STATISTICS IN TRANSITION new series, September 2018
Vol. 19, No. 3, pp. 433–452, DOI 10.21307/stattrans-2018-024

DEALING WITH HETEROSKEDASTICITY WITHIN THE MODELING OF THE QUALITY OF LIFE OF OLDER PEOPLE

Katarzyna Jabłońska[1]

ABSTRACT

Using the estimation method of ordinary least squares leads to unreliable results in the case of heteroskedastic linear regression model. Other estimation methods are described, including weighted least squares, division of the sample and heteroskedasticity-consistent covariance matrix estimators, all of which can give estimators with better properties than ordinary least squares. The methods are presented giving the example of modelling quality of life of older people, based on a data set from the first wave of the COURAGE – Poland study. The comparison of estimators and their practical application may teach how to choose methodologically the most appropriate estimation tool after detection of heteroscedasticity.

Key words: heteroskedasticity, linear regression, HC-estimators, quality of life.

1. Introduction

Homogeneity of error variance, called homoskedasticity, is one of the main assumptions of linear regression. Many models, especially based on cross-sectional data, do not satisfy it (Greene, 2012, p.297). Such a situation is called heteroskedasticity. Then, parameters estimation with ordinary least squares method (OLS) does not give optimal results. There are many alternative methods which are either resistant to disturbance of homoskedasticity or they transform a model into a new one, which is henceforth homoskedastic.

Our aim is to discuss methods of parameters estimation in linear regression models in the case of heteroskedasticity and to focus on their strengths, weaknesses and important properties of obtained estimators. While weighted least squares method and the division of the sample are well known, HC-estimators are not commonly used, which may be surprising in light of the fact that new ones are still being created, improving previous ones. A comparison of those methods can be valuable for professional sociologists and practitioners to help them choose an appropriate estimation tool in the occurrence of heteroscedasticity in linear regression model.

[1] Department of Medical Sociology, Chair of Epidemiology and Preventive Medicine, Jagiellonian University Medical College, Kopernika 7a, 31-034 Krakow, Poland. Phone number: (+48) 12 423 1003. E-mail address: k.jablonska@cm-uj.krakow.pl.

Methodological considerations will be presented giving the example of modelling of quality of life of older people depending on psychosocial, demographical and other factors, based on data from the first wave of the COURAGE – Poland population-based study from 2011. Due to the population aging, the group of older people has been recently a subject of great interest and its analysis can be crucial in understanding how quality of life is affected by, in particular, psychosocial factors in an old age. We investigate a group without chronic diseases to detect what accompanies healthy aging. Analyses are done separately for a group of men for whom regression model is heteroskedastic and for a group of women for whom regression model is homoskedastic. The division into gender is justified because there are some significant gender-related differences in effects of psychosocial factors on older people's quality of life, as shown by Tobiasz-Adamczyk (et al., 2017).

In Section 2 properties of estimators from the ordinary least squares method are recalled. Alternative methods of parameters estimation are presented in Sections 3-6. The next section contains an empirical example of older people's quality of life models using previously described methods, separately for men and women. Results for both genders are discussed in Subsection 7.5. Section 8 includes conclusions and indications as to the proper choice of the method of estimation in the case of heteroskedasticity.

2. Linear regression model and the method of least squares estimation

In a classic linear regression model we have $Y = \beta X + \varepsilon$. Given a vector $Y = (y_1, \dots, y_n)$ of n-observations, called dependent variable, and a matrix $X = (x_1, .., x_p)$ with p-independent variables, where $\forall_{j=1,\dots,p}: x_j = (x_{j1}, \dots, x_{jn})$, we want to find the value of an unknown parameter $\beta = (b_0, b_1, .., b_p)$ to be able to predict values of Y with a random error ε, called residual.

Unknown parameters must be estimated, which means approximated in a sufficiently good way. Fortunately, there exist some objective measures of such sufficiency goodness: consistency, unbiasedness and effectiveness (the last is considered in a specified class of estimators). Consistency is a stochastic convergence to the estimated parameter; unbiasedness means that its expected value is equal to the value of the parameter which is estimated; effectiveness characterizes an estimator with the least variance within the specified class of estimators.

A classic way to estimate parameters is to use the method of ordinary least squares (OLS), where the value of the sum of squares of errors $(\sum_{i=1}^{n} \varepsilon_i^2)$ is minimized. To make a model and its estimation purposeful, a number of assumptions are required. The first is homogeneity of error variance, called homoskedasticity. If, in addition, a model has a correct linear structure, a matrix X of fixed independent variables has rank p, the size of a sample is greater than the number of all parameters (n>p+1), random errors have mean zero and they are uncorrelated, then, on the basis of Gauss-Markov theorem (Dodge, 2008, p. 217-218), the OLS estimator will be linear, unbiased and effective among all linear and unbiased estimators. It will also be consistent (Verbeek, 2004). This estimator is expressed by the formula $\hat{\beta} = (X^T X)^{-1} X^T Y$. By the same theorem, the estimator of covariance matrix of the examined parameter, expressed by $\hat{V}(\hat{\beta}) = s^2 (X^T X)^{-1}$, where $s^2 = \frac{1}{n-(p+1)} \sum_{i=1}^{n} \varepsilon_i^2$, will also be unbiased. It gives important information about approximated standard errors of components of $\hat{\beta}$. If we additionally assume that errors are normally distributed, then significance tests (F test and Student's t-tests) will be possible to conduct. On this basis, one can determine which elements of vector $\hat{\beta}$, thereby, which independent variables, have significant relationship with variable Y.

If the assumption of homoskedasticity is violated, then we are talking about heteroskedasticity. To detect it, diagnostic tests should be conducted, for example three most popular: Breusch-Pagan test (Breusch and Pagan, 1979), which tests hypothesis that the error variance is linearly dependent with variables from the model; White's test (White, 1980), which finds out whether error variance is constant or Goldfeld-Quandt test (Goldfeld and Quandt, 1965), which checks whether heteroskedasticity is due to the one specified variable. Heteroskedasticity can also be detected with the help of OLS regression plots: errors and squared errors against predicted values, as well as errors against independent variables.

If we conclude that the analysed model is heteroskedastic, then OLS estimator of β is still consistent and unbiased, but no longer effective (Verbeek, 2004). Also, covariance matrix estimator $\hat{V}(\hat{\beta})$ is biased and inconsistent, and there is a problem with conducting statistical tests of the significance of parameters, because test statistics do not have required distributions (Verbeek, 2004). OLS estimation becomes unfounded in the case of heteroskedasticity, because there is a risk of both incorrect parameters approximation and untrustworthy tests results. Therefore, other methods of estimation should be used.

3. Weighted least squares method

A model with heteroskedasticity differs from a classic one in that consecutive observations have distinct values of error variance, that is $Var(\varepsilon) = \sigma^2 \Omega$ with different positive numbers $w_1, \dots, w_n$ (called weights) on the main diagonal of matrix Ω. Then, the sum $\sum_{i=1}^{n} \varepsilon_i^2 w_i$ is to be minimized. If Ω is known, estimator $\hat{\beta} = (X^T \Omega^{-1} X)^{-1} X^T (\Omega)^{-1} Y$ is effective among its unbiased estimators (Verbeek, 2004), covariance matrix for $\hat{\beta}$ equals

$$V(\hat{\beta}) = \sigma^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}$$

and its unbiased estimator is

$$\hat{V}(\hat{\beta}) = s^2 (X^T X)^{-1} X^T \Omega X (X^T X)^{-1}.$$

A method which uses weights is called weighted least squares method (WLS).

Knowledge of a matrix Ω is an unrealistic assumption and its values can only be approximated. We can use the model $\varepsilon^2 = \alpha X + c$ or eventually modify it to receive predicted values of ε^2, which we use as the diagonal of $\hat{\Omega}$. Then we can complete the main model of Y with weights equal to reciprocals of elements from the diagonal of $\widehat{\Omega}$. WLS estimator with the weights estimated before can be asymptotically more effective than classic OLS (Davidian, Carroll, 1987), assuming we used a proper and well-fitted model to predict ε^2. Otherwise, there is a big risk that the new model will still be heteroskedastic. It is one of the biggest WLS disadvantages, but this method has one important strength. It helps to detect presumptive cause of heteroskedasticity in a model. Assuming error squares regression was analysed and a variable significantly dependent on ε^2 was found, we can suspect that we detected the reason of the problem. It leads us to the next method of dealing with lack of homoskedasticity.

4. Division into subsamples

Having a variable significantly dependent on error squares, we can try to divide the sample into subsamples which depend on its values. It is obvious that we look for a division such that error variances among both subsamples are constant. Similar approach was considered by Goldfeld and Quandt (1965), and yield a test of heteroskedasticity based on the assumption that variances heterogeneity is due to the one specified variable.

It is crucial to select the division with a strong theoretical justification. Only then our original aim, which is drawing conclusions about the whole population based on its randomly selected part, will be preserved. It is much easier to isolate subsamples relying on a factorial variable than on a continuous one (in the latter case we have to arbitrary impose cut-off points). Other good idea is to divide the sample based on factorial variables like: gender, marital status, group of age and so on. Such divisions are almost always justified, but we should not forget to check homoskedasticity of new subsample models – the division had no sense without its occurrence.

5. Heteroskedasticity-consistent covariance matrix estimators

There is another estimation method which can be applied without any assumption about the error variance – heteroskedasticity-consistent covariance matrix estimators (HC-estimators). This method uses OLS to estimate β and one of HC-estimators to estimate its covariance matrix (then, standard errors) and to conduct tests. The main purpose of their use is to minimize the violation of inference caused by heteroskedasticity.

The first HC-estimator was HC0 proposed by White (1980), called a Sandwich estimator. Next were: HC1 by Hinkley (1977), HC2 by MacKinnon and White (1985), HC3 by Efron (1982), HC4 by Cribari-Neto (2004), HC5 by Cribari–Neto, Souza and Vasconcellos (2007), HC4m by Cribari-Neto and da Silva (2011) and there are still being created new ones, like the newest HC5m by Li et al. (2017), each improving previous ones. Formulas of all eight are given below.

$$\mathrm{HC0} = (X^T X)^{-1} \begin{pmatrix} \hat{\varepsilon}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\varepsilon}_n^2 \end{pmatrix} (X^T X)^{-1},$$

$$\mathrm{HC1} = \frac{n}{n-p-1}(X^T X)^{-1}\begin{pmatrix} \hat{\varepsilon}_1^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{\varepsilon}_n^2 \end{pmatrix}(X^T X)^{-1},$$

$HC2$ to $HC5m$ are equal to $(X^T X)^{-1}\begin{pmatrix} \frac{\hat{\varepsilon}_1^2}{(1-h_{11})^{\delta_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{\hat{\varepsilon}_n^2}{(1-h_{nn})^{\delta_n}} \end{pmatrix}(X^T X)^{-1}$ with different values of δ_i ($i = 1, \dots, n$) for each:

- HC2: $\delta_i = 1$,
- HC3: $\delta_i = 2$,
- HC4: $\delta_i = \min\left\{\frac{nh_{ii}}{p+1}, 4\right\}$,
- HC4m: $\delta_i = \min\left\{\frac{nh_{ii}}{p+1}, \gamma_1\right\} + \min\left\{\frac{nh_{ii}}{p+1}, \gamma_2\right\}$,
- HC5: $\delta_i = \min\left\{\frac{nh_{ii}}{p+1}, \max\left\{4, \frac{n \cdot k \cdot h_{max}}{p+1}\right\}\right\}$,
- HC5m: $\delta_i = k_1 \min\left\{\frac{h_{ii}}{\bar{h}}, \gamma_1\right\} + k_2 \min\left\{\frac{h_{ii}}{\bar{h}}, \gamma_2\right\} + k_3 \min\left\{\frac{h_{ii}}{\bar{h}}, \max\{4, \frac{kh_{max}}{\bar{h}}\}\right\}$,

where:

$\hat{\varepsilon}_i = y_i - \widehat{y_i} = y_i - x_i\widehat{\beta_i}$,

h_{ii} is i-th element of a matrix $H = X(X^T X)^{-1}X^T$ (i-th leverage),

$h_{max} = \max\{h_{ii} \mid i \in \{1, .., n\}\}$,

$\bar{h} = \frac{\sum_{i=1}^{n} h_{ii}}{n}$,

$k \in [0,1]$ (Cribari-Neto, Souza and Vasconcellos (2007) recommend k=0.7),

$\gamma_1, \gamma_2 > 0$ (Cribari-Neto and da Silva (2011) recommend $\gamma_1 = 1$, $\gamma_2 = 1.5$),

$k_1, k_2, k_3 \geq 0$ (Li (et al., 2017) recommend $k_1 = 1, k_2 = 0, k_3 = 1$).

The fact of the existence of so many HC-estimators inclines to make a comparison of their strengths and weaknesses, which will help to choose best estimators in specific situations.

HC1, in comparison with HC0, is corrected for degrees of freedom, whereas HC2 takes into account values of leverage, which has been improved by HC3 and later. All HC estimators do not demand homoskedasticity and are asymptotically consistent, nonetheless they have their own disadvantages. It is common for HC0 to become severely biased as mentioned by Cribari-Neto, Ferrari and Cordeiro (2000), especially in the case of a small sample (Long and Ervin, 2000) or occurrence of many high-leverage observations (Chesher and Jewitt, 1987).

Moreover, t-tests with HC0-estimator are liberal, which means that it is easy for them to achieve significant results, and similar can be said about HC1 and HC2.

In the case of a small sample, n<250, Long and Ervin (2000) recommend the use of HC3. For large samples HC0, HC1 and HC2 estimators should behave almost the same as HC3, but in the case of occurrence of high-leverage observations they become more biased than HC3. As it was shown by Cribari-Neto and Zarkos (2001), high-leverage observations can have even bigger influence on tests conservation, hence, their reliability, and estimator properties (primarily, their bias) than severity of heteroskedasticity. That is why HC4-estimator was proposed and, as it was shown by Cribari-Neto (2004), it has an advantage over HC3 in the case of many high-leverage observations.

The first HC4 modification, called HC5-estimator, was presented by Cribari–Neto, Souza and Vasconcellos (2007) and their innovation was taking into account the maximal leverage instead of only individual ones. Numerical evaluations showed that HC5-based inferences are much more reliable than HC3 and HC4-based: they are less size-distorted and tests are less liberal. HC5 can be crucial when observations are very strongly leveraged.

Another approach to HC4 modification was HC4m-estimator presented by Cribari-Neto and da Silva (2011). It improves the squared residuals discounting dependently on values of leverage: a heavier discount for low leverage observations and inversely for high leverage ones. HC4m is also a better alternative for HC4 in the case of non-normal errors (Cribari-Neto and da Silva, 2011). HC4m tries to fix HC4 and HC5 weaknesses in the case of a low degree of leverage, but it is worse than them when the high degree of leverage occurs.

The most recently presented HC5m-estimator by Li (et al., 2017), combines strengths of HC4m for low degree of leverages and HC5 for high degree of leverages. Simulations performed by Li (et al., 2017) showed that HC5m-based tests are reliable at points both with low or with high leverages and they have the smallest size distortions among tests based on all of HC3, HC4, HC4m and HC5.

It is worth pointing out that in the case of homoskedasticity HC-estimators can have worse properties than OLS estimator – almost all are biased then (Kauerman and Carroll, 2001). However, only HC2 is an unbiased estimator for homoskedastic data (Hayes and Cai, 2008), which gives it a supremacy in a situation when it is hard to confirm with certainty that there is a lack of variance homogeneity.

The last very important thing is that newer HC-estimators make significance tests less and less liberal. It means that a significant result received with a newer estimator is much more reliable than the one received with the older one, however, it is hard itself to achieve a significant result with the help of newest estimators. When there is no proven need to use a more conservative estimator, researchers can choose a bit less conservative one, provided that the model was analysed in detail and its use is fully justified. The newest does not always mean the best – it depends on model's properties.

6. Other methods

There are many other methods applied for heteroskedastic data. One of the most frequently mentioned in the literature is connected with transforming variables (Carroll and Ruppert, 1984), (Box and Cox, 1964), however, it can be very problematic in regard to interpreting its results (Sakia, 1992).

The other option is the general method of moments – GMM (Cragg, 1983), but as it was shown in (Kiviet and Feng, 2015), it has huge defects, and some modifications are proposed. Other ideas are General Linear Models, Penalized Least Squares Method (Wagener and Dette, 2012) or Residual Maximum Likelihood Estimation (Smyth, 2002).

The latest ideas concern Bayesian regression (Startz, 2017) and Generalized Least Squares based on machine-learning (Miller and Startz, 2017). They still need a deeper exploration, but give hope to accurate estimation without regard to the sample size and the values of leverages.

7. Empirical example of quality of life model

7.1. Statistical analysis

Linear regression will be used to model the assessment of quality of life of people aged 60 or over, who have not been diagnosed with any serious disease, including depression and chronic diseases (angina, arthritis, asthma, POCHP, diabetes and stroke), depending on: age, body mass index (BMI), assessment of activities of daily living (ADL) on Katz's scale described by Wieczorowska-Tobis and Talarska (2010), social network (Zawisza, Gałaś and Tobiasz-Adamczyk, 2014), loneliness (Hughes et al., 2004), social support (Dalgard, 1996) and two types of participation: relations with other people and activity in a local community. Models are additionally adjusted into education level, marital status and having children.

The data come from the first wave of the COURAGE - Poland population-based study from 2011. Values of quality of life are based on the Polish version of WHOQOL-AGE scale (Caballero et al., 2013; Zawisza, Gałaś and Tobiasz-Adamczyk, 2016), ranged from 0 to 100 points, and higher score of WHOQOL-AGE is interpreted as better health-related quality of life. Also loneliness, social support and social network range from 0 to 100 points. Higher score of ADL assessment means that more problems with daily living activities were reported.

We firstly consider a model for men. At the beginning, OLS estimation is conducted, but after detection of heteroskedasticity other methods are used. Finally, the most proper estimator is chosen. The same model for group of women is analysed with OLS and HC2-estimator, because both have good properties in the case of homoskedasticity, which was observed in this model. Then, a comparison of results for women and men is presented.

Analyses are conducted with SAS 9.4. To calculate standard errors and p values with HC4-estimator, we use a macro created by Hayes and Cai (2007) given in (Hayes and Cai, 2007, Appendix), while for HC4m, HC5 and HC5m we created new SAS macros, just like we did to conduct a Goldfeld-Quandt test. The adopted level of significance is 0.05.

7.2. Men's quality of life model

We conducted OLS estimation for the linear regression model of older men's quality of life with the sample size: n=366. Results are given in Table 1. Five psychosocial variables (ADL, social network, loneliness, participation - relations and social support) turned out to be significant with p values below 0.05. The model meets assumptions of errors normality with mean equals 0, significant linear structure of the model, no autocorrelation of errors, no correlation of independent variables, but it has a problem with homoskedasticity. Results of White's test (p=**0.001**) and Breusch-Pagan test (p=**0.009**) show that there is possibly a relationship between error variance and one or more independent variables.

Table 1. Results of OLS estimation of parameters from men's quality of life model, adjusted into: education level, marital status, having children

	OLS ($R^2=0.283$, $p<0.001$)		
	$\hat{\beta}$	Std. error	P
BMI	-0.095	0.133	.478
Age	-0.057	0.073	.434
ADL	-1.169	0.286	<.001
Social network	0.189	0.054	.001
Loneliness	-0.114	0.032	<.001
Participation - local community	0.580	0.765	.449
Participation - relations	1.654	0.777	.034
Social support	0.150	0.041	<.001

The scatter plot of predicted values of quality of life against error squares with a regression line (Figure 1) confirms that presumably error variance is not constant and seems to decrease with an increase in predicted quality of life. We need the other method of estimation.

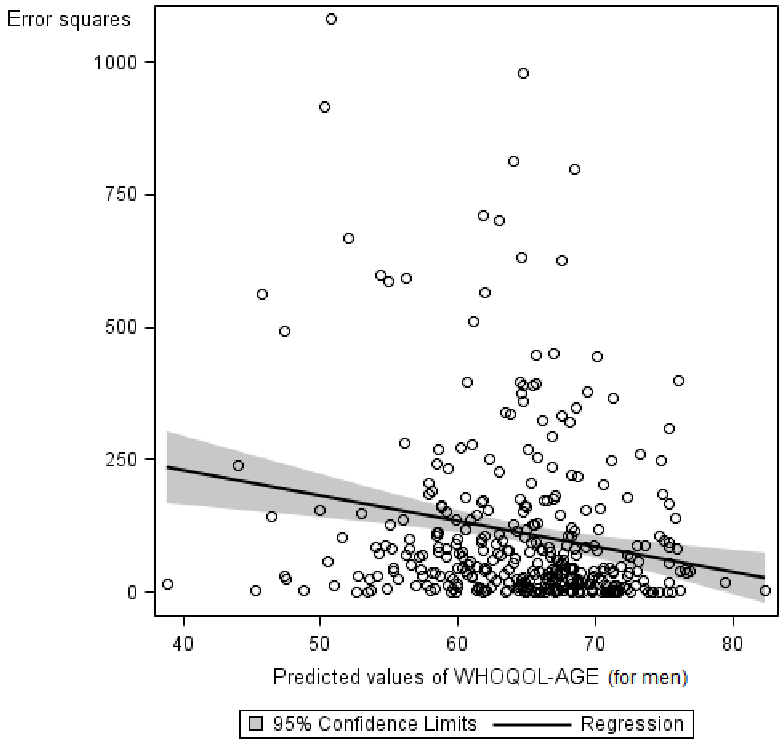

Figure 1. Regression line of error squares against predicted quality of life (WHOQOL-AGE scale) for men

Error squares regression with stepwise selection was conducted with all independent variables used in the quality of life model. It helped to choose their best linear model (Table 2), which turned out to be the one with social support (p=0.001) and loneliness (p=0.067). Their additive effect makes the whole model significant (R^2=0.044; p<0.001), but it has many failures, like non-normally distributed errors with non-zero mean.

Table 2. Results of the OLS estimation from the error squares model

	$\hat{\beta}$	Std. error	P
Social support	-1.950	0.566	.001
Loneliness	0.826	0.449	.067

One unit higher feeling of social support ($\hat{\beta} = -1.950$; $p = 0.001$) is associated with almost two units lower error variance and this variable can be suspected to cause heteroskedasticity in the quality of life model. Indeed, error variance is much higher for men who have the lowest level of social support (below its first quartile) compared to groups with higher levels (Table 3). Unfortunately, after adding weights to the basic quality of life model and conducting WLS estimation, a new model is still heteroskedastic. However, a strong relationship between error squares and social support could be seen and a natural idea is to divide the sample into subsamples relying on values of social support. We propose to create a dichotomous variable with value 1 if social support is below or equal to its first quartile (<=54.55) and value 0 for the other case.

Table 3. Error variances for social support quartile-based levels from quality of life OLS model

Social support	<=54.55 (Q1)	<=63.64 (Q2)	<=72.73 (Q3)	<=100
N	121	91	75	73
Error variance	144.12	102.51	89.65	83.99

To see whether such a division has a chance to be proper, let us conduct a Goldfeld-Quandt test. A test statistics is of the form

$$F[n_2 - (p+1), n_1 - (p+1)] = \frac{\frac{s_2}{n_2-(p+1)}}{\frac{s_1}{n_1-(p+1)}},$$

where p is the number of independent variables, n_i is an i-th subsample size, s_i^2 – i-th subsample sum of squares, where i=1 is for a subsample with relatively small error variance (higher social support in our case) and i=2 is for a subsample with relatively large error variance (lower social support). It has a chi-square distribution when the equality of variance of both subsample hypotheses is fulfilled, as followed by Goldfeld and Quandt (1965). In our case p=**0.010,** which means that error variances are significantly different among both subsamples and the division is reasonable. Quality of life model will be study separately for them.

For men with a low social support, Breusch-Pagan test indicates it still has a problem with heteroskedasticity (White: p=0.294; Breusch-Pagan: p=**0.043**) and the same can be told about White's test for the group with a higher social support (White: p=**0.019**; Breusch-Pagan: p=0.218). The division is not appropriate and we should try another one or consider a different method of estimation.

Results of estimation with HC-estimators for men's quality of life model are given in Table 4. The same five psychosocial variables as with OLS method turned out to be significant after estimations with HC0, HC1, HC2, HC3, HC4 and HC5. Standard errors are usually greater for HC than for OLS and the same can be told about p values. Participation – relations was just below 0.05 with HC3, HC4 and HC5, but crossed this line after HC4m and HC5m estimations. For HC5m also ADL assessment was not significant (p=0.067). HC4 and HC5 gave almost exactly the same results, which clearly indicates that the degree of leverage must not be very high in our model.

Table 4. Results of HC estimation of standard errors with p values from men's quality of life models, adjusted into: education level, marital status, having children

	HC0		*HC1*		*HC2*		*HC3*		*HC4*		*HC5*		*HC4m*		*HC5m*	
	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P	Std. error	P
BMI	0.133	.478	0.135	.485	0.137	.490	0.141	.503	0.144	.511	0.144	.511	0.143	.507	0.162	.560
Age	0.075	.448	0.076	.455	0.077	.458	0.078	.468	0.079	.469	0.079	.469	0.079	.471	0.085	.503
ADL	0.361	.001	0.367	.002	0.384	.003	0.409	.005	0.461	.012	0.461	.012	0.422	.006	0.636	.067
Social network	0.062	.003	0.063	.003	0.064	.004	0.066	.005	0.068	.006	0.068	.006	0.067	.005	0.076	.014
Loneliness	0.033	.001	0.034	.001	0.034	.001	0.035	.001	0.035	.001	0.035	.001	0.035	.001	0.039	.003
Participation-local community	0.682	.396	0.694	.404	0.697	.406	0.712	.415	0.708	.413	0.708	.413	0.714	.417	0.741	.434
Participation-relations	0.800	.039	0.813	.043	0.818	.044	0.837	.049	0.834	.048	0.834	.048	0.842	.050	0.877	.060
Social support	0.043	.001	0.044	.001	0.044	.001	0.045	.001	0.045	.001	0.044	.001	0.045	.001	0.046	.001

An observation is called an outlier if the absolute value of its studentized residual is greater than 2, whereas it is called a high-leveraged point if its leverage is greater than $\frac{2(p+1)}{n} \approx 0.0655$. Outliers and high leveraged observations are shown in Figure 2. There are 16 outliers (4%), 20 high-leverage observations (5%) and 3 points with both these features (0.8%). In our case, $h_{max} = 0.162$ and $\frac{n \cdot 0.7 \cdot h_{max}}{p+1} < 4$, which means that the degree of maximum leverage is moderate and not very high.

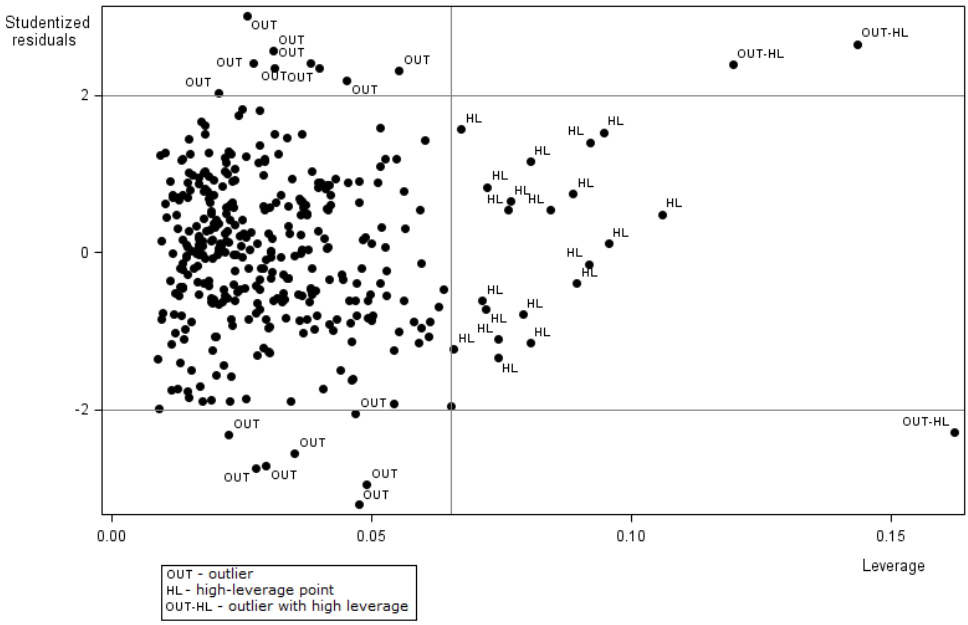

Figure 2. Scatter plot of leverages towards studentized residuals

The questions is which of HC-estimators is the most suitable in our case. We have a strong evidence of heteroskedasticity, the sample size is not very small and the maximum degree of leverage is moderate. HC4m, HC5m or even HC3 are better adapted for the situation of low or moderate leverages than HC4 and HC5, which are better in the case of very high leverages, so we can restrict to those 3 estimators. As it was outlined earlier in Section 5, the conservativeness of tests increases with every consecutive estimator. HC5m makes them severely conservative, but there is no need to use it for our model. In our opinion, the best option is to choose HC4m, which is still much more conservative than somewhat liberal HC3. As a result of use of HC4m-estimator, four variables (ADL assessment, social network, loneliness and social support) are considered significant in the context of men's quality of life.

7.3. Women's quality of life model

We will now consider the same model of quality of life for the group of women aged 60 or over (sample size: n=519), which meets all linear regression assumptions including homoskedasticity (White: p=**0.452**; Breusch-Pagan: p=**0.590**). The scatter plot of predicted values of quality of life against error squares with a regression line (Figure 3) confirms that error variance is rather constant.

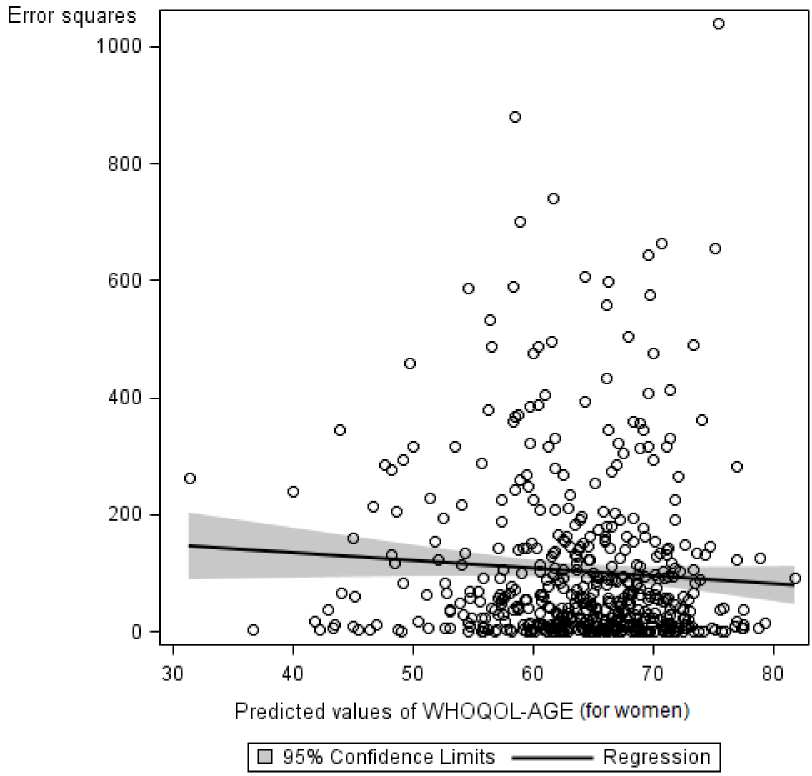

Figure 3. Regression line of error squares against predicted quality of life (WHOQOL-AGE scale) for women

Table 5 presents the results of OLS estimation of women's quality of life. There are four significant variables: ADL, social network, loneliness and social support. In the light of facts previously presented in Section 5, it might be valuable to compare OLS results with an unbiased, but a slightly more conservative HC2-estimator. Despite both OLS and HC2 variance estimators are unbiased, values of standard errors estimators and results of significance tests are not equally same, but they are very similar. This is mainly due to the fact that in both cases the variances are estimated and square roots are computed to finally get standard errors. While p value for participation–local community is equal to the level of 0.05 for HC2-estimator, it is much higher for OLS (p=0.071) and we consider this variable insignificant. Despite HC-estimators almost always increases the conservativeness of tests, it is not a set rule for all variables in a given model.

Table 5. Results of estimation with OLS and HC2-estimator of women's quality of life, adjusted into: education level, marital status and having children

	OLS *(R^2=0.343, p<0.001)*			*HC2*	
	$\hat{\beta}$	Std. error	P	Std. error	P
BMI	-0.020	0.086	.819	0.084	.813
Age	-0.114	0.063	.071	0.064	.078
ADL	-1.121	0.157	<.001	0.174	<.001
Social network	0.128	0.040	.002	0.040	.001
Loneliness	-0.076	0.023	.001	0.025	.003
Participation – local community	1.274	0.704	.071	0.648	.050
Participation - relations	0.887	0.586	.131	0.596	.137
Social support	0.112	0.031	<.001	0.032	.001

7.4. Discussion on the results from quality of life models

Men's quality of life model turned out to have a problem with heteroskedasticity. OLS estimation method was unreliable, therefore we tried methods of WLS and division into subsamples, both of which did not result in homoskedasticity. Then, HC-estimators were used and HC4m was found the most appropriate. Thanks to the estimation with HC4m four variables are considered to have a significant relationship with quality of life among older men: social network, loneliness, social support and ADL assessment.

Men who feel more lonely have worse quality of life assessment ($\hat{\beta} = -0.114;\ p = 0.001$), which means that one unit increase in the feeling of loneliness results in 0.114 unit decrease of quality of life. The more problems connected with daily living activities were reported, the significantly lower quality of life was detected ($\hat{\beta} = -1.169;\ p = 0.006$). In turn, the better the assessment of social network, the higher quality of life was reported $(\hat{\beta} = 0.189;\ p = 0.005)$, as well as for social support $(\hat{\beta} = 0.150;\ p = 0.001)$.

For women, the applicability of the OLS method was fully justified. As a result, the same four variables turned out to have a significant relationship with quality of life: ADL $(\hat{\beta} = -1.121;\ p < 0.001)$, social network $(\hat{\beta} = 0.128;\ p = 0.002)$, social support $(\hat{\beta} = 0.112;\ p < 0.001)$ and loneliness $(\hat{\beta} = -0.076;\ p = 0.001)$. Results of estimation with unbiased HC2-estimator were very similar to OLS results.

As it can be seen, the same four variables turned out to be significant in the context of quality of life both for men and women and trends for all four are very similar. Nonetheless, absolute values of all estimated parameters are higher in the case of men. Especially feeling of loneliness has 1.5 times stronger effect on quality of life among men than among women. While it has been shown before that QOL can be significantly reduced by loneliness among older people (Musich et al., 2015), there are very few studies on gender differences in older people quality of life affected by feeling of loneliness (Tobiasz-Adamczyk et al., 2017). Despite women seem to suffer from loneliness more frequently than men (Beal, 2006; I. Thomopoulou, D. Thomopoulou and Koutsouki, 2010), men's reaction to a higher feeling of loneliness, resulting in poorer quality of life assessment, can be stronger than women's. Deeper sociological inferences were not the main scope of this paper, but we hope that our results will encourage sociologists to make further analysis.

8. Conclusion

Giving the example of men's quality of life model, we could observe what are possible consequences of ignoring heteroskedasticity. We could also investigate how to choose the best alternative method of estimation.

Following Hayes and Cai (2007), if there is rationale for stating that our model does not meet the homoskedasticity assumption, it is recommended to have a very critical view on the results obtained by the OLS and to use other estimation methods. However, not all of them work equally well. WLS requires the form of heteroskedasticity to be known, which is usually difficult to meet and we are not assured that a new model will be free of lack of homoskedasticity. The latter can also be said about the division of a sample into subsamples.

It seems that the best idea is to use heteroskedasticity-consistent variance matrix estimators. Some HC-estimators are offered by statistical programs, for example SAS (HC0-HC3) and R (it offers also HC4m and HC5 in the *Sandwich* package presented in 2017).

Despite HC-estimators asymptotically have desirable properties, some problems with their credibility can occur. Generally, when it comes to choose the best HC-estimator, we recommend the following criteria:

- HC2 is the best option when heteroskedasticity is not clear or it is relatively low.
- If the sample size is small, HC3 and later estimators are preferred. The degree of leverages should be taken into account to choose the best among them.
- In the case of the occurrence of some high leveraged observations, HC3, HC4, HC5 or HC5m are recommended. If the degree of leverage is very high, HC5 and HC5m prevail over others.
- If the degree of leverage is low or moderate $\left(\frac{0.7 \cdot n \cdot h_{max}}{p+1} < 4\right)$, the best choices are HC4m or HC5m.
- In the occurrence of non-normal errors, use of HC4m is preferable.

Since we now have the knowledge and such an easy access to more precise tools, we should use HC-estimators in the occurrence of heteroskedasticity to verify the significance of model parameters and to analyse them in depth.

Acknowledgments

This work was supported by the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant number 223071 (COURAGE in Europe) and Polish Ministry for Science and Higher Education under Grant number 1277/7PR/ UE/2009/7, 2009-2012.

REFERENCES

BEAL, C., (2006). Loneliness in older women: a review of the literature. Issues in Mental Health Nursing, 27 (7), pp. 795–813.

BOX, G.E.P., COX, D. R., (1964). An Analysis of Transformations. Journal of the Royal Statistical Society, Series B (26), pp. 211–252.

BREUSCH, T. S., PAGAN, A. R., (1979). A Simple Test for Heteroscedasticity and Random Coefficient Variation, Econometrica, 47 (5), pp. 1287–1294.

CABALLERO, F. F., MIRET, M., POWER, M., CHATTERJI, S., TOBIASZ-ADAMCZYK, B., KOSKINEN, S., AYUSO-MATEOS, J. L., (2013). Validation of an instrument to evaluate quality of life in the aging population: WHOQOL-AGE. Health Qual Life Outcomes, 11 (177).

CARROLL, R. J., RUPPERT, D., (1984). Power Transformations When Fitting Theoretical Models to Data. Journal of the American Statistical Association, 79 (386).

CHESHER, A., JEWITT, I., (1987). The bias of a heteroskedasticity consistent covariance matrix estimator. Econometrica, 55 (5), pp. 1217–1222.

CRAGG, J. G., (1983). More efficient estimation in the presence of heteroscedasticity of unknown form. Econometrica, 51 (3), pp. 751–763.

CRIBARI-NETO, F., (2004). Asymptotic inference under heteroskedasticity of unknown form. Computational Statistics & Data Analysis, 45, pp. 215–233.

CRIBARI-NETO, F., DA SILVA, W. B., (2011). A new heteroskedasticity-consistent covariance matrix estimator for the linear regression model. Advances in Statistical Analysis, 95, pp. 129–146.

CRIBARI-NETO, F., FERRARI, S. L. P., CORDEIRO, H. M., (2000). Improved heteroscedasticity-consistent covariance matrix estimators. Biometrika, 87 (4), pp. 907–918.

CRIBARI–NETO, F., SOUZA, T. C., VASCONCELLOS, K. L. P., (2007). Inference Under Heteroskedasticity and Leveraged Data. Communications in Statistics—Theory and Methods, 36, pp. 1877–1888.

CRIBARI-NETO, F., ZARKOS, S. P., (2001). Heteroskedasticity-consistent covariance matrix estimation: White's estimator and the bootstrap. Journal of Statistical Computation and Simulation, 68, pp. 391–411.

DALGARD, O. S., (1996). Community health profile: a tool for psychiatric prevention, in Promotion of Mental Health. D. R. Trent, C.A. Reed (eds). Aldershot: Avebury Press, pp. 681–695.

DAVIDIAN, M., CARROLL, R. J., (1987). Variance Function Estimation. Journal of the American Statistical Association. 82 (400), pp. 1079-1091.

DODGE, Y., (2008). Gauss-Markov Theorem, in The Concise Encyclopedia of Statistics. Springer New York, pp. 217–218.

EFRON, B., (1982). The Jackknife, the Bootstrap and Other Resampling Plans. Society for Industrial and Applied Mathematics, Philadelphia.

GOLDFELD, S. M., QUANDT, R. E., (1965). Some Tests for Homoscedasticity. Journal of the American Statistical Association, 6 (310), pp. 539–547.

GREENE, W. H., (2012). Econometric Analysis (7th Edition). Pearson Education Limited.

HAYES, A. F. CAI, L., (2007). Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation. Behavior Research Methods, 39 (4), pp. 709–722.

HAYES, A. F., CAI, L., (2008). A New Test of Linear Hypotheses in OLS Regression Under Heteroscedasticity of Unknown Form. Journal of Educational and Behavioral Statistics, 33 (1), pp. 21–40.

HINKLEY, D. V., (1977). Jackknifing In Unbalanced Situations. Technometrics, 19 (3).

HUGHES, M. E., WAITE, L. J., HAWKLEY, L. C., CACIOPPO, J. T., (2004). A Short Scale for Measuring Loneliness in Large Surveys: Results From Two Population-Based Studies. Research on aging, 26 (6), pp. 655–672.

KAUERMAN, G., CARROLL, R. J., (2001). A note on the efficiency of sandwich covariance matrix estimation. Journal of the American Statistical Association, 96, pp. 1387–1396.

KOJIMA, M., KOJIMA, T., ISHIGURO, N., OGUCHI, T., OBA, M., TSUCHIYA, H., TOKUDOME, S., (2009). Psychosocial factors, disease status, and quality of life in patients with rheumatoid arthritis. Journal of Psychosomatic Research, 67, pp. 425–431.

KIVIET, J.F., FENG, Q., (2015). Efficiency Gains by Modifying GMM Estimation in Linear Models under Heteroskedasticity. Econometrics Working Papers, 14 (06).

LI, S., ZHANG, N., ZHANG, X., WANG, G., (2017). A new heteroskedasticity-consistent covariance matrix estimator and inference under heteroscedasticity. Journal Of Statistical Computation And Simulation, 87 (1) pp. 198–210.

LONG, J. S., ERVIN, L.H., (2000). Using Heteroscedasticity Consistent Standard Errors in the Linear Regression Model. The American Statistician, 54, pp. 217–224.

MACKINNON, J. G., WHITE, H., (1985). Some Heteroskedasticity-Consistent Covariance Matrix Estimators with Improved Finite Sample Properties. Queen's Economics Department Working Paper, 537.

MILLER, S., STARTZ, R., (2017). Feasible Generalized Least Squares Using Machine Learning. Social Science Research Network [online] (6th February 2018). Available at: <https://ssrn.com/abstract=2966194> [Accessed on 27 March 2018].

MUSICH, S., WANG, S. S., HAWKINS, K., YEH, C. S., (2015). The Impact of Loneliness on Quality of Life and Patient Satisfaction Among Older. Sicker Adults. Gerontology & Geriatric Medicine, 1.

SAKIA, R. M., (1992). The Box-Cox transformation technique: a review. The Statistician, 41, pp. 169–178.

SMYTH, G. K., (2002). An Efficient Algorithm for REML in Heteroscedastic Regression. Journal of Graphical and Computational Statistics, 11, pp. 836–847.

STARTZ, R., (2017). Bayesian Heteroskedasticity-Robust Regression, 2015. The Economics Department of UCSB, [online] (Revised February 2015). Available

at: <http://econ.ucsb.edu/~startz/Bayesian%20Heteroskedasticity-Robust%20Regression.pdf> [Accessed on 27 March 2018].

THOMOPOULOU, I., THOMOPOULOU, D., KOUTSOUKI, D., (2010). The differences at quality of life and loneliness between elderly people. Biology of Exercise, 6 (2), pp. 13–28.

TOBIASZ-ADAMCZYK, B., GALAS, A., ZAWISZA, K., CHATTERJI, S., HARO, J. M., AYUSO-MATEOS, J. L., LEONARDI, M., (2017). Gender-related differences in the multi-pathway effect of social determinants on quality of life in older age—the COURAGE in Europe project. Quality of Life Research, 26 (7), pp. 1865–1878.

VERBEEK, M., (2004). A Guide to Modern Econometrics (2nd edition), John Wiley & Sons Ltd.

VILHENA, E., PAIS-RIBEIRO, J., SILVA, I., PEDRO, L., MENESES, R., CARDOSO, H., MENDONÇA, D., (2014). Psychosocial factors as predictors of quality of life in chronic Portuguese patients. Health and Quality of Life Outcomes, 12 (3).

WAGENER, J., DETTE, H., (2012). Bridge Estimators and the Adaptive Lasso under Heteroscedasticity. Mathematical Methods of Statistics, 21 (2), pp. 109–126.

WHITE, H., (1980). A heteroskedasticity consistent covariance matrix estimator and a direct test for heteroscedasticity. Econometrica, 48 (4), pp. 817–838.

WIECZOROWSKA-TOBIS, K., TALARSKA, D., (2010). Geriatria i pielęgniarstwo geriatryczne [Geriatrics and geriatric nursing]. Wydawnictwo Lekarskie PZWL, Warszawa.

ZAWISZA, K., GAŁAŚ, A., TOBIASZ-ADAMCZYK, B., (2014). Polska wersja Courage Social Network Index — skali do oceny poziomu sieci społecznych [Polish version of The Courage Social Network Index — the scale for the assessment of social networks]. Gerontologia Polska, 22 (1), pp. 31–41.

ZAWISZA, K., GAŁAŚ, A., TOBIASZ-ADAMCZYK, B., (2016). Walidacja polskiej wersji skali oceny jakości życia WHOQOL-AGE w populacji osób starszych [Validation of the Polish version of the WHOQOL-AGE scale in older population]. Gerontologia Polska, 1, pp. 7–16.