Econometrics I

Workshop V

Mar 7, 2023

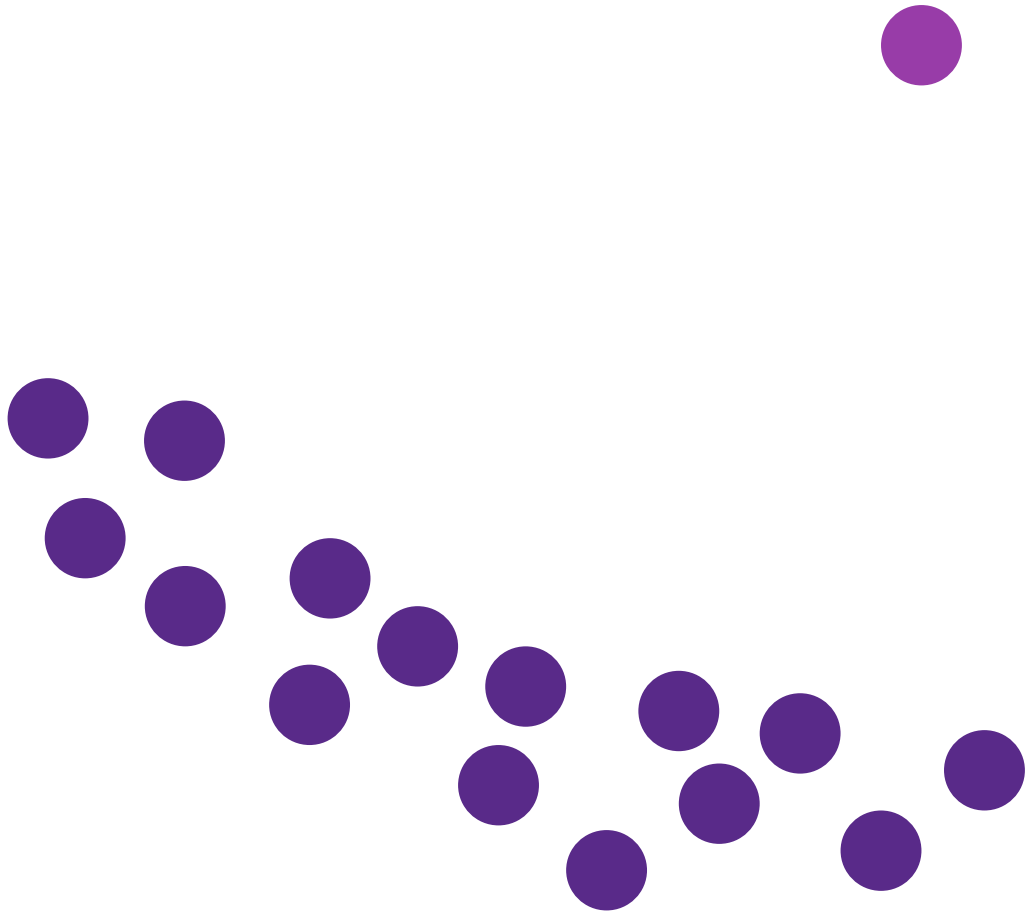
Homoscedasticity

First of all…Why is it important the assumption of normality?

Normality in residuals ensure that OLS estimator is consistent and efficient.

Several tests such as t and F are calculated from the normal distribution assumption.

However, in practice we find some distribution shocks…



If there is an outlier caused by a special situation that is outside from the model, it can provoke a disruption in error distribution.

A datapoint is considered an *outlier* if the value for that point for any variable substantially differs from the rest of the observation's pattern.

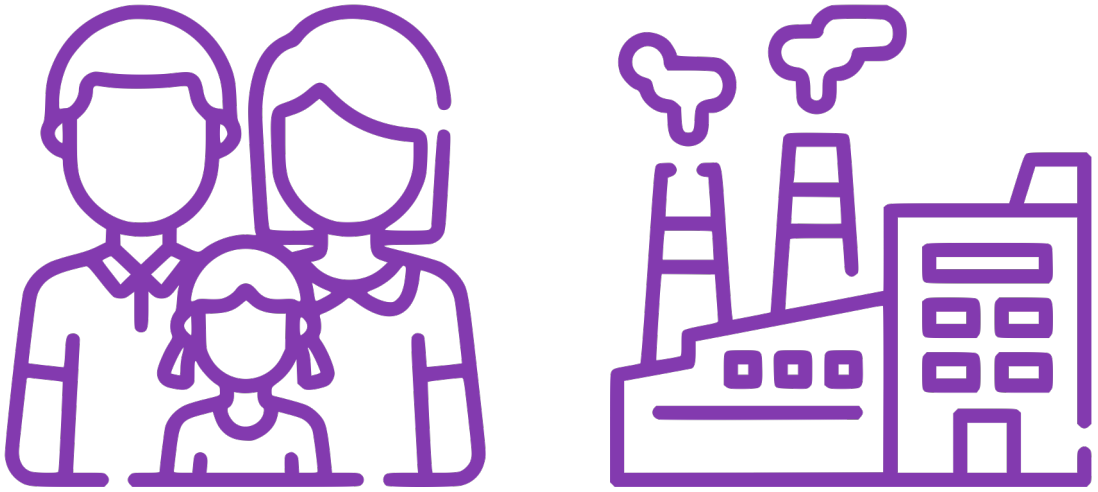
Now, the question is…What can we do with these anormal datapoints?

If you have paid enough attention to classes, you must remember that one of the assumptions under linear regression was:

Homogeneity in residuals' variance

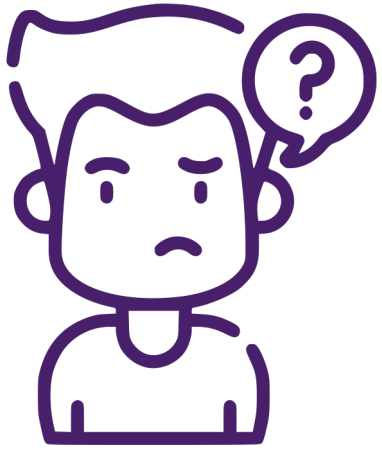
If variance in residuals is not constant, then variance in residuals expose heteroscedasticity

It is quite usual in cross-sectional data



We work with members of a population in a specific moment, such as families or industries, which can have different sizes.

We've got two cases:



Pure Heteroscedasticity

We specify the correct model and yet we observe non-constant variance in residuals

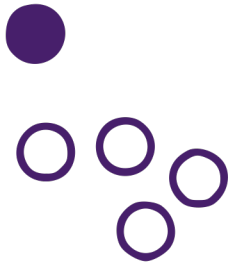
Impure Heteroscedasticity

We incorrectly specify the model, causing the non-constant variace

Causes of **homoscedasticity**

Explanatory variables with an asymetric distribution.

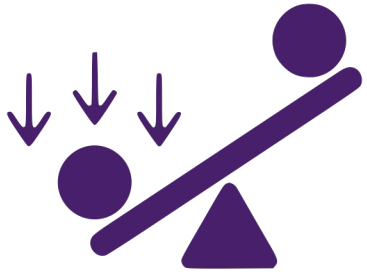
When we omit a variable, it will rely in stochastic term, perhaps causing its own variation.



Outliers necessarily imply an imbalance in disturbance variance.

An outlier can be considered as a sampling element belonging to another distribution (different variance.)

Consequences of **homoscedasticity**



Low heteroscedasticity (Estimator errors are biased)

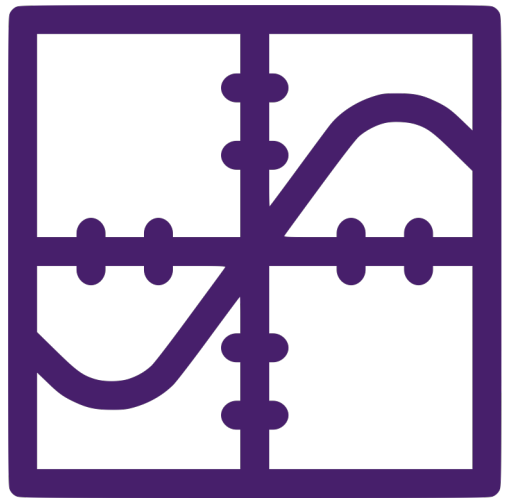
In presence of heteroscedasticity usual statistics in hypothesis testing under the Gauss-Markov assumptions are not applicable.

Due to $var(u|X)$ is not longer constant, OLS estimator is not BLUE and not asymptotically efficient.

It is possible to find estimators that are more efficient than the OLS estimator, but it is necessary to kwow the shape of heteroscedasticity

Defective estimate of parameters

OLS estimator keeps linear, unbiasedness and consistent but is not longer efficient. Homoscedasticity of error term does not play any relevant role in biasedness or consistency



Incorrect computation of variances and inefficient parameters

Variances of OLS estimator despite the fact of not being minimal it cannot be calculated with the expression used in presence of homoscedasticity

Let's prove that (`api00`) depends on free meals percentage given to students (`meals`), students currently learning English (`ell`) and percentage professors with recent accreditations (`emer`)

```
// Database: elemapi2.dta

// Run regression
regress api00 meals ell emer

// Plots for regress
rvfplot, yline(0)
```

```
// Rename variables
rename oldvar newvar

// Change their label
label variable variablename labelname
```

```
// Run regression
regress api00 meals ell emer

// Apply Breusch-Pagan
// (Null hypothesis assumes variance in errors is constant)
estat hettest
```

Advantages

- Easy to apply
- Does not require to know the functional form of heteroscedasticity

Disadvantages

- Relies on the error normality assumption
- Auxiliary equation is not exempted of specification errors from any regression

```
// Run regression
regress api00 meals ell emer

// Apply White test
// (Null hypothesis assumes variance in errors is constant)
estat hettest, white
```

Advantages

- It's a general test
- Easy to apply

Disadvantages

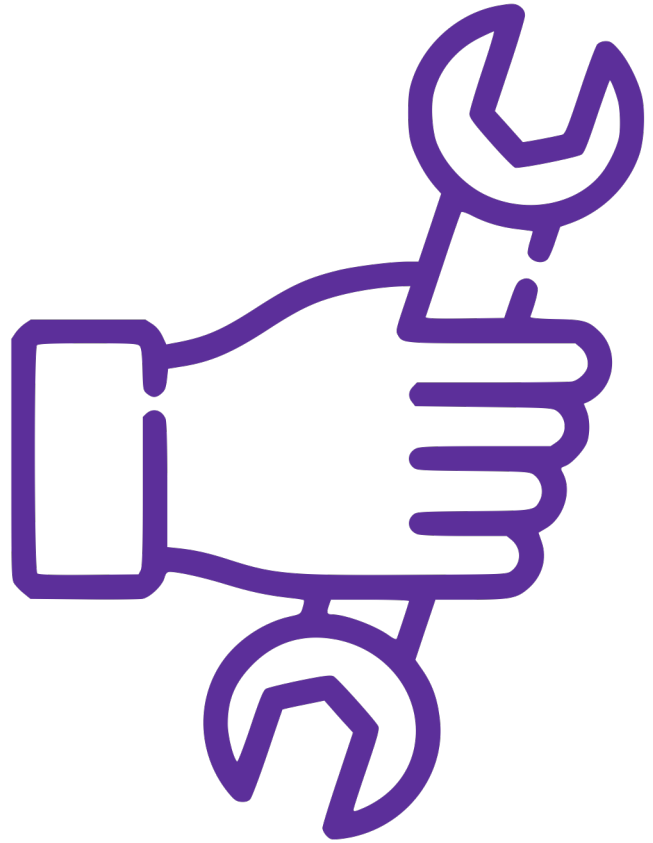
- Auxiliary equations may include too many independent variables
- Auxiliary equation is not exempted of specification errors from any regression

How to fix it?

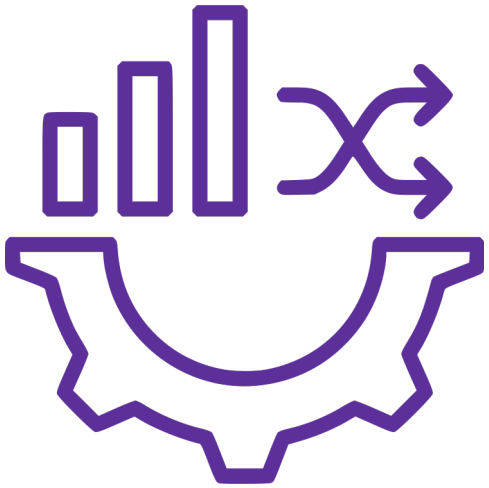
In presence of heteroscedasticity, the OLS estimator is linear, unbiasedness and consistent, but not efficient.

We can add the omitted variable, or to modify the structural form, to model with robust errors, or even to change the estimation method.

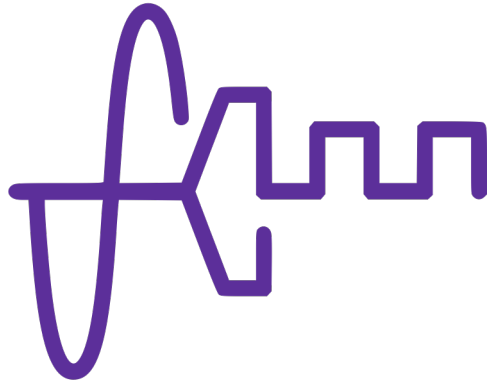
In the case of pure heteroscedasticity, the remedy may be complex…



Dealing with heteroscedasticity



Specify the model again / transform variables



Use Standard Robust Errors

```
// Data: hprice1
// Run regression
reg price lotsize sqrft bdrms. + heteroscedasticity price
```

An advantage for using logarithmic form of dependent variable is that it may reduce heteroscedasticity.

References

- **Salvatore, D., & Sarmiento, J. C.** (1983). *Econometría* (No. HB141 S39). McGraw-Hill.
- **Gujarati, D. N.** (2009). Basic econometrics. Tata McGraw-Hill Education.
- **Wooldridge, J.M.** (2016). *Introductory Econometrics*, Cengage Learning, 6th edition.