## II.9 Regression Models with Heteroscedastic Errors

If assumption (39) (homoscedastic errors) is violated, one has to deal with heteroscedastic errors, i.e. the variance differs among the observations:

$$\mathrm{Var}\,(u_i|X_1,\ldots,X_K) = \sigma_i^2. \tag{75}$$

- Standard errors of OLS estimation are no longer valid; efficiency of OLS estimation is lost.
- OLS estimations looses optimality, better estimation methods exist.

# Case Study Profit

Demonstration in EVIEWS, workfile `profit`:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i,$$

$$y_i \ldots \text{profit 1994}$$

$$x_{1,i} \ldots \text{profit 1993}$$

$$x_{2,i} \ldots \text{turnover 1994}$$

Variances increases with size of the firm

# OLS Estimation under Heteroscedasticity

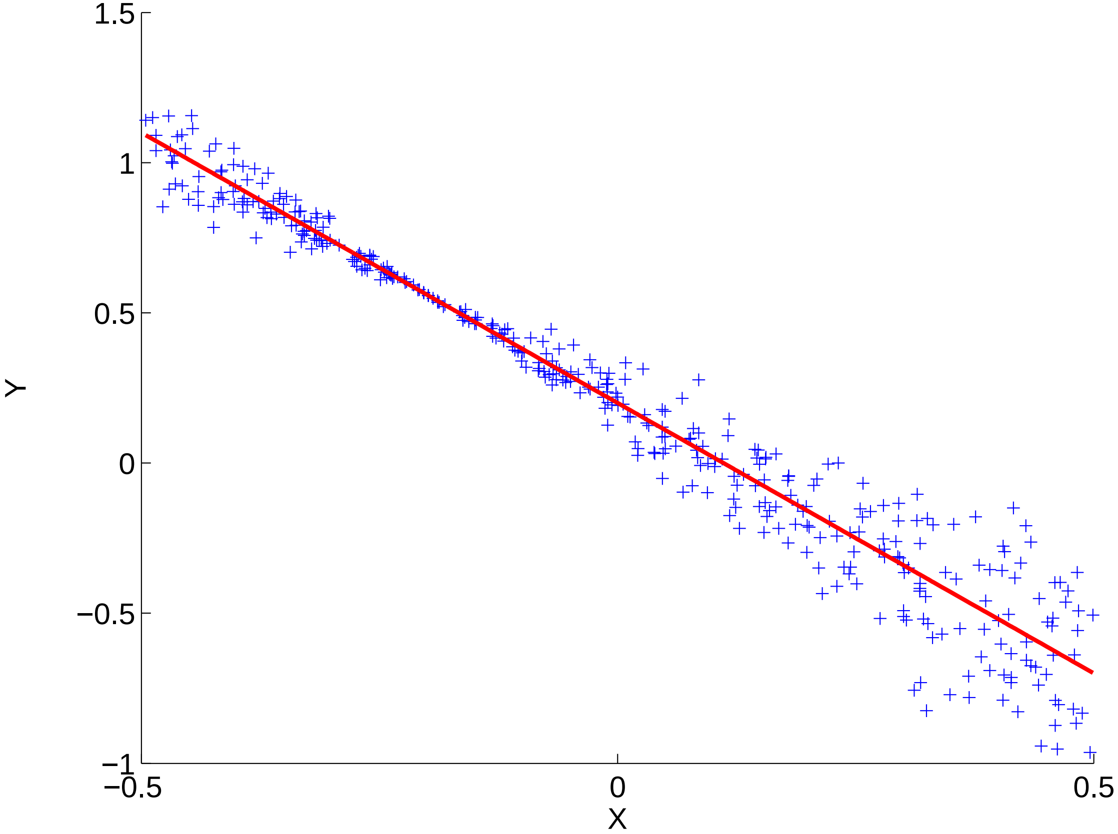
Simulate data from a regression model with $\beta_0 = 0.2$ and $\beta_1 = -1.8$ and heteroscedastic errors:

$$y_i = 0.2 - 1.8x_i + u_i, \quad u_i \sim \text{Normal}\left(0, \sigma_i^2\right),$$
$$\sigma_i^2 = \sigma^2 \cdot (0.2 + x_i)^2.$$

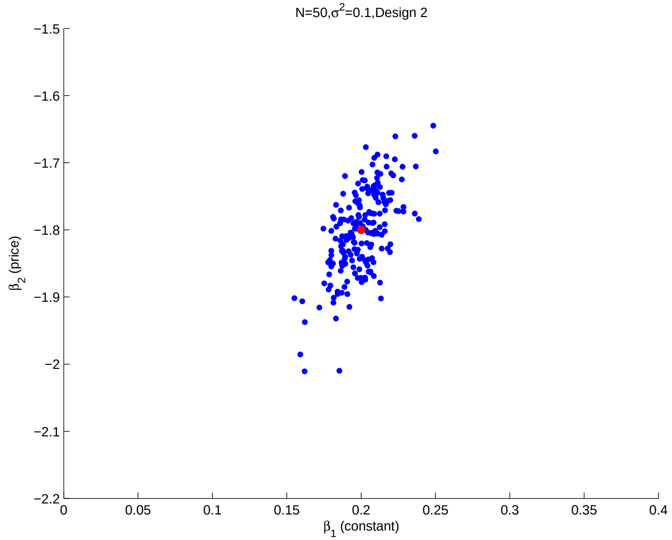
$\Rightarrow$

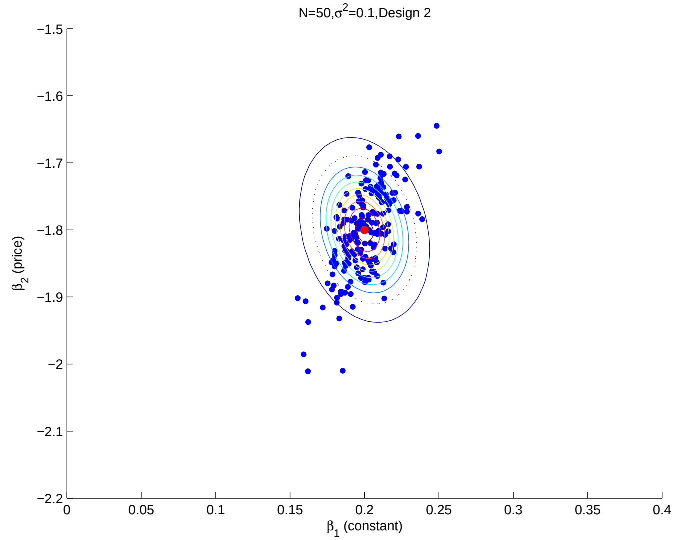
**MATLAB Code:** `reghet.m`

# OLS Estimation under Heteroscedasticity



# OLS Estimation under Heteroscedasticity




Left hand side: estimation errors obtained from a simulation study with 200 data sets (each $N = 50$ observations); right hand side: contours show estimation error according to OLS estimation

# Weighted Least Square Estimation

If the variance increases with an observed variable $Z_i$,

$$\mathrm{Var}\,(u_i) = \sigma_i^2, \quad \sigma_i^2 = \sigma^2 Z_i,$$

then a simple transformation leads to a model with homoscedastic variances:

$$u_i^\star = \frac{u_i}{\sqrt{Z_i}},$$
$$\mathrm{Var}\,(u_i^\star) = \sigma^2.$$

# Weighted Least Square Estimation

Therefore a simple transformation of regression model

$$y_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_K x_{K,i} + u_i,$$

leads to a model with homoscedastic variances:

$$\frac{y_i}{\sqrt{Z_i}} = \beta_0 \frac{1}{\sqrt{Z_i}} + \beta_1 \frac{x_{1,i}}{\sqrt{Z_i}} + \ldots + \beta_K \frac{x_{K,i}}{\sqrt{Z_i}} + u_i^\star. \qquad (76)$$

Regression model (76) has identical parameters as the original model, but a transformed response variable as well as transformed predictors.

# Weighted Least Square Estimation

Rewrite model (76) as

$$y_i^\star = \beta_0 x_{0,i}^\star + \beta_1 x_{1,i}^\star + \ldots + \beta_K x_{K,i}^\star + u_i^\star, \tag{77}$$

where

$$y_i^\star = \frac{y_i}{\sqrt{Z_i}}, \qquad x_{0,i}^\star = \frac{1}{\sqrt{Z_i}},$$
$$x_{j,i}^\star = \frac{x_{j,i}}{\sqrt{Z_i}}, \qquad \forall j = 1, \ldots, K.$$

Note that model (77) fulfills assumption (39), i.e. it is a model with homoscedastic errors.

# Weighted Least Square Estimation

Use OLS estimation for the transformed model (77):

$$y_i^\star = \beta_0 x_{0,i}^\star + \beta_1 x_{1,i}^\star + \ldots + \beta_K x_{K,i}^\star + u_i^\star,$$

and minimize the sum of squared residuals in the transformed model:

$$\mathrm{SSR} = \sum_{i=1}^{N} (u_i^\star)^2 .$$

# Weighted Least Square Estimation

Due to the following relation

$$u_i^\star = \frac{u_i}{\sqrt{Z_i}},$$

the OLS estimator of the transformed model is equal to a weighted least square estimator in the original model:

$$\mathrm{SSR} = \sum_{i=1}^{N} (u_i^\star)^2 = \sum_{i=1}^{N} u_i^2 \frac{1}{Z_i}.$$

# Weighted Least Square Estimation

Residuals $u_i$ for observations with big variances are down-weighted, while residuals for observations with small variances obtain a higher weight. Hence the name weighted least square estimation.

There is no "intercept" in the model (77), only covariates. Using the matrix formulation of the multiple regression model (77), we obtain following matrix of predictors and observation vector:

$$\mathbf{X}^\star = \mathrm{Diag}\,(w_1 \cdots w_N)\,\mathbf{X}, \qquad \mathbf{y}^\star = \mathrm{Diag}\,(w_1 \cdots w_N)\,\mathbf{y}.$$

where

$$w_i = \frac{1}{\sqrt{Z_i}}, \qquad i = 1, \ldots, N.$$

# Weighted Least Square Estimation

The OLS estimator is computed for the transformed model, i.e.

$$\hat{\boldsymbol{\beta}} = ((\mathbf{X}^\star)'\mathbf{X}^\star)^{-1}(\mathbf{X}^\star)'\mathbf{y}^\star.$$

This is equal to following WLS estimator, which is expressed entirely in terms of the original variables:

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{y}, \tag{78}$$

where $\mathbf{W} = \mathrm{Diag}\left(w_1^2 \cdots w_N^2\right)$.

# Testing for Heteroscedasticity

- Classical tests of heteroscedasticity are based on the squared OLS-residuals $\hat{u}_i^2$, e.g. the White or the Breusch-Pagan heteroscedasticity test: test for dependence of the squared residuals on any of the predictor variables using a regression type model:

$$\hat{u}_i^2 = \gamma_0 + \gamma_1 x_{1,i} + \ldots + \gamma_K x_{K,i} + \xi_i,$$

  and test, if $\gamma_0 = \ldots = \gamma_K = 0$ using the F-test.

- Problem: test not reliable, as the errors $\xi_i$ of this regression model are not normal

# Case Study Profit

Demonstration in EVIEWS, workfile `profit`:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i.$$

- Discuss classical tests of heteroscedasticity
- Possible choice for $Z_i$: $Z_i = x_{2,i}$ (`um94`)
- Show how to estimate the transformed model
- Perform residual diagnostics for the transformed model