of order one, i.e.,

$$\mathcal{H}_0: \quad \rho_1 = 0. \tag{63}$$

- The test statistic:

$$LM = T \times R^2 \sim \chi^2_{(1)}. \tag{64}$$

- The AR(k) model for error term:

$$e_t = \rho_1 e_{t-1} + \rho_2 e_{t-2} + \ldots + \rho_k e_{t-k} + \nu_t. \tag{65}$$

- The auxiliary regression:

$$\hat{e}_t = \gamma_0 + \gamma_1 x_t + \rho \hat{e}_{t-1} + \hat{\rho}_2 e_{t-2} + \ldots + \hat{\rho}_k e_{t-k} + \nu_t. \tag{66}$$

- The null is about no autocorrelation up to k-th order:

$$\mathcal{H}_0 : \quad \rho_1 = \rho_2 = \ldots = \rho_k = 0. \tag{67}$$

- The test statistic:

$$LM = T \times R^2 \sim \chi^2_{(k)}. \tag{68}$$

Estimation with Serially Correlated Errors

Dealing with serial correlation

Four strategies can be considered:

1. Least squares estimation with HAC (heteroskedasticity and autocorrelation consistent) standard errors.
2. Generalized squares estimation (the Cochrane-Orcutt estimator).
3. Dynamic model (TBC on next classes)

- The variance of the least squares estimator (in the simply regression model, i.e., $y_t = \beta_0 + \beta_1 x_t + e_t$):

$$var(\hat{\beta}_1) = \sum_t w_t var(e_t) + \sum_t \sum_{t \neq s} w_t w_s cov(e_t, e_s), \tag{69}$$

where

$$w_t = \frac{(x_t - \bar{x})}{\sum_t (x_t - \bar{x})^2}. \tag{70}$$

- If there is no serial correlation then the variance

$$var(\hat{\beta}_1) = \sum_t w_t^2 var(e_t), \tag{71}$$

is very similar to the heteroskedasticity-consistent (HC) variance estimator.

- In practice, we estimate the Newey-West robust standard errors by limiting (truncating) the number of lags (the second term of the HAC). The results (i.e. standard errors) can be very sensitive the this choice.
- The common practice is to use the prewhitening of explanatory variables.
  - It allows to eliminate the persistence of explanatory variable that could be essential in constructing weights (w_t).

- AR(1) error model:

$$e_t = \rho e_{t-1} + \nu_t, \tag{72}$$

where $|\rho| < 1$, $\nu_t \sim \mathcal{N}(0, \sigma_\nu^2)$ and $cov(\nu_t, \nu_s) = 0$ for $t \neq s$.

- The mean and variance of the error term:

$$\mathbb{E}(e_t) = 0, \qquad var(e_t) = \sigma_e^2 = \frac{\sigma_\nu^2}{1-\rho}. \tag{73}$$

- The covariance and autocorrelation (of order i-th) of the error term:

$$cov(e_t, e_{t-k}) = \frac{\rho^k \sigma_\nu^2}{1-\rho^2}, \qquad \rho_i = \rho^i. \tag{74}$$

Simple regression with AR(1) errors

SGH

- When the error term follows AR(1) then the simple regression can be expressed:

$$y_t = \beta_0 + \beta_1 x_t + \rho e_{t-1} + \nu_t. \tag{75}$$

- For the period $t-1$ the error term can be expressed as:

$$e_{t-1} = y_{t-1} - \beta_0 - \beta_1 x_{t-1}. \tag{76}$$

- Combining above facts we get:

$$y_t = \beta_0 (1-\rho) + \beta_1 x_t + \rho y_{t-1} - \rho\beta_2 x_{t-1} + \nu_t. \tag{77}$$

The Cochrane-Orcutt estimator

- Alternatively, we can use the Cochrane-Orcutt estimator.
- This is a special case of GLS transformation, i.e.,

$$z_t^* = z_t - \rho z_{t-1}, \tag{78}$$

where $z \in \{y_t, x_t, e_t\}$.
- This transformation is called *quasi-differencing*.
- To get estimates of ρ we can use sample correlation of residuals.
- By construction the error term is not autocorrelated:

$$e_t^* = e_t - \rho e_{t-1} = \rho e_{t-1} + \nu_t - \rho e_{t-1} = \nu_t. \tag{79}$$