Intermediate Econometrics

Heteroskedasticity

Text: Wooldridge, 8

July 17, 2011

Heteroskedasticity

- Assumption of **homoskedasticity**,

$$\mathsf{Var}(u_i|x_{i1},\ldots,x_{ik}) = \mathsf{E}(u_i^2|x_{i1},\ldots,x_{ik}) = \sigma^2.$$

- That is, the variance of u does not depend on the explanatory variables.

- Under **heteroskedasticity**, the variance of the error u_i depends on $\boldsymbol{x}_i = [x_{i1}, x_{i2}, \ldots, x_{ik}]$, that is

$$\mathsf{E}(u_i^2|x_{i1},\ldots,x_{ik}) = h(\boldsymbol{x}_i) = \sigma_i^2.$$

- Assume that the other Gauß–Markov assumptions hold.

- As an example, consider a model for average monthly credit card expenditures for $n = 72$ individuals,

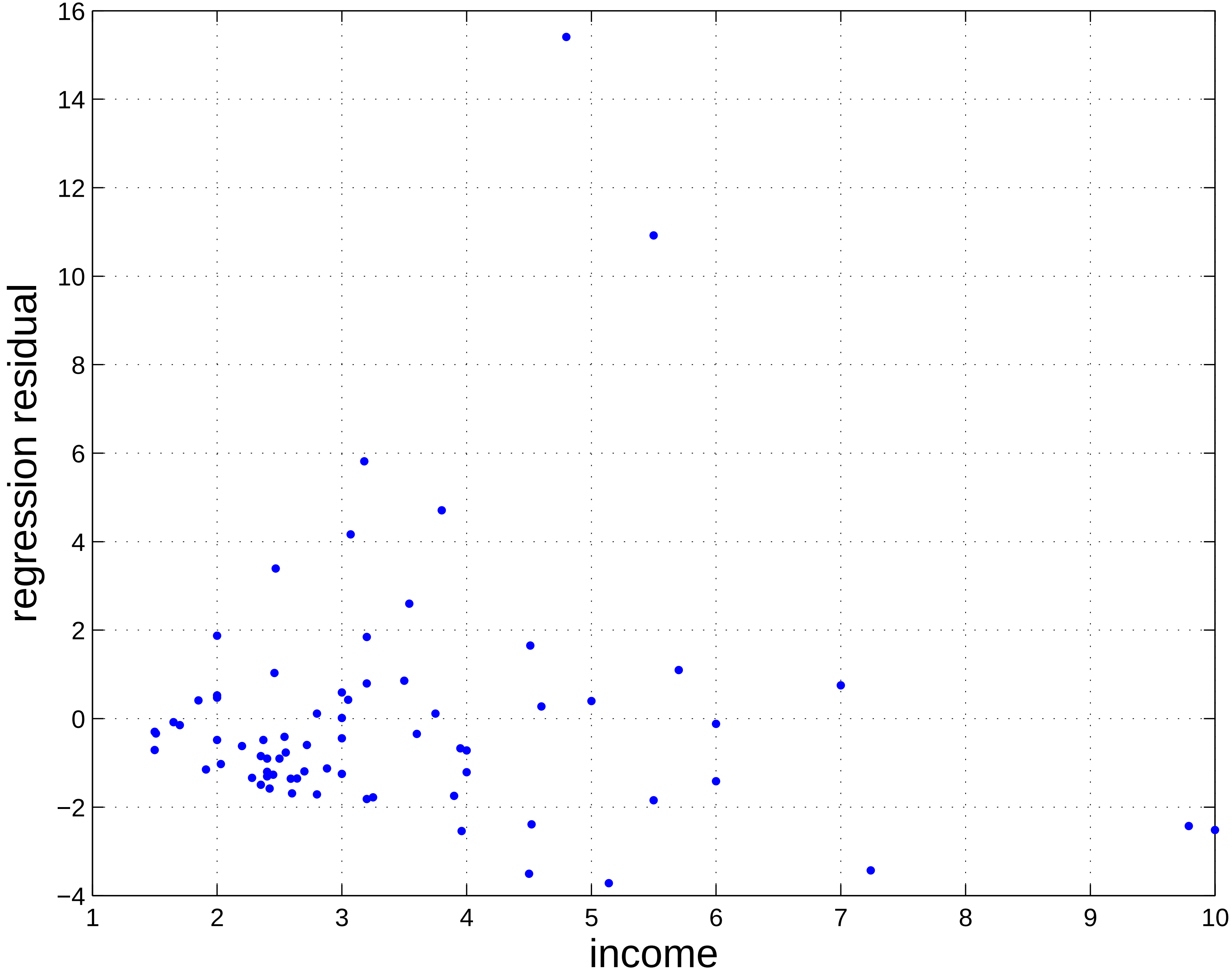
$$ccexp = \beta_0 + \beta_1 income + \beta_2 income^2 + \beta_3 age + \beta_4 ownrent + u, \quad (1)$$

  where

  - $ccexp$ are credit card expenditures
  - $ownrent$ is a dummy variable for home ownership

- We may plot the residuals of this regression against the independent variables to see if a distinctive pattern appears.

- Variance appears to increase with the income.



16
14
12
10
8
6
4
2
0
−2
−4
1
2
3
4
5
6
7
8
9
10
regression residual
income

Consequences of Heteroskedasticity

- Provided the other Gauß–Markov Assumptions are valid, then the OLS estimator is still
  - unbiased
  - consistent
  - asymptotically normally distributed.
- However, the usual standard errors are biased, and therefore t– and F–tests based on these standard errors are not valid.
- The OLS estimator is also no longer efficient (not BLUE).

- To illustrate, consider the slope estimator in the simple linear regression model

$$y_i = \beta_0 + \beta_1 x_i + u_i, \tag{2}$$

i.e.,

$$\widehat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \beta_1 + \frac{\sum_{i=1}^{n}(x_i - \bar{x})u_i}{n s_x^2} = \beta_1 + \sum_{i=1}^{n} w_i u_i,$$

where

$$w_i = \frac{x_i - \overline{x}}{n s_x^2}, \quad s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{3}$$

- This shows that

$$\mathsf{Var}(\widehat{\beta}_1 | \boldsymbol{X}) = \sum_{i=1}^{n} w_i^2 \mathsf{Var}(u_i | x_i) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sigma_i^2}{n^2 s_x^4}. \tag{4}$$

- The variance

$$\mathsf{Var}(\widehat{\beta}_1|\boldsymbol{X}) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2\sigma_i^2}{n^2 s_x^4}. \tag{5}$$

reduces to our previous formula

$$\mathsf{Var}(\widehat{\beta}_1|\boldsymbol{X}) = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} = \frac{\sigma^2}{n s_x^2} \tag{6}$$

only if $\sigma_i^2 = \sigma^2$ for $i = 1, \ldots, n$.

- For the general model,

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \quad i = 1, \ldots, n,$$

or

$$\underbrace{\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}}_{=\boldsymbol{y}} = \underbrace{\begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1k} \\ 1 & x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{nk} \end{bmatrix}}_{=\boldsymbol{X}} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix}}_{=\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}}_{=\boldsymbol{u}},$$

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$$

with

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y},$$

we have

$$
\begin{aligned}
\mathsf{Cov}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) &= \begin{bmatrix} \mathsf{Var}(\widehat{\beta}_0|\boldsymbol{X}) & \mathsf{Cov}(\widehat{\beta}_0,\widehat{\beta}_1|\boldsymbol{X}) & \cdots & \mathsf{Cov}(\widehat{\beta}_0,\widehat{\beta}_k|\boldsymbol{X}) \\ \mathsf{Cov}(\widehat{\beta}_0,\widehat{\beta}_1|\boldsymbol{X}) & \mathsf{Var}(\widehat{\beta}_1|\boldsymbol{X}) & \cdots & \mathsf{Cov}(\widehat{\beta}_1,\widehat{\beta}_k|\boldsymbol{X}) \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Cov}(\widehat{\beta}_0,\widehat{\beta}_k|\boldsymbol{X}) & \mathsf{Cov}(\widehat{\beta}_k,\widehat{\beta}_1|\boldsymbol{X}) & \cdots & \mathsf{Var}(\widehat{\beta}_k|\boldsymbol{X}) \end{bmatrix} \\
&= \mathsf{E}[(\widehat{\boldsymbol{\beta}} - \mathsf{E}(\widehat{\boldsymbol{\beta}}))(\widehat{\boldsymbol{\beta}} - \mathsf{E}(\widehat{\boldsymbol{\beta}}))'|\boldsymbol{X}] \\
&= \mathsf{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}|\boldsymbol{X}\right] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathsf{E}(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}.
\end{aligned}
$$

- If $\mathsf{E}(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X}) = \sigma^2\boldsymbol{I}$, this reduces to

$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1},$$

which was the basis for the inferential procedures discussed so far.

- Under heteroskedasticity, we have

$$\mathsf{E}(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X}) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix} =: \boldsymbol{\Omega}.$$

- The covariance matrix of the errors is still diagonal.

- The appropriate way to deal with heteroskedasticity depends on whether we (believe that we) know how the variance depends on the independent variables, at least approximately.

- We first consider the situation where we do not want to make an assumption about the functional relationship

$$\mathsf{E}(u_i^2|x_{i1},\ldots,x_{ik}) = h(x_{i1},\ldots,x_{ik}).$$

- This leads to the use of *heteroskedasticity–consistent* standard errors.

Heteroskedasticity–Robust Standard Errors

- The basic idea is to still use the ordinary least squares (OLS) estimator but to appropriately adjust the standard errors used in hypotheses tests.

- Last week, when discussing the asymptotic normality of

$$\widehat{\beta}_1 = \beta_1 + \frac{\frac{1}{n}\sum_i (x_i - \overline{x})u_i}{\frac{1}{n}\sum_i (x_i - \overline{x})^2}, \tag{7}$$

  we have seen that

$$\sqrt{n}(\widehat{\beta}_1 - \beta_1) \xrightarrow{d} \mathsf{N}\left(0, \frac{\mathsf{E}\{(x_i - \mu_x)^2 u_i^2\}}{\mathsf{Var}(x)^2}\right). \tag{8}$$

- In case of homoskedasticity, i.e., $\mathsf{E}(u_i^2|x_i) = \sigma^2$, it turns out that

$$\mathsf{E}\{(x_i - \mu_x)^2 u_i^2\} = \sigma^2 \mathsf{E}(x_i - \mu_x)^2 = \sigma^2 \mathsf{Var}(x). \tag{9}$$

- When we have heteroskedasticity, a natural consistent estimator for the asymptotic variance of $\widehat{\beta}_1$ is obtained by replacing population moments with sample moments, i.e.,

$$\widehat{Var}(\widehat{\beta}_1) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 u_i^2}{n\left(\frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2\right)^2}. \tag{10}$$

- However, (10) is not feasible since the errors u_i, $i = 1, \ldots, n$, in the numerator are not observable.

- However, it turns out (White, 1980) that to obtain a feasible consistent estimator we can estimate E$\{(x_i - \mu_x)^2 u_i^2\}$ via

$$\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2 \widehat{u}_i^2,$$

where $\widehat{u}_i$ is the standard OLS residual.

- Thus, the **heteroskedasticity–consistent (HC) standard error** of the slope coefficient $\widehat{\beta}_1$ in the simple linear regression model is

$$\widehat{\sigma}^2_{\widehat{\beta}_1,HC} = \widehat{\mathsf{Var}}(\widehat{\beta}_1) = \frac{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2\widehat{u}_i^2}{ns_x^4}, \quad s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2. \tag{11}$$

- (11) consistently estimates the asymptotic variance both in case of homo– and heteroskedasticity.

- That is, in large samples, we can treat $\widehat{\beta}_1$ as

$$\widehat{\beta}_1 \overset{a}{\sim} \mathsf{N}(\beta_1, \widehat{\sigma}^2_{\widehat{\beta}_1,HC}), \tag{12}$$

where $\overset{a}{\sim}$ denotes "approximately in large samples".

- Heteroskedasticity–robust t statistics can be calculated based on (12).

- They are calculated in the same way as before, with the exception that robust standard errors are used.

- Note that their use is based on asymptotic arguments.

- Such standard errors are often referred to as *White standard errors* due to White (1980).[1]

[1] H. White (1980): A Heteroskedasticity–Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, *Econometrics*, 48, 817-838.

Robust standard errors for the multiple regression model

- The covariance matrix is

$$\mathsf{Cov}(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathsf{E}(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1},$$

- Now matrix

$$\boldsymbol{X}'\mathsf{E}(\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X})\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\mathsf{E}(u_i^2|\boldsymbol{x}_i),$$

where $\boldsymbol{x}_i = [x_{i1}, \ldots, x_{ik}]'$, is replaced with

$$\sum_{i=1}^{n} \boldsymbol{x}_i\boldsymbol{x}_i'\widehat{u}_i^2$$

to obtain an asymptotically valid covariance matrix estimator under heteroskedasticity (and homoskedasticity).

- Thus we estimate the asymptotic covariance matrix via

$$\widehat{\mathsf{Cov}}(\widehat{\beta}) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}, \tag{13}$$

where

$$\widehat{\Omega} = \begin{bmatrix} \widehat{u}_1^2 & 0 & \cdots & 0 \\ 0 & \widehat{u}_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widehat{u}_n^2 \end{bmatrix}. \tag{14}$$

- Once heteroskedasticity–consistent standard errors have been computed, **heteroskedasticity–consistent t statistics** can be calculated in the usual way.

- The only difference between the usual t statistic and the heteroskedasticity–robust t statistic is in how the standard error in the denominator is calculated.

- Multiple (linear) hypotheses can likewise be tested, although the test statistics used so far (i.e., the F test) are no longer appropriate.

- The relevant test is known as a *Wald test*.[2]

[2] For the form of the test statistic and an example, see, for example, Greene, *Econometric Analysis*, Ch. 11.2.

Testing for Heteroskedasticity

- Informal "test": Plot OLS residuals against the independent variables, as in the credit card example.

- In principle, a plot against *all* the independent variables may be required.

Goldfeld–Quandt Test

- This compares the variance of two groups by means of an F test, where under the null hypothesis of homoskedasticity,

$$H_0 : \sigma_1^2 = \sigma_2^2, \quad H_1 : \sigma_1^2 \neq \sigma_2^2. \tag{15}$$

- For example, we may suppose that the variance differs by gender.

- Calculate the regression of interest separately for both groups (with n_1 observations in Group 1 and n_2 observations in Group 2).

- From these regressions, estimate the error variances

$$\widehat{\sigma}_1^2 = \frac{\mathsf{SSR}_1}{n_1 - k - 1}, \quad \widehat{\sigma}_2^2 = \frac{\mathsf{SSR}_2}{n_2 - k - 1}.$$

- Under the null hypothesis (15) of equal variances, the test statistic

$$F = \frac{\widehat{\sigma}_1^2}{\widehat{\sigma}_2^2} = \frac{\mathsf{SSR}_1/(n_1 - k - 1)}{\mathsf{SSR}_2/(n_2 - k - 1)} \tag{16}$$

has an (exact[3] or approximate) F distribution with $n_1 - k - 1$ degrees of freedom in the numerator and $n_2 - k - 1$ degrees of freedom in the denominator.

- The larger residual variance is used in the numerator, i.e., in (16), it has been assumed that $\widehat{\sigma}_1^2 > \widehat{\sigma}_2^2$.

- E.g., we tested for gender–specific slope coefficients in the wage equation

$$\log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 exper^2 + \beta_4 tenure + \beta_5 tenure^2.$$

- For the Chow test (for differences in regression functions) to be valid, we need to test whether the error variance is the same for both groups.

[3] This holds if the errors are normal, i.e., the assumptions of the classical linear model apply.

- The sample has $n_1 = 274$ men and $n_2 = 252$ women ($n = n_1 + n_2 = 526$).

- From Tables 3 and 4 of the last slide set, we have

$$\widehat{\sigma}_1^2 = \frac{43.2453}{274 - 6} = 0.1614, \quad \widehat{\sigma}_2^2 = \frac{36.6751}{252 - 6} = 0.1491,$$

so the test statistic (16) is

$$F = \frac{0.1614}{0.1491} = 1.0824.$$

- The critical value of the appropriate F distribution is 1.2262, so we cannot reject the null of homoskedasticity.

- The Goldfeld–Quandt test can also be applied in cases where we suppose the variance to depend on the value of a quantitative variable, e.g.,

$$\sigma_i^2 = \sigma^2 x_{ij}, \quad i = 1, \ldots, n.$$

- Then we rank all the observations based on this x_j and thereby separate the observations into those with low and high variances.

- In the credit card example, we sort the $n = 72$ observations according to income ($= x_1$), and then the regression is calculated separately both for the first and the second 36 observations.

- The sum of squares for the first and the second regression is 32.6247 and 489.4130, respectively, with gives rise to a Goldfeld–Quandt test statistic

$$F = \frac{\mathsf{SSR}_1/(n_1 - k - 1)}{\mathsf{SSR}_2/(n_2 - k - 1)} = \frac{489.4130/(36 - 4 - 1)}{32.6247/(36 - 4 - 1)} = 15.0013,$$

which can be compared with the 5% (1%) critical value of the $F(31, 31)$ distribution, given by 1.8221 (2.3509).

Breusch–Pagan Test

- The Breusch–Pagan test can be used when we have an idea about which variables have an impact on the error term variance.

- Let this set of regressors be $x_1, \ldots, x_p$.

- We may reasonably assume that the magnitude of the OLS residuals, $\widehat{u}_i$, has something to say about σ_i^2.

- Thus specify the regression

$$\widehat{u}_i^2 = \delta_0 + \delta_1 x_{i1} + \cdots + \delta_p x_{ip} + v_i. \tag{17}$$

where the $\widehat{u}_i$ have been obtained from standard OLS.

- Then calculate the coefficient of determination R^2 from the regression (17).

- The test statistic to test the null hypothesis

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_p = 0$$

is

$$\text{BP} = nR^2 \overset{asy}{\sim} \chi^2(p).$$

- For the credit card example, we might specify the test based on

$$\widehat{u}_i^2 = \delta_0 + \delta_1 income_i + \delta_2 income_i^2 + v_i. \tag{18}$$

- (18) gives rise to a coefficient of determination $R^2 = 0.0859$, so

$$\text{BP} = n \times R^2 = 72 \times 0.0859 = 6.1869,$$

so we would reject homoskedasticity by comparison with the 5% critical value of the χ^2 distribution with two degrees of freedom, which is 5.9915.

White Test

- The White test is based on a comparison between the OLS covariance matrix under homoskedasticity and under a general form of heteroskedasticity.

- It is carried out along the following lines.

- An auxiliary regression is calculated as in the Breusch–Pagan test, but the $\widehat{u}_i$s are regressed, in addition to all independent variables and an intercept, on all squares and all cross products of the independent variables.

- For example, in a model with $k = 3$ independent variables, the auxiliary regression is

$$\begin{aligned} \widehat{u}^2 &= \delta_0 + \delta_1 x_1 + \delta_2 x_2 + \delta_3 x_3 + \delta_4 x_1^2 + \delta_5 x_2^2 + \delta_6 x_3^2 \\ &\quad + \delta_7 x_1 \cdot x_2 + \delta_8 x_1 \cdot x_3 + \delta_9 x_2 \cdot x_3 + v, \end{aligned} \tag{19}$$

  and we test

$$H_0 : \delta_1 = \delta_2 = \cdots = \delta_9 = 0$$

  as in the Breusch–Pagan test (i.e., comparing with the critical value of the appropriate χ^2 distribution).

- Note that there may be redundant terms in the regression (19), e.g., if squares of a variable are included in the original regression.

- In this case, the redundant terms are dropped and the degrees of freedom are reduced appropriately.

- The White test is very general. Its power, however, may be rather low against certain alternatives, in particular if the number of observations is small.

- For the credit card example, we have the constant plus 12 variables, since $ownrent^2 = ownrent$, and $income \times income = income^2$.

- The regression gives $R^2 = 0.1990$.

- Thus LM $= 72 \times 0.1990 = 14.3290$, which is, however, not significant when compared to the 5% critical value of the $\chi^2(12)$ distribution, which is 21.0261.

- If we have an idea about the nature of heteroskedasticity (i.e., the variables affecting the error term variance), the Breusch–Pagan test has more power in detecting heteroskedasticity.

Generalized (or Weighted) Least Squares Estimation

- Suppose that the heteroskedastic pattern of the variance is known up to a multiplicative constant, i.e., in

$$\mathsf{E}(u_i^2|\boldsymbol{x}_i) = \sigma^2 h(\boldsymbol{x}_i) = \sigma^2 h_i, \tag{20}$$

  the function $h(\boldsymbol{x}_i)$ is known.

- In this case, it is possible to obtain an efficient (BLUE) estimator along with valid t and F statistics (which have known exact distributions under the Gaussianity assumption) by means of **generalized least squares**.

- The idea is to transform the linear equation such that it has homoskedastic errors.

- Consider the equation

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + u_i, \tag{21}$$

where

$$\mathsf{E}(u_i^2|\boldsymbol{x}_i) = \sigma^2 h_i. \tag{22}$$

- As we have assumed h_i to be known, we can consider the transformed residual $u_i^\star = u_i/\sqrt{h_i}$, with variance

$$\mathsf{E}(u_i^{\star 2}|\boldsymbol{x}_i) = \mathsf{E}\left[\left(\frac{u_i}{\sqrt{h_i}}\right)^2 \middle| \boldsymbol{x}_i\right] = \frac{\mathsf{E}(u_i^2|\boldsymbol{x}_i)}{h_i} = \frac{\sigma^2 h_i}{h_i} = \sigma^2,$$

so $u_i^{\star 2}$ is homoskedastic, i.e.,

$$\mathsf{E}(u_i^{\star 2}|\boldsymbol{x}_i) = \sigma^2 \quad \text{for all } i. \tag{23}$$

- Thus, we divide equation (21) by $\sqrt{h_i}$ to obtain

$$y_i/\sqrt{h_i} = \beta_0/\sqrt{h_i}+\beta_1 x_{i1}/\sqrt{h_i}+\beta_2 x_{i2}/\sqrt{h_i}+\cdots+\beta_k x_{ik}/\sqrt{h_i}+u_i/\sqrt{h_i}, \tag{24}$$

or

$$y_i^\star = \beta_0 x_{i0}^\star + \beta_1 x_{i1}^\star + \beta_2 x_{i2}^\star + \cdots + \beta_k x_{ik}^\star + u_i^\star, \tag{25}$$

where

$$x_{i0}^\star = 1/\sqrt{h_i}, \quad x_{ij}^\star = x_{ij}/\sqrt{h_i}, \quad j = 1, \ldots, k, \quad y_i^\star = y_i/\sqrt{h_i}. \tag{26}$$

- Equation (25) satisfies the Gauß–Markov assumptions in view of (23).

- Thus, estimating Equation (25) gives **b**est **l**inear **u**nbiased **e**stimators (BLUE).

- Moreover, if u is normal, $u^\star$ is also normal, and Equation (25) satisfies the assumptions of the classical linear model (Gauß–Markov + normality).

- In this case, exact t and F tests can be conducted based on the transformed variables.

- That is, after the variables have been transformed as in (24), standard OLS inferential procedures can be applied.

- The multiplicative constant σ^2 can be estimated as

$$\widehat{\sigma}^2 = \frac{1}{n-k-1}\sum_{i=1}^{n} \widehat{u}_i^{\star 2}.$$

- *Parameter estimates and test results, however, have to be interpreted in terms of the original equation, i.e., Equation (21).*

- Also note that the R^2 calculated from the transformed equation may not be that interesting, since it measures the variation in $y^\star$ explained by $x_j^\star$, $j = 1, \ldots, k$.[4]

- This procedure is an example of **generalized least squares** (GLS), which can generally applied when the covariance matrix of the u_is is not equal to $\sigma^2 \boldsymbol{I}$.

[4] It may be useful for doing F tests, however.

- The GLS estimator correcting for heteroskedasticity (with diagonal error term covariance matrix) is also referred to as **weighted least squares**, since it minimizes the weighted sum of squared residuals, where each squared residual is weighted by the inverse of its variance.

- This can be seen by noting that OLS estimation of Equation (25) amounts to minimizing

$$
\begin{aligned}
\sum_{i=1}^{n} \widehat{u}_i^{\star 2} &= \sum_{i=1}^{n} (y_i^\star - \widehat{\beta}_0 x_{i0}^\star - \widehat{\beta}_1 x_{i1}^\star - \widehat{\beta}_2 x_{i2}^\star - \cdots - \widehat{\beta}_k x_{ik}^\star)^2 \\
&= \sum_{ii=1}^{n} \frac{(y_i - \widehat{\beta}_0 x_{i0} - \widehat{\beta}_1 x_{i1} - \widehat{\beta}_2 x_{i2} - \cdots - \widehat{\beta}_k x_{ik})^2}{h_i}.
\end{aligned}
$$

- This intuition is that less weight is given to observations with a relatively high error variance, and thus they have a smaller influence on the estimates.

- Note that in the transformed model all variables are transformed, including the constant, which implies that the transformed model (usually) does not have an intercept.[5]

[5] There will be an intercept in the transformed equation if $h_i = x_{ij}^2$ for one of the independent variables x_j.

- As an example, in the simple linear regression model through the origin,

$$y_i = \beta_1 x_i + u_i,$$

the slope estimator is

$$\widehat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i^{\star} y_i^{\star}}{\sum_{i=1}^{n} x_i^{\star 2}} = \frac{\sum_{i=1}^{n} x_i y_i / h_i}{\sum_{i=1}^{n} x_i^2 / h_i} = \frac{\sum_{i=1}^{n} x_i (x_i \beta_1 + u_i) / h_i}{\sum_{i=1}^{n} x_i^2 / h_i} \tag{27}$$

$$= \beta_1 + \frac{\sum_{i=1}^{n} x_i u_i / h_i}{\sum_{i=1}^{n} x_i^2 / h_i}. \tag{28}$$

- The variance is

$$\mathsf{Var}(\widehat{\beta}_1) = \frac{\sum_i x_i^2 / h_i^2 \mathsf{Var}(u_i)}{\left(\sum_{i=1}^{n} x_i^2 / h_i\right)^2} = \frac{\sum_i x_i^2 / h_i^2 (\sigma^2 h_i)}{\left(\sum_{i=1}^{n} x_i^2 / h_i\right)^2} = \frac{\sigma^2}{\sum_{i=1}^{n} x_i^2 / h_i}.$$

- The OLS estimator,

$$\widehat{\beta}_1^{OLS} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} = \beta_1 + \frac{\sum_i x_i u_i}{\sum_i x_i^2},$$

- has variance

$$\mathsf{Var}(\widehat{\beta}_1^{OLS}) = \frac{\sum_i x_i^2 \mathsf{Var}(u_i)}{\left(\sum_i x_i^2\right)^2} = \frac{\sum_i x_i^2 (\sigma^2 h_i)}{\left(\sum_i x_i^2\right)^2} = \sigma^2 \frac{\sum_i x_i^2 h_i}{\left(\sum_i x_i^2\right)^2},$$

and

$$\frac{\sigma^2}{\sum_i x_i^2 / h_i} \leq \sigma^2 \frac{\sum_i x_i^2 h_i}{\left(\sum_i x_i^2\right)^2} \Leftrightarrow \left(\sum_i x_i^2\right)^2 \leq \left(\sum_i x_i^2 / h_i\right)\left(\sum_i x_i^2 h_i\right),$$

- The last inequality holds by the Cauchy–Schwarz inequality, stating that

$$\left(\sum_i w_i v_i\right)^2 \leq \left(\sum_i w_i^2\right)\left(\sum_i v_i^2\right). \tag{29}$$

(In (29), take $v_i = x_i/\sqrt{h_i}$, and $w_i = x_i\sqrt{h_i}$.)

- Proof of the Cauchy–Schwarz inequality: For any constant λ, we have

$$\sum_i (w_i - \lambda v_i)^2 = \sum_i w_i^2 - 2\lambda \sum_i w_i v_i + \lambda^2 \sum_i v_i^2 \geq 0. \tag{30}$$

Now let $\lambda = (\sum_i w_i v_i)/(\sum_i v_i^2)$.

Then (30) becomes

$$\sum_i w_i^2 - 2\frac{\left(\sum_i w_i v_i\right)^2}{\sum_i v_i^2} + \frac{\left(\sum_i w_i v_i\right)^2}{\sum_i v_i^2} \geq 0 \tag{31}$$

$$\sum_i w_i^2 - \frac{\left(\sum_i w_i v_i\right)^2}{\sum_i v_i^2} \geq 0. \tag{32}$$

Rearranging gives the result.

Matrix formulation of generalized least squares (GLS)

- If the variances are known up to a multiplicative constant, as in (20), let

$$\Omega = \begin{bmatrix} h_1 & 0 & \cdots & 0 \\ 0 & h_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n \end{bmatrix}. \tag{33}$$

- We transform the linear equation $y = X\beta + u$:

$$\begin{aligned} y &= X\beta + u & (34) \\ \Omega^{-1/2} y &= \Omega^{-1/2} X\beta + \Omega^{-1/2} u & (35) \\ y^\star &= X^\star \beta + u^\star, & (36) \end{aligned}$$

where

$$y^\star = \Omega^{-1/2} y, \quad X^\star = \Omega^{-1/2} X, \quad u^\star = \Omega^{-1/2} u,$$

and

$$\Omega^{-1/2} = \begin{bmatrix} h_1^{-1/2} & 0 & \cdots & 0 \\ 0 & h_2^{-1/2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & h_n^{-1/2} \end{bmatrix}. \tag{37}$$

- GLS estimation amounts to OLS estimation of Equation (36), i.e.,

$$\begin{aligned} \widehat{\beta}^{GLS} &= (X^{\star\prime} X^{\star})^{-1} X^{\star\prime} y^{\star} \\ &= (X'\Omega^{-1/2}\Omega^{-1/2}X)^{-1}X'\Omega^{-1/2}\Omega^{-1/2}y \\ &= (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \end{aligned}$$

($\Omega^{-1/2}\Omega^{-1/2} = \Omega^{-1}$), and the covariance matrix of $\widehat{\beta}^{GLS}$ is

$$\mathrm{Var}(\widehat{\beta}^{GLS}|X) = \sigma^2(X^{\star\prime}X^{\star})^{-1} = \sigma^2(X'\Omega^{-1}X)^{-1}. \tag{38}$$

- Note that the error term covariance matrix is $\sigma^2\Omega$.

Unknown Pattern of Heteroskedasticity: Feasible Generalized Least Squares (FGLS)

- In practice, the exact form of the heteroskedasticity function $h(\boldsymbol{x})$ may often (or usually) not be known.

- In this case, we may try to model this function by means of a relatively low–dimensional parameter vector.

- This results in a sequence of estimates $\widehat{h}_i$, $i = 1, \ldots, n$, which are then used for calculating the transformed regression equation.

- This is referred to as **feasible generalized least squares** (FGLS).

- Provided the parameters in the variance function h_i are consistently estimated, the FGLS estimator has (very generally) the same properties as the GLS estimator asymptotically.

- However, the small sample properties of GLS are not shared by FGLS.

- For example, in general, FGLS is not unbiased.

- To illustrate, consider the FGLS estimator of the slope coefficient in the simple linear model, where, replacing σ_i^2 with $\widehat{\sigma}_i^2$ in (27),

$$\widehat{\beta}_1 = \frac{\sum_i \widehat{x}_i^\star \widehat{y}_i^\star}{\sum_i \widehat{x}_i^{\star 2}} = \frac{\sum_i x_i y_i / \widehat{h}_i}{\sum_i x_i^2 / \widehat{h}_i} = \beta_1 + \sum_i \frac{x_i / \widehat{h}_i}{\sum_i x_i^2 / \widehat{h}_i} u_i$$

- Now $\widehat{h}_i$, $i = 1, \ldots, n$, is estimated using the data at hand and thus depends on $u_1, \ldots, u_n$, so that

$$\mathsf{E}\left\{\sum_i \frac{x_i / \widehat{h}_i}{\sum_i x_i^2 / \widehat{h}_i} u_i\right\} \neq \sum_i \frac{x_i / \widehat{h}_i}{\sum_i x_i^2 / \widehat{h}_i} \mathsf{E}(u_i) = 0.$$

- In small samples, due the unavoidable estimation error in the $\widehat{h}_i$, there is no guarantee that the FGLS estimator outperforms OLS, although this may be the case in situations of sufficiently strong heteroskedasticity.

Multiplicative Heteroskedasticity

- A common (and very flexible) parametric form of heteroskedasticity is **multiplicative heteroskedasticity**.

- This assumes that

$$\mathsf{E}(u^2|\boldsymbol{x}) = \sigma^2 h_i = \sigma^2 \exp\{\delta_1 z_1 + \cdots + \delta_\ell z_\ell\}, \quad (39)$$

where $z_1, \ldots, z_\ell$ are functions of $x_1, \ldots, x_2$.

- For example, if the variance if proportional to an unknown power α of x_1 (clearly then it must be that $x_1 > 0$), then

$$\sigma_i^2 = \sigma^2 x_{i1}^\alpha = \sigma^2 x_{i1}^\alpha = \sigma^2 \exp\{\alpha \log(x_{i1})\}.$$

- To determine the parameters in (39), we use the fact that the magnitude of the regression residuals $\widehat{u}_i$ contains some information about σ_i^2, $i = 1, \ldots, n$.

- In fact, a consistent estimator of parameters $\delta_1, \ldots, \delta_\ell$ in (39) can be obtained by OLS estimation of

$$\log(\widehat{u}_i^2) = \alpha + \delta_1 z_{i1} + \cdots + \delta_k z_{i\ell} + e_i, \tag{40}$$

  with e_i is an error term.

- Thus the steps of the procedure are as follows:
  - First run OLS to get the residuals $\widehat{u}_1, \ldots, \widehat{u}_n$, and calculate $\log \widehat{u}_i^2$, $i = 1, \ldots, n$
  - Then run the regression (40) and calculate $\widehat{h}_i = \exp\{\widehat{\delta}_1 x_{i1} + \cdots + \widehat{\delta}_k x_{ik}\}$.
  - Use these estimates $\widehat{h}_i$, $i = 1, \ldots, n$, to apply the generalized least squares method, i.e., run OLS on the transformed equation (25).

- This is referred to as a **two–step estimator**.

- We consider various specifications for the credit card example.

Table 1: Credit Card Expenditures, $x_1 = income$

	OLS		$\sigma_i^2 = \sigma^2 x_{i1}$		$\sigma_i^2 = \sigma^2 x_{i1}^{\alpha}$	
variable	estimate	std. error	estimate	std. error	estimate	std. error
$const$	-2.3715	1.9935	-1.8187	1.6552	-1.9333	1.7108
$income$	2.3435	0.8037	2.0217	0.7678	2.0888	0.7720
$income^2$	-0.1500	0.0747	-0.1211	0.0827	-0.1277	0.0808
age	-0.0308	0.0551	-0.0294	0.0460	-0.0296	0.0476
$ownrent$	0.2794	0.8292	0.5049	0.6988	0.4736	0.7214

In the rightmost model, $\widehat{\alpha} = 0.8193$.

Note I

- As noted above, if the parameters in the variance function h_i are consistently estimated, the FGLS estimator has the same properties as the GLS estimator asymptotically.

- E.g., its asymptotic covariance matrix is (38), where in practice Ω is replaced by $\widehat{\Omega}$.

- However, if the variance function is (seriously) misspecified, it may be inferior to OLS even for large samples.

- Thus OLS may often be actually preferable: Although it may be less efficient than FGLS when the variance function is (approximately) known, it is robust when used with appropriate (heteroskedasticity–robust) standard errors, even if we have no concrete ideas about the variance function.

Note II

- Often it may be possible to overcome heteroskedasticity in advance, e.g., by transforming the data.

- For example, in the credit card example, taking logs reduces the variation in the credit card expenditures for higher incomes.

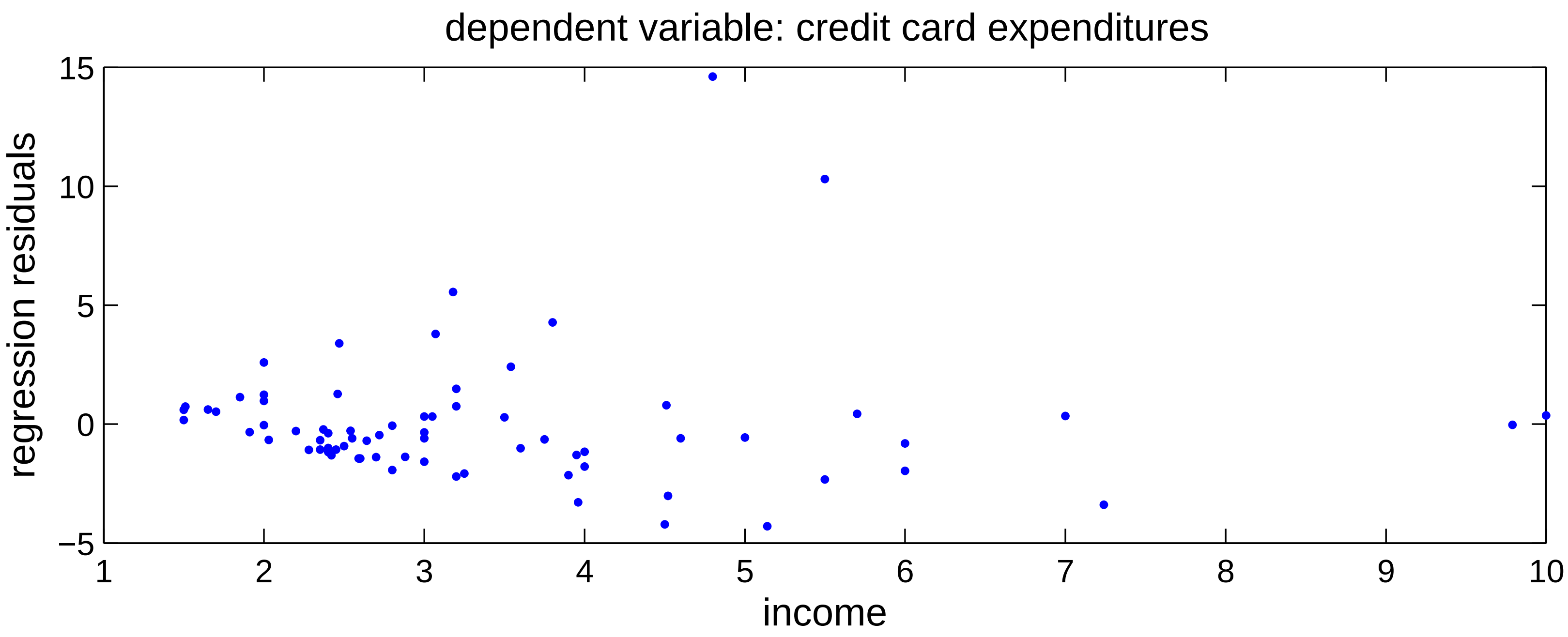
- See the figure on the next slide.

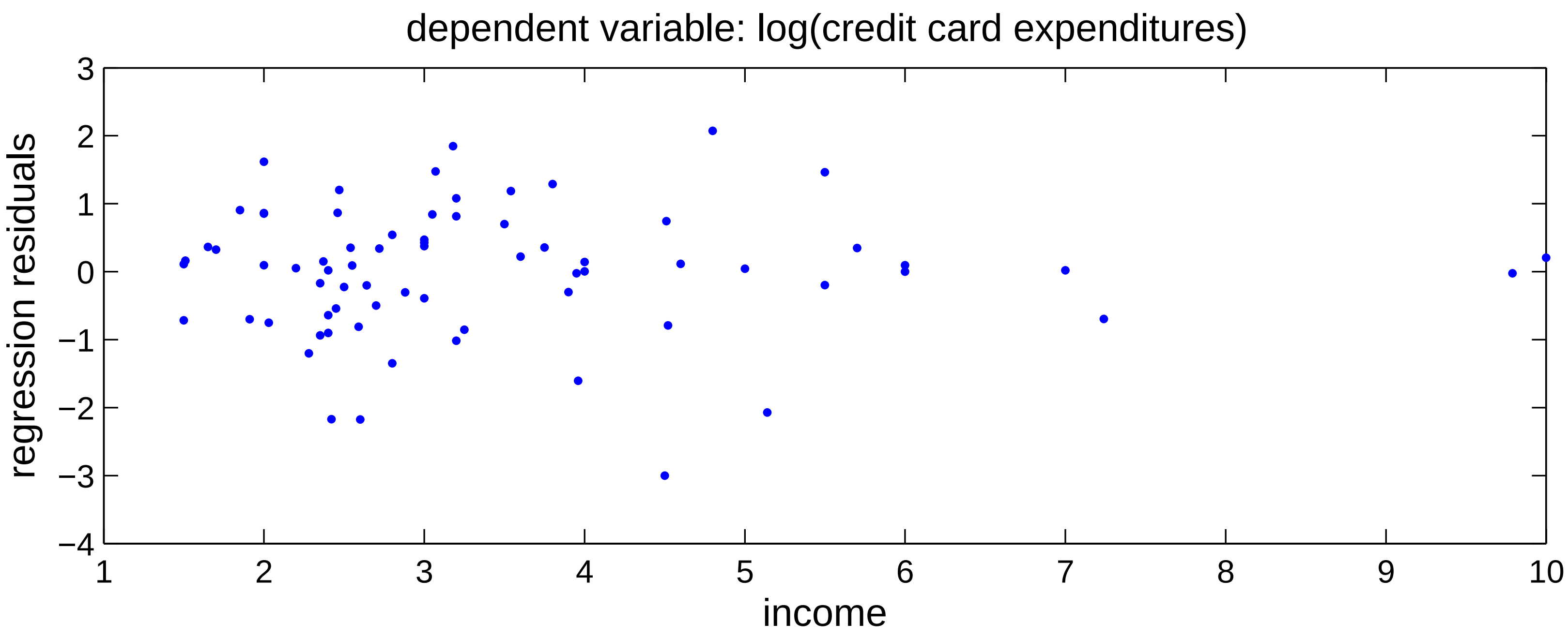
- The upper panel shows the residuals of regression

$$ccexp = \beta_0 + \beta_1 income + \beta_2 income^2 + \beta_3 age + \beta_4 ownrent + u, \quad (41)$$

- and the lower panel shows those of

$$\log(ccexp) = \beta_0 + \beta_1 income + \beta_2 income^2 + \beta_3 age + \beta_4 ownrent + u. \quad (42)$$


dependent variable: credit card expenditures
regression residuals
15
10
5
0
−5
1
2
3
4
5
6
7
8
9
10
income


dependent variable: log(credit card expenditures)
regression residuals
3
2
1
0
−1
−2
−3
−4
1
2
3
4
5
6
7
8
9
10
income

The Linear Probability Model (LPM)

- We have seen that heteroskedasticity is unavoidable in the LPM.

- One way to proceed is to use heteroskedasticity–consistent standard errors.

- We may also use weighted least squares by employing the formula

$$\mathsf{E}(u^2|\boldsymbol{x}) = p(\boldsymbol{x})(1 - p(\boldsymbol{x})),$$

  but this is difficult when some of the probabilities are actually negative.

- So first alternative would perhaps be appropriate in case of the model of labor market participation considered above.

Table 2: labor market participation

variable	coefficient	standard se	White se
constant	0.5855	0.1542	0.1514
other income	-0.0034	0.0014	0.0015
education	0.0380	0.0074	0.0072
experience	0.0395	0.0057	0.0058
experience^2	-0.0006	0.0002	0.0002
age	-0.0161	0.0025	0.0024
# kids under 6 years	-0.2618	0.0335	0.0316
# kids 6–18	0.0130	0.0132	0.0135

"se" is standard error, "White se" refers to the heteroskedasticity–robust standard errors.

- Note that the standard errors are very similar.

Table 3: 95% Quantiles of the F distribution (= 5% critical values) (ν_1 numerator degrees of freedom; ν_2 denominator degrees of freedom)

ν_2/ν_1	2	3	4	5	6	7	8	9	10
10	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782
15	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437
20	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479
25	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365
30	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646
35	3.2674	2.8742	2.6415	2.4851	2.3718	2.2852	2.2167	2.1608	2.1143
40	3.2317	2.8387	2.6060	2.4495	2.3359	2.2490	2.1802	2.1240	2.0772
45	3.2043	2.8115	2.5787	2.4221	2.3083	2.2212	2.1521	2.0958	2.0487
50	3.1826	2.7900	2.5572	2.4004	2.2864	2.1992	2.1299	2.0734	2.0261
60	3.1504	2.7581	2.5252	2.3683	2.2541	2.1665	2.0970	2.0401	1.9926
70	3.1277	2.7355	2.5027	2.3456	2.2312	2.1435	2.0737	2.0166	1.9689
80	3.1108	2.7188	2.4859	2.3287	2.2142	2.1263	2.0564	1.9991	1.9512
90	3.0977	2.7058	2.4729	2.3157	2.2011	2.1131	2.0430	1.9856	1.9376
100	3.0873	2.6955	2.4626	2.3053	2.1906	2.1025	2.0323	1.9748	1.9267
∞	2.9957	2.6049	2.3719	2.2141	2.0986	2.0096	1.9384	1.8799	1.8307

Table 4: Quantiles of the χ^2 distribution (ν denotes degrees of freedom)

ν	0.9	0.95	0.975	0.99
1	2.7055	3.8415	5.0239	6.6349
2	4.6052	5.9915	7.3778	9.2103
3	6.2514	7.8147	9.3484	11.3449
4	7.7794	9.4877	11.1433	13.2767
5	9.2364	11.0705	12.8325	15.0863
6	10.6446	12.5916	14.4494	16.8119
7	12.0170	14.0671	16.0128	18.4753
8	13.3616	15.5073	17.5345	20.0902
9	14.6837	16.9190	19.0228	21.6660
10	15.9872	18.3070	20.4832	23.2093
11	17.2750	19.6751	21.9200	24.7250
12	18.5493	21.0261	23.3367	26.2170
13	19.8119	22.3620	24.7356	27.6882
14	21.0641	23.6848	26.1189	29.1412
15	22.3071	24.9958	27.4884	30.5779