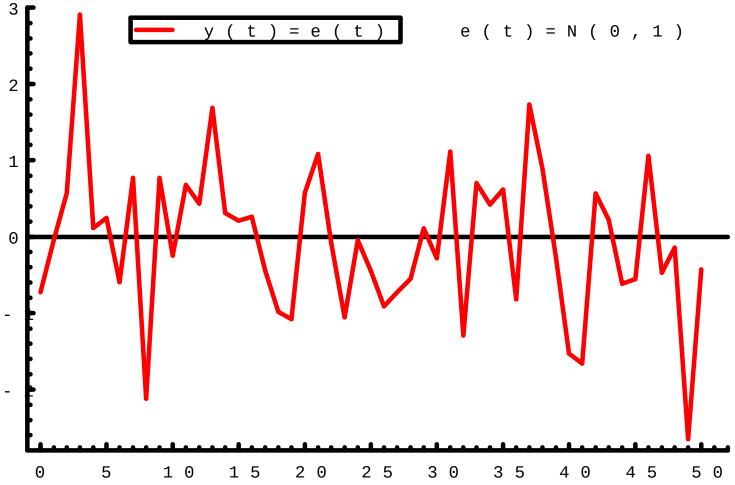
Econometrics

Jerzy Mycielski

2013

Autocorrelation and heteroskedasticity

- White noise



- Assume that the following is true

$$\operatorname{Var}(\boldsymbol{\varepsilon}) = \cdot = \sigma^2 \mathbf{V}$$

but other assumptions of CLRM are fullfiled

- We say that in the we have in our model:
  - aitocorrelation if

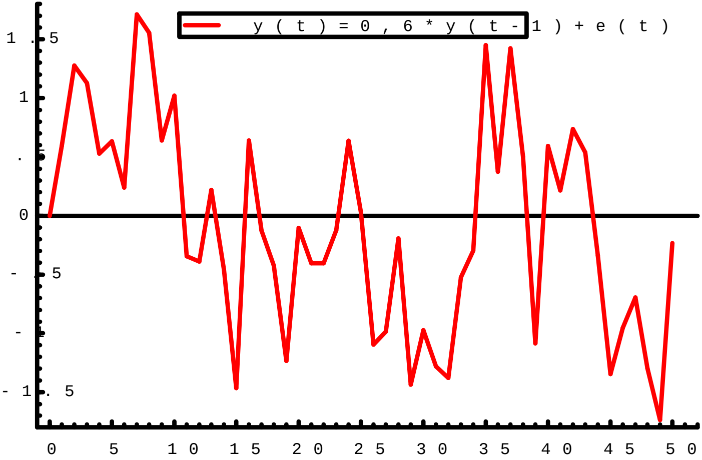
$$\operatorname{Cov}(\varepsilon_i, \varepsilon_j) = \operatorname{E}(\varepsilon_i \varepsilon_{j-s}) \neq 0 \quad \text{for some} \quad s \in 1, 2, \ldots$$

  - heteroscedasticity if

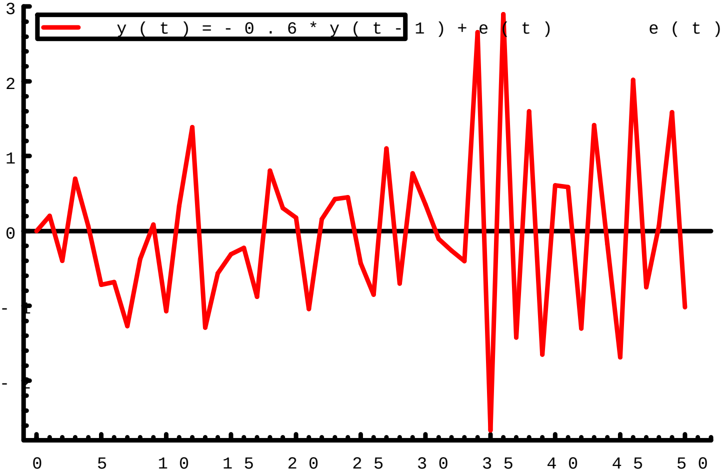
$$\operatorname{Var}(\varepsilon_i) = \operatorname{E}(\varepsilon_i^2) \neq \text{const}$$

- We are saying that random elements are nonspherical if we have autocorrelation or heteroscedasticity in the model

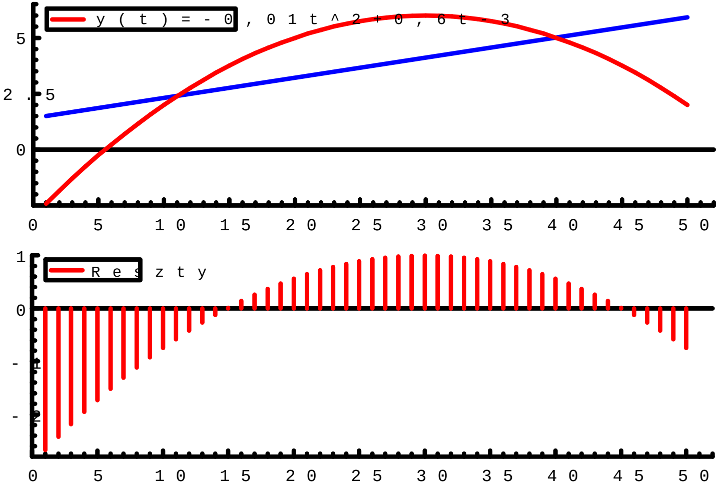
Positive autocorrelation (change of sign is too rare)



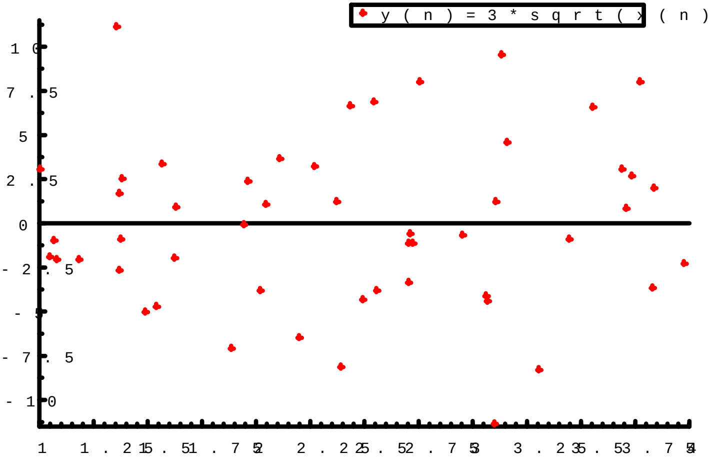
Negative autocorrelation (change of sign is too often)



- Residuals from regressions of t on t^2 - strong autocorrelation!



- Residuals from model with heterskedasticitz depending on x_i



Example

Model explaining the logarithm of expenditure for food(*lq*) in employee families w ith 2 childrens , explanatory variables - logarithm of income (*linc*), class of residence (*klm*):

```
      Source |       SS       df       MS              Number of obs =    3346
-------------+------------------------------           F(  6,  3339) =  181.40
       Model |  89.6721174     6  14.9453529           Prob > F      =  0.0000
    Residual |   275.09173  3339   .08238746           R-squared     =  0.2458
-------------+------------------------------           Adj R-squared =  0.2445
       Total |  364.763848  3345  .109047488           Root MSE      =  .28703

------------------------------------------------------------------------------
          lq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        linc |   .3540467   .0113581    31.17   0.000     .3317772    .3763162
     _Iklm_2 |  -.0334229   .0197196    -1.69   0.090    -.0720866    .0052408
     _Iklm_3 |  -.0584767   .0213409    -2.74   0.006    -.1003194   -.0166341
     _Iklm_4 |  -.0325534   .0176063    -1.85   0.065    -.0670736    .0019668
     _Iklm_5 |  -.0423542    .019028    -2.23   0.026    -.0796619   -.0050465
     _Iklm_6 |  -.0203535   .0189354    -1.07   0.283    -.0574796    .0167726
       _cons |   3.749705    .090792    41.30   0.000     3.571692    3.927719
------------------------------------------------------------------------------
```

- Breusch-Pagan test result for *klm*

```
chi2(5)      =    14.45
Prob > chi2  =   0.0130
```

- Regression of e_i^2 on *klm*

resid2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Iklm_2	.0068093	.0091102	0.75	0.455	-.0110528	.0246715
_Iklm_3	-.016938	.0098739	-1.72	0.086	-.0362975	.0024215
_Iklm_4	-.0129282	.0080402	-1.61	0.108	-.0286926	.0028361
_Iklm_5	-.0088624	.0086485	-1.02	0.306	-.0258194	.0080945
_Iklm_6	-.0133894	.0084823	-1.58	0.115	-.0300205	.0032416
_cons	.0906276	.0066867	13.55	0.000	.0775171	.1037381

- Joint test for significance of *klm*

```
F(  5,  3340) =    2.19
     Prob > F =    0.0526
```

- Regression e_i^2 on *linc*

resid2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
linc	.017435	.005053	3.45	0.001	.0075278	.0273423
_cons	-.0504988	.0385321	-1.31	0.190	-.1260476	.02505

- Test for significance of *linc*

```
F(  1,  3344) =   11.91
      Prob > F =    0.0006
```

- Conclusion: heteroscedasticity is present and depends on household income

Weighted Least Squares

Example

Assume that heteroskedasticity is of the form $\sigma_i^2 = \sigma^2 z_i^2$, where z_i is some known variable. In this case $y_i^* = \frac{y_i}{z_i}$ i $x_{s,i}^* = \frac{x_{s,i}}{z_i}$

Example

Consumption function

$$C_i = a + bY_i + \varepsilon_i$$

where C_i consumption in country i, Y_i is GDP country i.

- Precision of estimates can be measured with ratio $\frac{\sigma}{y_i}$ Value of goods consumed in Lithuania is about 1000 times smaller then consumption in USA.
- If estimated relation $\frac{\sigma}{C_i}$ would be equal 1 for Lithuania then for USA this proportion is equal to $0,001$
- Homoscedasticity assumption in this case imply, that precision measured with $\frac{\sigma}{C_i}$ is much higher forr large countries than for small countries.

Example

cont.

- More realistic is a model on variables expressed in per capita terms

$$C_i^* = a\frac{1}{z_i} + bY_i^* + \varepsilon_i^*$$

  where $C_i^* = \frac{C_i}{z_i}$, $Y_i^* = \frac{Y_i}{z_i}$ and z_i is the population of a given country
- This model is equivalent to *GLS* if standard errors are proportional to population
- Notice that constant term is transformed into $\frac{1}{z_i}$

Example

Spendings for food. We assumed that the variance of error term is given by $\sigma_i^2 = \alpha_0 + \alpha linc$. From regression of lq on cosntant and linc we obtained e_i. Next e_i^2 was regressed on *linc*.

resid2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
linc	.017435	.005053	3.45	0.001	.0075278	.0273423
_cons	-.0504988	.0385321	-1.31	0.190	-.1260476	.02505

- From this regression we obtain fitted values $\widehat{\sigma}_i = \sqrt{\widehat{\alpha}_0 + \widehat{\alpha}_1 linc}$
- Dependent variable and all independent variables were divided by $\widehat{\sigma}_i$

- Regression results for transformed variables:

Source	SS	df	MS
Model	1682677.41	7	240382.487
Residual	3345.60972	3339	1.00197955
Total	1686023.02	3346	503.892116

Number of obs	3346
F(7, 3339)	.
Prob > F	0.0000
R-squared	0.9980
Adj R-squared	0.9980
Root MSE	1.001

t_lq	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t_linc	.3603508	.0114135	31.57	0.000	.3379727	.3827289
t_Iklm_2	-.0330761	.0199833	-1.66	0.098	-.0722569	.0061048
t_Iklm_3	-.0565737	.0216039	-2.62	0.009	-.0989319	-.0142156
t_Iklm_4	-.0290265	.0178167	-1.63	0.103	-.0639593	.0059063
t_Iklm_5	-.0374901	.0191635	-1.96	0.051	-.0750634	.0000832
t_Iklm_6	-.017306	.0190144	-0.91	0.363	-.054587	.0199751
t_cons	3.699076	.0906672	40.80	0.000	3.521307	3.876845

- Estimate of s^2 should be equal to 1.
- Standard value of F jest is invalid (transformed constatnt treated as one of the explanatory variables). Correct joint test should test significance of all variables apart from transformed constant

```
F(  6,  3339) =  183.74
      Prob > F =    0.0000
```

- R^2 from this regression cannot be intepreted - dependent variable is artificial
- Correct R^2 should be calculated for original dependent variable and fitted values form GLS dopasowanych

$$R^2_{GLS} = \frac{\sum \left(\widehat{y}_{GLS,i} - \overline{\widehat{y}}_{GLS} \right)^2}{\sum \left(y_i - \overline{y} \right)^2} = \rho^2_{\widehat{y},y}$$

where $\widehat{y}_{GLS} = \mathbf{x}_i \boldsymbol{\beta}_{GLS}$.

- Calculated value $R^2_{GLS} = .28674733$
- It is a descriptive statistics cannot be used to compare OLS and GLS

- Result of the Breusch-Pagan test:

```
Breusch-Pagan LM statistic:  .0090563  Chi-sq( 1)  P-value
=   .9242
```

- Heteroskedastisicty is eliminated!
- Notice small differences in estimates of parameters and standard errors between *OLS* and *GLS*

Example

Expenditure for food cd. Results with robust variance matrix:

```
Regression with robust standard errors               Number of obs =    3346
                                                     F(  6,  3339) =  174.56
                                                     Prob > F      =  0.0000
                                                     R-squared     =  0.2458
                                                     Root MSE      =  .28703

------------------------------------------------------------------------------
             |               Robust
          lq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        linc |   .3540467    .011677    30.32   0.000     .3311518    .3769415
     _Iklm_2 |  -.0334229   .0210545    -1.59   0.113     -.074704    .0078582
     _Iklm_3 |  -.0584767   .0213145    -2.74   0.006    -.1002676   -.0166859
     _Iklm_4 |  -.0325534   .0181496    -1.79   0.073    -.0681388     .003032
     _Iklm_5 |  -.0423542   .0196662    -2.15   0.031    -.0809132   -.0037951
     _Iklm_6 |  -.0203535   .0194412    -1.05   0.295    -.0584714    .0177644
       _cons |   3.749705   .0933026    40.19   0.000     3.566769    3.932641
------------------------------------------------------------------------------
```