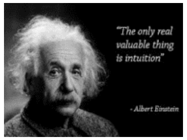
Weighted Least Squares (WLS)

EV(✘) – Error variance is not constant

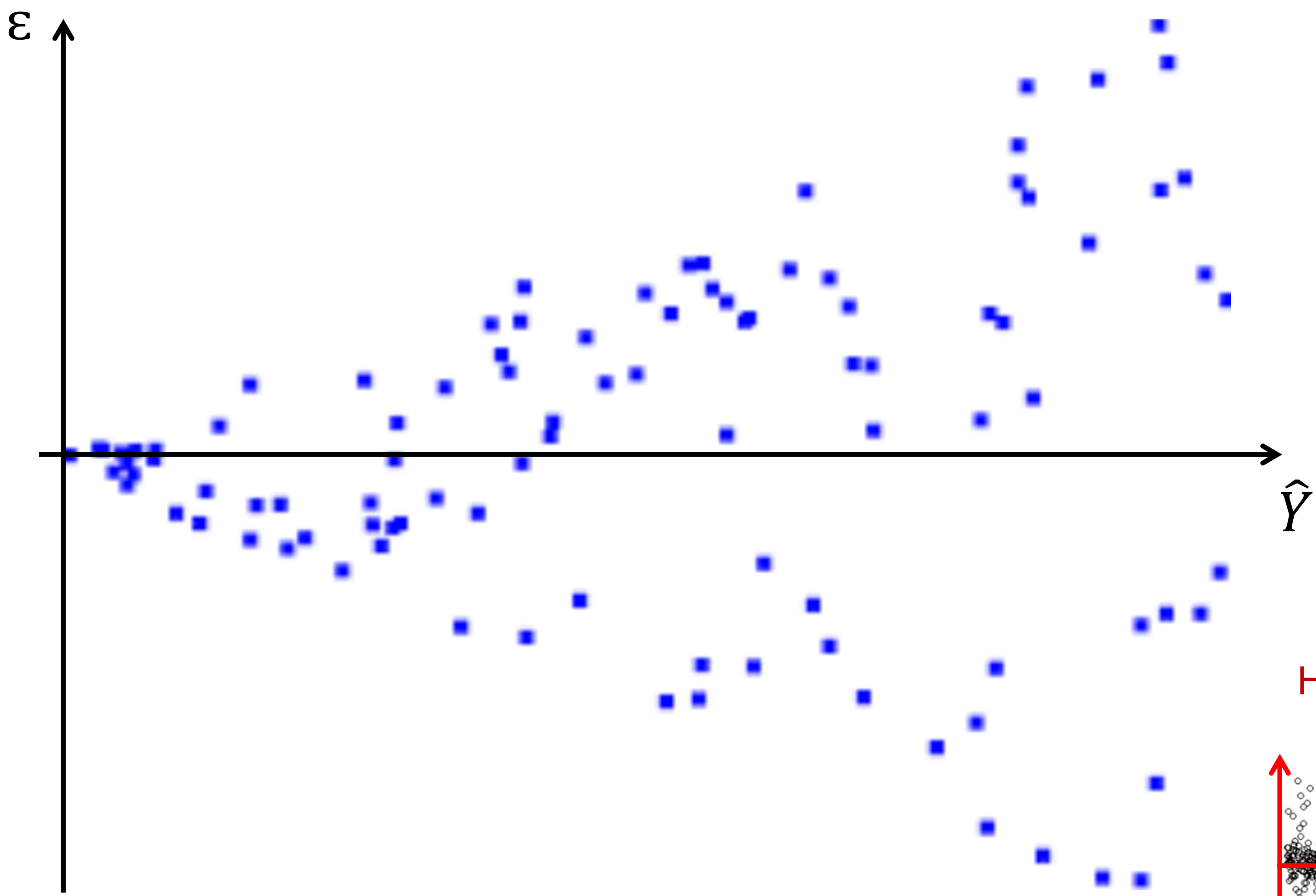
(Note: this notation indicates the OLS assumption not met, resolved by the featured model)

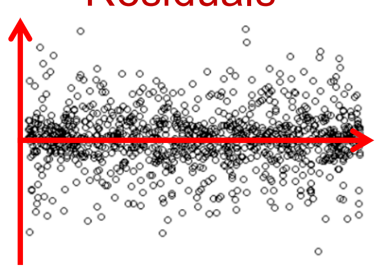


Heteroscedasticity: Intuition

- The **OLS** estimation method fits a line that **minimizes the SSE**.
- But if the errors **grow** or **shrink** systematically along one of the predictor variables it causes the squared errors to be too large (or too small) in some parts of the regression line
- If so, **OLS** is not the **most efficient** (least variance) estimator
- This problem of **"uneven"** error variance is referred to as **"heteroscedasticity"** or **"non-spherical residuals"**
- If you plot the errors along the predicted values $\hat{Y}$, the errors should look like an even cloud – i.e., **"homoscedastic"** and not show a pattern.
- If you have a **business reason** to believe that errors increase or decrease with one or more variables (e.g., as people learn they make smaller errors; there outliers in the data; the closer to the city the more unpredictable the traffic) you should inspect for heteroscedasticity.
- The presence of heteroscedasticity can be **tested**

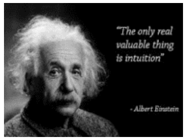
Heteroscedasticity Illustration





Testing for Heteroscedasticity

- Intuitively, heteroscedasticity is **easy to identify**
- You can do it **visually** by plotting residuals against predicted values and observe if there is a **pattern** on the residuals or a cloud
- There are some **formal tests** for Heteroskedasticity
- Most of these tests use a **regression** method
- Regress the residuals on the predicted values → $\varepsilon = \boldsymbol{\beta_0} + \boldsymbol{\beta_{\hat{Y}}}(\hat{Y})$
- There are various test, but the most common are:
  - ➢ **Breusch-Pagan Test:** run the regression above and if the **p-value** of the residual regression is significant, then the errors are correlated with the predicted values → heteroscedasticity is present
  - ➢ **White's Test:** examine the **R^2** of the residual regression
- Again, **R** does all of this for us



Weighted Least Squares (WLS): Intuition

- Regression models with heteroskedastic residuals can be easily **corrected** with **WLS**
- WLS is a generic regression estimation method that minimizes a **"weighted" sum of squares**, rather than just the sum of squares.
- The challenge is to find the **appropriate weights** to use with WLS. You may have a number of business reasons for selecting a particular set of weights.
- **The standard method for WLS to correct for heteroscedasticity is to first run OLS, then compute the errors, and then use the inverse of the squared errors as the weights**
- This method penalizes observations with **large errors** by giving them **lower weight** in the weighted SSE calculation
- Fortunately, **R** takes care of the WLS weighting for us
- Before you go through this trouble, better **test** for Heteroskedasticity

R Tips

`install.packages("lmtest")` → This package has several tests for lm objects, including `bptest()` below

`bptest(lm.fit, data=dataName)` → To perform a Breusch-Pagan test for heteroscedasticity

Use WLS if `p-value` in `bptest` is significant → Errors are heteroscedastic

WLS in 2 Steps:

1. `lm.fit <- lm(y~x1+x2+etc.,data=dataName)` → Fit linear
2. `lm.fit.wls=lm(y~x1+x2+etc.,data=dataName, weights=1/lm.fit$residuals^2)` → Use the squared value of the inverse residuals of the OLS model for the new weighted regression

KOGOD SCHOOL
of
BUSINESS