Econometrics **August 18, 2026**

CLRM (Gauss-Markov) Assumptions

Instructor: Kalyan Kolukuluri *Content: Lecture Notes*

Contents

1 CLRM Assumptions **3**

2 Assumption 1: Linear regression model. **3**

3 Assumption 2: Explanatory (X) are non-stochastic **4**

4 Assumption 3: No Perfect collinearity **4**

5 Assumption 4: No Endogeneity (Zero Conditional Mean) **4**

6 Assumption 5:Homoskedasticity **5**

6.1 Heteroskedasticity . . . 5

6.2 Visualizing Heteroskedastic Population . . . 5

6.3 Consequences of Heteroskedasticity . . . 6

6.4 Detect Heteroskedasticity (Visual) . . . 6

6.5 Detect Heteroskedasticity - Statistical tests . . . 6

6.5.1 Breusch-Pagan Test . . . 7

6.5.2 White's Test . . . 8

6.5.3 Goldfeld-Quandt Test . . . 8

6.6 Remedy for Heteroskedasticity . . . 9

7 Assumption 6: No Serial Dependence of disturbances **9**

8 Assumption 7: Normal errors **10**

8.1 Why the normality assumption? . . . 10
8.2 Consequences of non-normal (Unbiased, yet, inefficient $\hat{\beta}$) . . . 11
8.3 Detect non-normal (Visual) . . . 12
8.4 Detect non-normal errors - Statistical tests . . . 13
8.5 Addressing non-normal errors . . . 13
8.6 Addressing non-normality . . . 13
8.7 Dealing with non-normality . . . 13

9 Summary **15**

1 CLRM Assumptions

CLRM Assumptions of the Population Linear Model

1. **MLR A1:** Linear regression model. The regression model is linear in the parameters
2. **MLR A2:** X values are fixed in repeated sampling. i.e. X is *non-stochastic*.
3. **MLR A3: No perfect multicollinearity:** NO perfect linear relationship between X's.
4. **MLR A4:** Zero mean value of disturbance $\mathbb{E}(\varepsilon_i \mid X_i) = 0 \implies \text{Cov}(X_i, \varepsilon_i) = 0$. No specification bias or specification error in the model.
5. **MLR A5:** *Homoskedasticity* $\text{Var}(\varepsilon_i|X_i) = \text{Var}(\varepsilon_i) = \sigma_\varepsilon^2$ Not all Y's are equally *reliable*. So, we woould prefer to sample X's close to $E(Y_i|X_i)$ and this will induce higher $\text{var}(\hat{\beta}_1)$
6. **MLR A6: Normality of** ε_i. Since distribution of β_0, β_1 is dependent on distribution of ε_i, any hypothesis tests are only valid as long as $\varepsilon_i \sim N(0, \sigma_\varepsilon^2)$
7. **MLR A7: No autocorrelation** between the disturbances. $\text{Cov}(\varepsilon_i, u_j|X_i, X_j) = 0$ The observations are sampled independently [1]

MLR A1 to MLR A5 are called Gauss Markov Assumptions.

Theorem 1 (Gauss Markov Theorem) *Given the assumptions of the CLRM, the least squares estimators in the class of unbiased linear estimators have minimum variance, i.e., they are **BLUE**.*

2 Assumption 1: Linear regression model.

MLR A1: Linear regression model.

Linear regression refers to a specification that is linear in the parameters.

$\mathbb{E}[Y_i \mid X_i] = \alpha + \beta X_i^2$
$\mathbb{E}[\log Y_i \mid X_i] = \alpha + \beta (Y_i)$

Both specify linear regressions.

[1] Challenging for Time Series data

3 Assumption 2: Explanatory (X) are non-stochastic

MLR A2: Regressor (X) are non-stochastic

X values are fixed in repeated sampling. i.e. X is *non-stochastic*. X values are independent of the error term, because they are considered *fixed* in repeated samples. In experimental settings, this is plausible as Xs are in control of the experimenter.
Even if X variables are stochastic, the OLS model mimics the *fixed regressor* mode; and many properties of OLS continue to hold.

4 Assumption 3: No Perfect collinearity

MLR A3: No Perfect collinearity

In the *sample and therefore in the population*, none of the independent variables is constant, and there are no exact linear relationships among the independent variables.

5 Assumption 4: No Endogeneity (Zero Conditional Mean)

MLR A4: Zero Conditional Mean of disturbance

Zero mean value of disturbance $\mathbb{E}(\varepsilon_i \mid X_i) = 0 \implies \text{Cov}(X_i, \varepsilon_i) = 0$. No specification bias or specification error in the model.
If X is non-stochastic, $\mathbb{E}[\varepsilon_i] = 0$. Factors not explicitly included in the model and therefore subsumed into ε_i do not systematically affect the mean of Y.

Note: If $\mathbb{E}(\varepsilon_i \mid X_i) = 0$ then $\text{Cov}(X_i, \varepsilon_i) = 0$ and so the two variables are uncorrelated. The *vice-versa* is not always true. When we assume X and ε are uncorrelated, then they may have separate and additive influences on Y. But, if X and ε are correlated, it is not possible to assess their individual effects on Y.

6 Assumption 5:Homoskedasticity

A5: Homoskedasticity

One of the CLRM model assumptions is the homoskedasticity of the residuals in the population.

$$\text{Var}(\varepsilon) = \sigma_\varepsilon^2 = \text{constant}$$

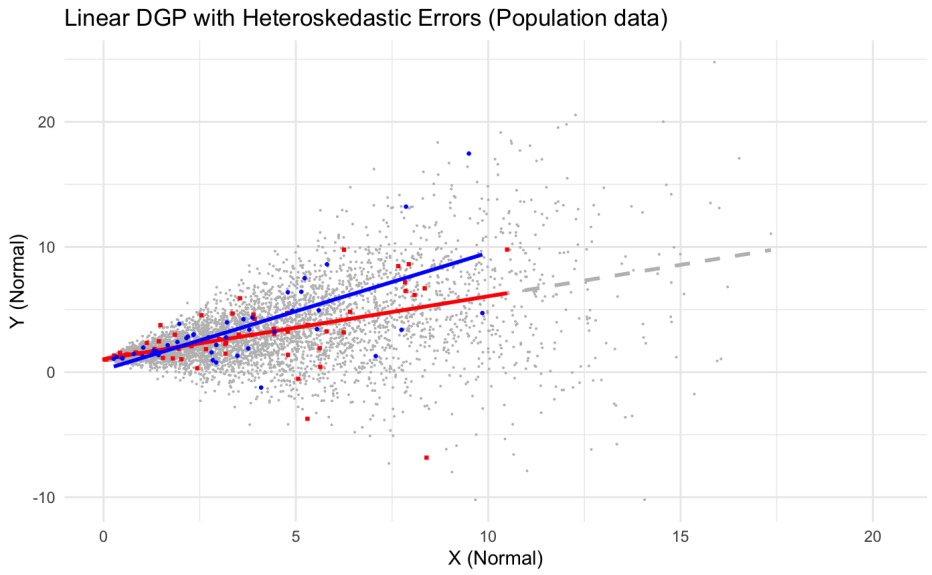
Why the homoskedastic assumption? This makes the different Y's that shall be sampled to be *unreliable*. Realiability is judged by how closely or distantly, the Y values are distributed around the conditional means ($E(Y \mid X)$). Under heteroskedasticity, we would prefer to sample from points lying close to the conditional means, rather than data points spread wide away from the mean.

6.1 Heteroskedasticity

If $\text{Var}(\varepsilon \mid X_i) = f(X_i) = \sigma_i^2 =$ **say,** $\sqrt{X_i}$ **in the population. This is called *heteroskedasticity*.**

6.2 Visualizing Heteroskedastic Population

Figure 1: Population with heteroskedastic errors



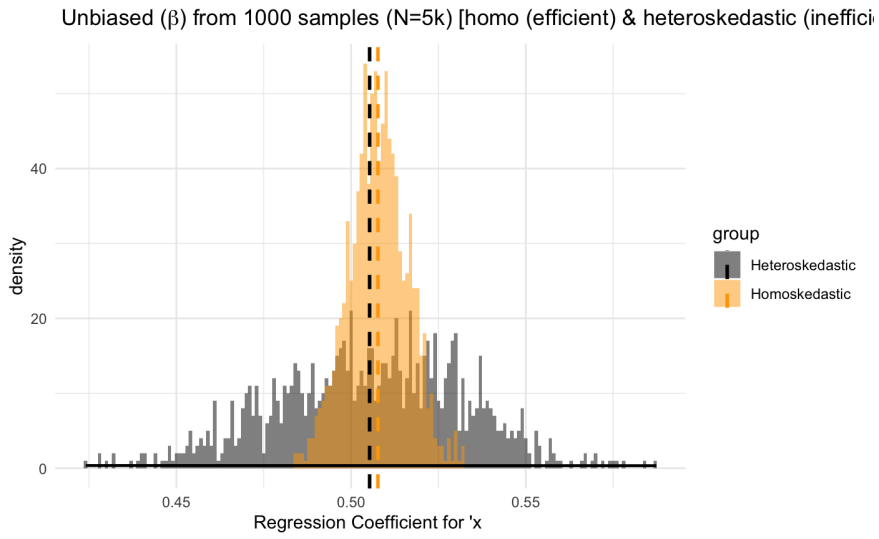
6.3 Consequences of Heteroskedasticity

Table 1: Consequence of Heteroskedasticity

Estimator	Unbiased	Efficiency	Consistent
$\hat{\beta}_0$	Y	N	Y
$\hat{\beta}_1$	Y	N	Y
S.E($\hat{\beta}_0$)			N
S.E($\hat{\beta}_1$)	N	N	N

Consequences of Heteroskedasticity (Unbiased, yet, inefficient $\hat{\beta}$)

Figure 2: Unbiased, yet, inefficient $\hat{\beta}_1$ estimates, in heteroskedastic data



Consequences of Heteroskedasticity (Inconsistent S.E($\hat{\beta}$))

If S.E($\hat{\beta}$) is inconsistent, then it is invalid to use the regular estimate. No hypothesis test would be valid, i.e., the hypothesis test will not indicate what it is supposed to indicate.

6.4 Detect Heteroskedasticity (Visual)

Quantile Plots of residuals and residuals vs fitted values will visually reveal pattern of heteroskedasticity. Estimate the model with simple OLS and do post-estimation checks visually on the spread of residuals.

6.5 Detect Heteroskedasticity - Statistical tests

Figure 3: Unbiased, yet, inefficient $\hat{\beta}_1$ estimates, in heteroskedastic data

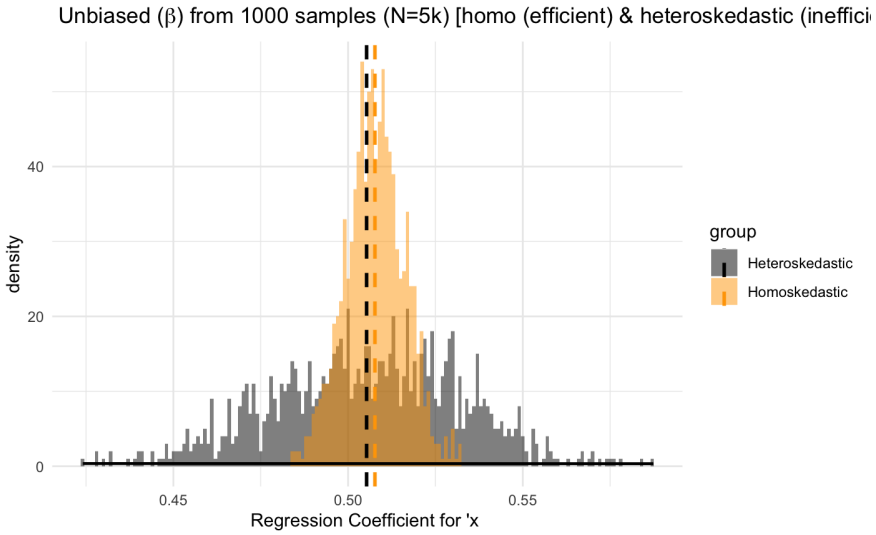
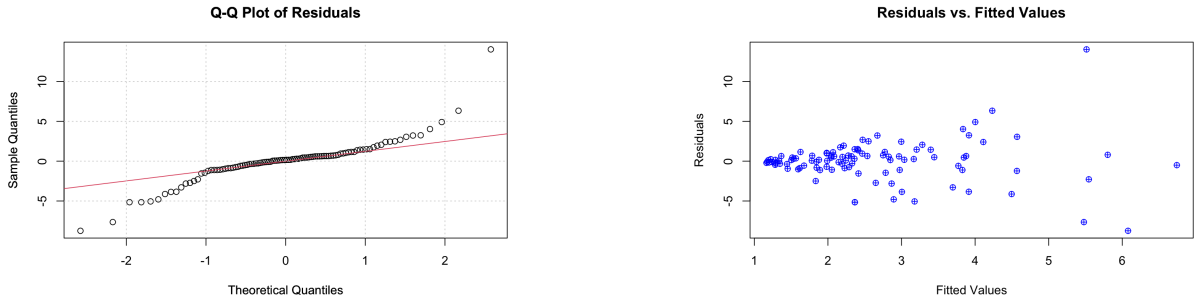


Figure 6: Visual detection of heteroskedasticity



6.5.1 Breusch-Pagan Test

The Breusch-Pagan test is a popular test for detecting heteroskedasticity. It assumes that the error variance can be modeled as a linear function of one or more independent variables.

Hypotheses:

H_0: Homoskedasticity (variance is constant)
H_1: Heteroskedasticity (variance is not constant)

Breusch-Pagan Procedure:

1. Estimate the regression model and obtain the residuals (ε_i^2).
2. Regress the squared residuals (ε_i^2) on the suspected independent variables ($X_1, X_2, ..., X_p$). These X variables can be the same as the original independent variables or different ones.

3. Calculate the test statistic, which is $N \cdot R^2$ from the auxiliary regression, where N is the sample size.
4. The test statistic follows a chi-squared distribution with p degrees of freedom under the null hypothesis of homoskedasticity.
5. If the calculated test statistic exceeds the critical value from the chi-squared distribution, we reject the null hypothesis and conclude that heteroskedasticity is present.

6.5.2 White's Test

The White's test is a more general test for heteroskedasticity that does not require a specific functional form for the variance.

Hypotheses:

H_0: Homoskedasticity
H_1: Heteroskedasticity

White Test Procedure:

1. Estimate the regression model and obtain the residuals (ε_i^2).
2. Regress the squared residuals (ε_i^2) on the original independent variables, their squares, and their cross-products.
3. Calculate the test statistic, which is $n \cdot R^2$ from the auxiliary regression.
4. The test statistic follows a chi-squared distribution with degrees of freedom equal to the number of regressors in the auxiliary regression (including squares and cross-products).
5. If the calculated test statistic exceeds the critical value, we reject the null hypothesis.

6.5.3 Goldfeld-Quandt Test

The Goldfeld-Quandt test is applicable when the suspected source of heteroskedasticity is related to one specific independent variable.

Hypotheses:

H_0: Homoskedasticity
H_1: Heteroskedasticity

Procedure:

1. Order the observations according to the suspected variable.
2. Divide the sample into two groups (e.g., high and low values of the variable).

3. Estimate separate regressions for each group.

4. Calculate the ratio of the residual sum of squares (RSS) from the two regressions: $F = \frac{RSS_2/(n_2-k)}{RSS_1/(n_1-k)}$, where RSS_1 and RSS_2 are the RSS for group 1 and 2, respectively, n_1 and n_2 are the sample sizes, and k is the number of parameters in the model.

5. The test statistic follows an F-distribution with $(n_2 - k)$ and $(n_1 - k)$ degrees of freedom.

6.6 Remedy for Heteroskedasticity

If heteroskedasticity is detected, it can lead to incorrect inferences. Several remedies exist, including using heteroskedasticity-consistent standard errors (e.g., White's robust standard errors) or transforming the dependent variable.

1. $\text{HC0} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\text{diag}\left[e_i^2\right]\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$ -Huber-White std. errors

2. $\text{HC1} = \left(\frac{N}{N-K}\text{HC0}\right)$

3. $\text{HC2} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\text{diag}\left[\frac{e_i^2}{1-h_{ii}}\right]\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$; Leverage $h_{ii} = x_i\,(\boldsymbol{X}'\boldsymbol{X})^{-1}\,x_i'$

4. $\text{HC3} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\text{diag}\left[\frac{e_i^2}{(1-h_{ii})^2}\right]\boldsymbol{X}\,(\boldsymbol{X}'\boldsymbol{X})^{-1}$

7 Assumption 6: No Serial Dependence of disturbances

A6: No Serial Dependence

This assumption can also be called *no autocorrelation between disturbances* assumption. Given, any two X records, say, X_i, X_j, the correlation between any two errors/ disturbances, $\varepsilon_i, \varepsilon_j$ is *zero*. In short, the observations are sampled independently.

$$Cov(\varepsilon_i, \varepsilon_j \mid X_i, X_j) = 0 \tag{1}$$

$$Cov(\varepsilon_i, \varepsilon_j) = 0 \quad \text{if } X \text{ is non-stochastic} \tag{2}$$

If data is time-series, this assumption is difficult to maintain. Positive ε_i is followed by a positive ε_j; negative errors are followed by negative error. Positive ε_i is followed by a negative ε_j and otherwise.

8 Assumption 7: Normal errors

MLR A7: Normal errors

One of the assumptions of the CLRM model is the normality of the residuals in the population.

$$\varepsilon \sim N(0, \sigma_\varepsilon^2)$$

Since,

$$Y = \beta_0 + \beta_1(X) + \varepsilon$$

and using $\hat{\beta}_1 = \frac{Cov(Y,X)}{Var(X)} = \sum[\frac{(X-\overline{X})}{\sum(X-\overline{X})^2}]Y$

$$\hat{\beta}_1 = \sum[\frac{(X-\overline{X})}{\sum((X-\overline{X})^2)}][\beta_0 + \beta_1(X) + \varepsilon]$$

Because X, and population parameters β_0, β_1 are non-stochastic, the sample estimate $\hat{\beta}_1$ is a linear function of the random variable ε. **So the nature of probability distribution of ε assumes an important role in hypothesis testing**.

8.1 Why the normality assumption?

a Any linear function of normally distributed variables is itself normally distributed. Now, $\hat{\beta}_0, \hat{\beta}_1$ are linear functions of $\varepsilon \sim N(0, \sigma_\varepsilon^2)$ this implies that, $\hat{\beta}_0 \sim N(0, Var(\hat{\beta}_0)), \hat{\beta}_1 \sim N(0, Var(\hat{\beta}_1))$ are also normally distributed.

b $Z = \frac{\hat{\beta}_0 - \beta_0}{S.E(\hat{\beta}_0)}; Z = \frac{\hat{\beta}_1 - \beta_1}{S.E(\hat{\beta}_1)}$ follows Std. normal distribution.

c So, if we are dealing with a *small or finite sample size, normality plays a critical role.* It helps us derive exact probability of distribution of OLS estimators $(\hat{\beta}_0, \hat{\beta}_1)$.

d If sample size is reasonably large, we may be able to relax the normality assumption.

In addition, Wooldridge (Introductory Econometrics, 3rd Ed, p.168) states the following: "In addition to finite sample properties, it is important to know the asymptotic properties or large sample properties of estimators and test statistics. These properties are not defined for a particular sample size; rather, they are defined as the sample size grows without bound. Fortunately, under the assumptions we have made, OLS has satisfactory large sample properties. One practically important finding is that even without the normality assumption (Assumption MLR.6), t and F statistics have approximately t and F distributions, at least in large sample sizes."

In matrix notation

$$E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$$
$$\text{var}(\hat{\boldsymbol{\beta}}) = \sigma^2 \left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

Thus, $\dfrac{\left(\hat{\beta}_j - \beta_j\right)}{\text{S.D}\left(\hat{\beta}_j\right)} \sim N(0,1)$, where $\text{S.D}\left(\hat{\beta}_j\right) = \sigma\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{jj}}$ is one practical problem with this: We don't know σ. We must replace σ with its estimator $\hat{\sigma}$ to get the standard error of $\hat{\beta}_j$,

$$\text{S.E}\left(\hat{\beta}_j\right) = \hat{\sigma}\sqrt{(\mathbf{X}'\mathbf{X})^{-1}_{jj}}$$

When we do that, the normal distribution becomes a t distribution with $(n-k-1)$ degrees of freedom:

$$\frac{\left(\hat{\beta}_j - \beta_j\right)}{\text{S.E}\left(\hat{\beta}_j\right)} \sim t_{n-k-1}$$

For simple regression, $\text{S.E}\left(\hat{\beta}_1\right) = \dfrac{\hat{\sigma}}{\sqrt{\sum_{i=1}^{n}\left(x_i - \overline{x}\right)^2}}$.

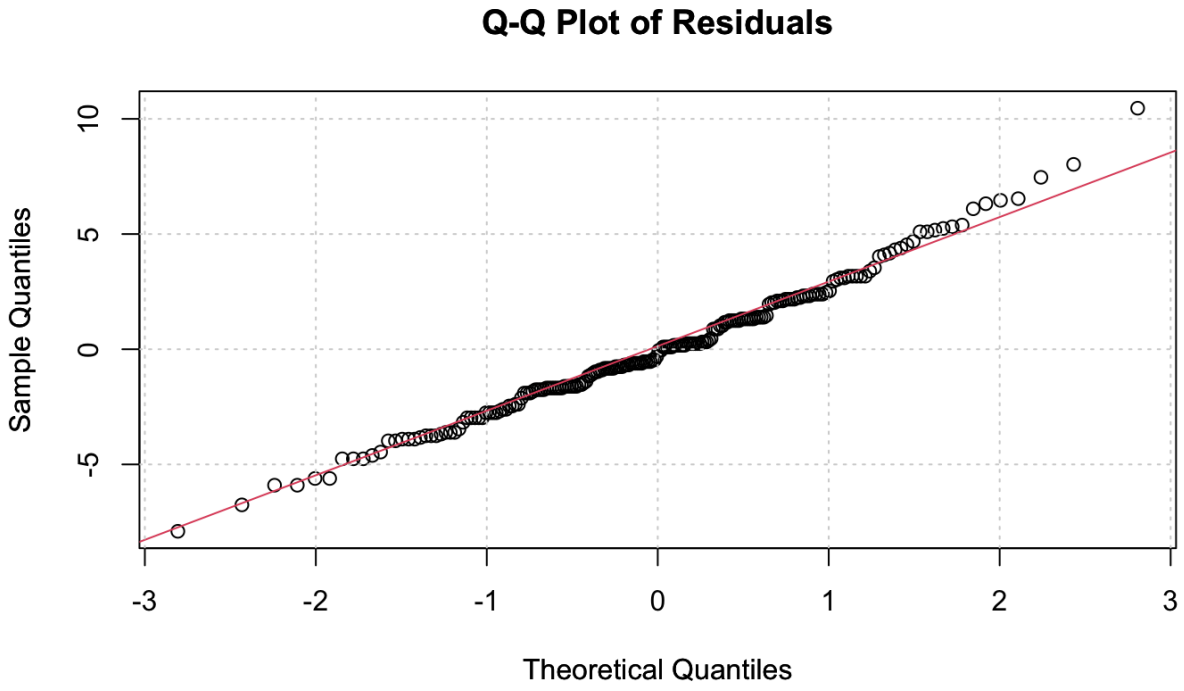
8.2 Consequences of non-normal (Unbiased, yet, inefficient $\hat{\beta}$)

When non-normality in ε is observed, two assumptions in the linear model are potentially unmet.

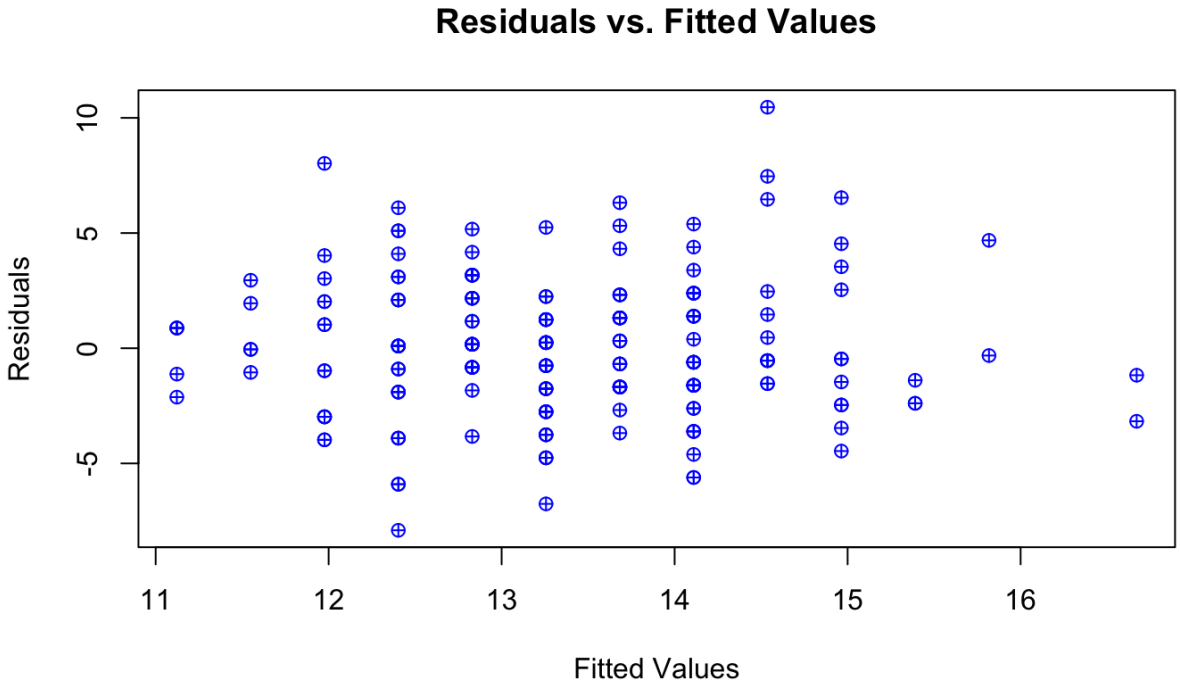
1. First, non-normality in $\hat{\varepsilon}$ suggests non-normality in $\varepsilon \implies$ inaccurate inferential results regarding p-values and CI coverage.

2. Second, the relationship between X and y may not be linear. If the unknown population functional form between X and y is non-linear and a linear model is fit, instead, the estimates of the linear model $\hat{\beta}$ are biased estimates of the unknown population parameters β.

8.3 Detect non-normal (Visual)



0.5



0.5

Estimate model with simple OLS, and do post-estimation checks visually on spread of residuals.

8.4 Detect non-normal errors - Statistical tests

1. Shapiro-Wilk test
2. Kolmogorov-Smirnov test
3. Jacque-Bera test
4. Anderson-Darling test

8.5 Addressing non-normal errors

Method	Linear model	Keep data as is	Non-normality is informative
CLT	✓	✓	x
HCCM	✓	✓	x
Bootstrap	✓	✓	x
Trim or Winsorize	✓	x	✓†
Transform	✓	x	Depends
Rank-based Nonparametric	x	$x \neq$	x
Nonlinear models	x	✓	x

Reproduced from Pek et al. (2018)

8.6 Addressing non-normality

Caution !!!
The process of Winsorizing or trimming the data follows from ordering the data such that extreme data points are replaced by less extreme points or are removed. This ordering creates dependency among the data points, violating the assumption of independence required for OLS estimation.
Winsorized or trimmed means are biased estimates of the arithmetic mean.
Once data is winsorized, i.e. when we believe the outliers are uninformative (nuisance), the dependency created by removing data one must employ ***M-class estimators*** (Max. likelihood type estimators)

8.7 Dealing with non-normality

In order to deal with non-normality in the errors, you could consider the following options:

- Use the $\log(Y)$ for regression

- For time series data use the first difference of the dependent variable, because you might have a stationarity issue.
- Use a different estimator. For example, if your dependent variable can only have positive values, you might be better off with a generalised linear model, which assumes Gamma-distributed errors. Or, for count data, you might want to use a Poisson estimator etc.

9 Summary

			$\hat{\beta}_{\text{OLS}}$		
	Assumptions	**Description**	**Unbiasedness**	**Efficiency**	**Consistency (Asymptotic Unbiasedness)**
Gauss Markov Assumptions MLR.1 to MLR.5	MLR.1	Linear in Parameters	Necessary	Necessary	Necessary
	MLR.2	Random Sampling	Necessary	Necessary	Necessary
	MLR.3	No Perfect Collinearity	Necessary	Necessary	Necessary
	MLR.4	Zero Conditional Mean	Necessary	Necessary	Necessary[a]
	MLR.5	Homoskedasticity	Not Necessary	Necessary	Necessary
	MLR.6	Normality of the error term	Not Necessary	Not Necessary	Not Necessary

[a] For this we need MLR. 4′ :$\mathrm{E}(\varepsilon) = 0 \quad$ and $\quad \mathrm{Cov}(X_j, \varepsilon) = 0 \quad \forall j \in (1, p)$

- Assumption MLR. 4′: $\mathbb{E}(\varepsilon) = 0 \quad$ and $\quad \mathrm{Cov}(X_j, \varepsilon) = 0 \quad \forall j \in (1, p)$, which is a weaker condition than MLR.4. It only suggests for no linear dependence between ε and each X_j.
- Assumption MLR4: $\mathbb{E}(\varepsilon \mid X_1, X_2, \cdots X_p) = 0$ suggests ε is independent of all X_j in any functional form.

References

Pek, J., Wong, O., and Wong, A. C. (2018). How to address non-normality: A taxonomy of approaches, reviewed, and illustrated. *Frontiers in Psychology*, 9:1–17.

The End