Chapter 6

Relaxing assumptions

Flow of study in this chapter

For this chapter, we are going to relaxing some assumptions we imposed since we estimate the first model. These topics will be covered: (1) multicollinearity (2) heteroscedasticity
(3) autocorrelation. For each section, we will explore

› The nature of assumption

› Effects on the estimated coefficients and variance, also standard error.

› How to detect the problem.

› What are remedial measure(s).

Further reading can be found in Gujarati and Porter, Chapter 10-13.

(1) The nature

This is a simplified version, compared to the book, of multicollinearity problem. Let λ_i be a constant, consider the following argument when X_{2i} and X_{3i} are linearly correlated perfectly, or **perfect multicollinearity**.

› $\lambda_2 X_{2i} + \lambda_3 X_{3i} = 0$

when $\lambda_i \neq 0$ for all i simultaneously. Another relation that describe a non-perfect collinearity, or **multicollinearity**, is

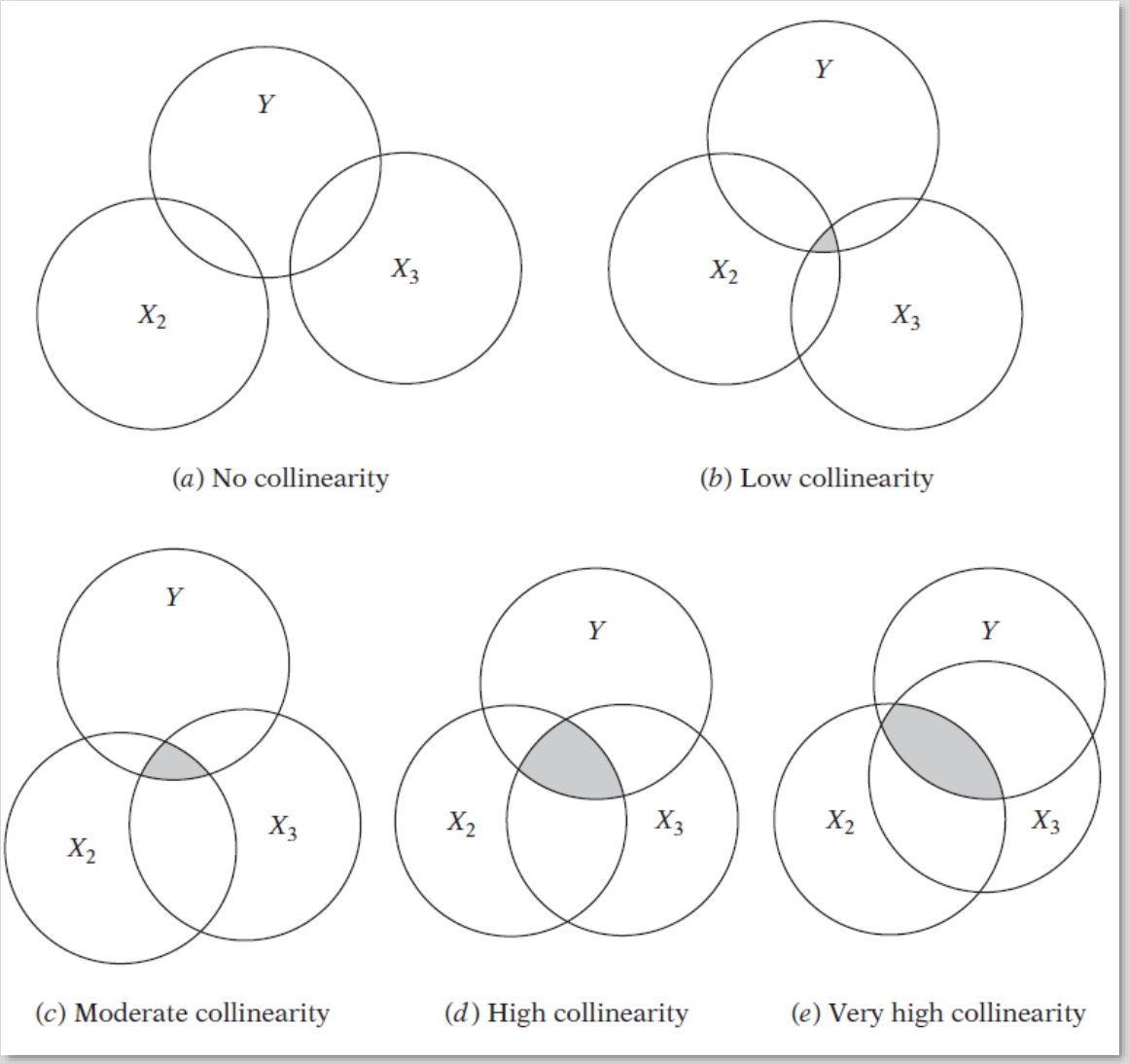
› $\lambda_2 X_{2i} + \lambda_3 X_{3i} + v_i = 0$

where v_i is a stochastic error term. Now assumed that $\lambda_2 \neq 0$ then,

› $X_{2i} = -\frac{\lambda_3}{\lambda_2} X_{3i}$ for the first equation.

› $X_{2i} = -\frac{\lambda_3}{\lambda_2} X_{3i} - v_i$ for the second equation.

(1) The nature


Y
X2
X3
(a) No collinearity
Y
X2
X3
(b) Low collinearity
Y
X2
X3
(c) Moderate collinearity
Y
X2
X3
(d) High collinearity
Y
X2
X3
(e) Very high collinearity

(1) The nature

Precisely taken from the book, here are some causes of multicollinearity.

› ***Data collection method*** may limit range of values taken in the independent variables.

› ***Constraints on the model or in the population being sampled.*** E.g. income and house size tend to be correlated.

› ***Model specification.*** E.g. including a polynomial term especially when the range of X is small.

› ***Overdetermined model*** or when k is larger than n.

› ***Common trend.*** E.g. a time-series data consist of consumption expenditure, income, wealth, and number of population.

(1) The nature

Looking from another perspective apart from stated above, multicollinearity is seen particularly as sampling problem, not a problem on a population, since when we postulate population regression, X variables included in a model have a separate or independent influence on Y.

Meaning that, as Goldberger coined the term, this may be considered as **micronumerosity** problem when our sampling may not be “rich” enough to capture X variability.

By the way, micronumerosity refers to the problem of small sample size.

Another important note is that multicollinearity is **much more common** in cross-sectional data compared to time-series data, in which autocorrelation is much more common.

(2) Effects on estimation

1. Perfect collinearity

Recall that the estimated coefficients are

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2}$$

If we assumed that $X_{3i} = \lambda X_{2i}$, replacing this into $\hat{\beta}_2$, we get,

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2) - (\lambda \sum y_i x_{2i})(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2) - \lambda^2 (\sum x_{2i}^2)^2} = \frac{0}{0}$$

(2) Effects on estimation

Example: Consider the weight model with height (hei_i) as a regressor, now let's create a perfectly correlated variable of height * 2 (defined as $hei2_i$). The model will be

› $wei_i = \beta_1 + \beta_2 sex_i + \beta_3 hei_i + \beta_4 hei2_i + u_i$

Throwing this model into STATA, we have the regression result as follows. We can see that STATA automatically rejects, or omits, one of these variables immediately because $\hat{\beta}_4$ cannot be estimated.

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.sex |  -5.383607   3.396954    -1.58   0.118    -12.16779    1.400578
        hei |   .5914144    .206873     2.86   0.006     .1782605    1.004568
       hei2 |           0  (omitted)
      _cons |  -36.09017    35.8754    -1.01   0.318    -107.7383    35.55795
------------------------------------------------------------------------------
```

(2) Effects on estimation

2. Multicollinearity

Given that $X_{3i} = \lambda X_{2i} + \nu_i$, replacing this into $\hat{\beta}_2$, we get,

- $\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2 + \sum \nu_i^2) - (\lambda \sum y_i x_{2i} + \sum y_i \nu_i)(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2 + \sum \nu_i^2) - \lambda^2 (\sum x_{2i}^2)^2}$

If ν_i is approaching zero, the more this will be closer to perfect multicollinearity.

We are going to skip lots of proof to get to the conclusion what are affected as follows.

(2) Effects on estimation

1. Coefficients estimated are still BLUE.

2. Large variances and covariances.

Recall that the variance of estimated coefficients can be written into this form.

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2(1-\text{r}_{23}^2)}$

› $\text{var}(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2(1-\text{r}_{23}^2)}$

where r_{23}^2 is the coefficient of correlation between X_2 and X_3. The value is between 0 and 1, as an absolute value.

We can see clearly that when the correlation between X_2 and X_3 gets **higher**, the denominator will be **smaller**, leading to **higher** variance.

(2) Effects on estimation

If we separate a part of $\text{var}(\hat{\beta}_2)$ like this,

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} \cdot \frac{1}{(1-r_{23}^2)}$

we can define the latter part as **variance-inflating factor** or VIF

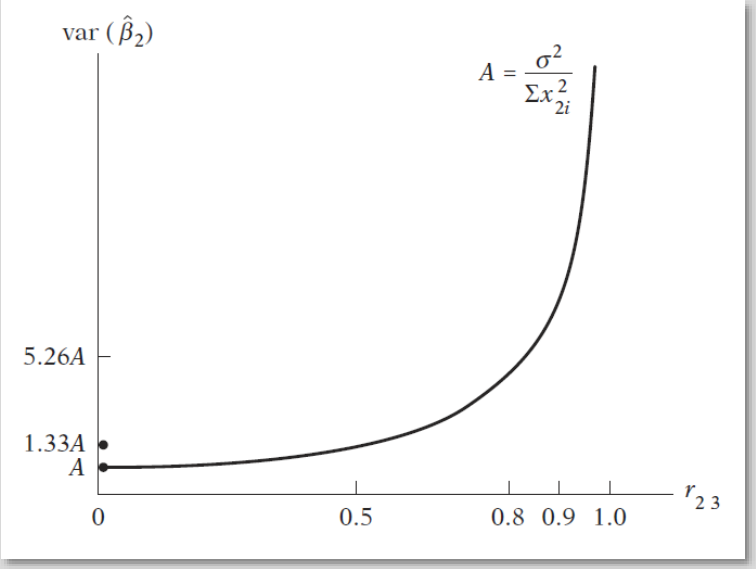
› $\text{VIF} = \frac{1}{(1-r_{23}^2)}$

The higher r_{23}^2, the higher it is for VIF. We can also define the inverse of VIF as **tolerance** or TOL as

› $\text{TOL} = \frac{1}{\text{VIF}} = (1 - r_{23}^2)$

Note that these definition can be generalized for any pair of regressors.

(2) Effects on estimation



3. Coefficients estimated are still BLUE.

Since the variance is used to construct confidence interval, CI is stretched outward and the t-curve is flatter, which will later affect

4. Large variances and covariances.

The acceptance region also becomes larger when variance is high. It is more likely that we would accept the null hypothesis when we should reject, causing more-likely type-II error.

(2) Effects on estimation

Example: Consider the weight model again, the first one take only height (hei_i) as a regressor while the second one include $hein_i$ which is height multiplied by a randomly generated number. The correlation between hei_i and $hein_i$ is exaggerate 0.9928 to the results as follows.

› First model

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.sex |  -5.383607   3.396954    -1.58   0.118    -12.16779    1.400578
        hei |   .5914144    .206873     2.86   0.006     .1782605    1.004568
      _cons |  -36.09017    35.8754    -1.01   0.318    -107.7383    35.55795
------------------------------------------------------------------------------
```

› Second model

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      1.sex |  -5.391169   3.411785    -1.58   0.119    -12.20699    1.424654
        hei |   1.359276   1.180214     1.15   0.254    -.9984733    3.717025
       hein |  -.7848297   1.187455    -0.66   0.511    -3.157043    1.587383
      _cons |  -33.34278   36.27082    -0.92   0.361    -105.8021    39.11651
------------------------------------------------------------------------------
```

(2) Effects on estimation

5. High R^2 but few significant t ratios

The coefficient of determination or R^2 from a model with multicollinearity is likely to be high, also the F-stat of overall model test, since the estimation is 'tricked' to have similar regressors with more explanatory power.

We can also see from the previous example that when we intentionally add a colinear variable into the regression, R^2 is higher.

However, this is not due to more explanatory power of regressors, but multicollinearity. Therefore, each coefficient is not very likely to be significant.

6. Sensitivity of coefficient due to small changes in data.

You can read for this example in page 331. In conclusion, a very slight change in data will affect the value of coefficient tremendously. Some of the direction of coefficient can be different from theoretical speculation.

It would be beneficial to read an illustrative example from page 332 to 337.

(3) Detecting multicollinearity

There are several ways to detect multicollinearity. Some of them are mentioned earlier. Some of them from the book are skipped.

Firstly, the phrase "**Rule of Thumb**" should be introduced. It refers to a specific level of criterion that is usually and mutually accepted as a threshold. More illustrative examples later here.

1. Conflicting test

This is already mentioned that when our estimation reports high R^2 or F value, but rarely coefficients are significant, we should suspect that there might be multicollinearity problem.

2. Pair-wise correlation among regressors

Another easy method to detect multicollinearity is to perform a pair-wise correlation on all regressors. (Easy when using STATA) The rule of thumb suggests that coefficient of correlation **exceeding 0.8** can be problematic and researcher may seek a remedial approach.

(3) Detecting multicollinearity

3. Auxiliary regressions

Regressing X_i on other X variables and obtain R_i^2 from the estimation then calculate

› $F_i = \frac{R^2_{xi \cdot x_2 x_3 \ldots x_k}/(k-2)}{(1-R^2_{xi \cdot x_2 x_3 \ldots x_k})/(n-k+1)}$

where k is the number of independent variables including intercept.

If F_i exceeds critical value from chosen level of significant, it means that X_i is collinear with other X.

Instead of testing all the auxiliary R_i^2, **Klein's rule of thumb** suggests that multicollinearity is troublesome if the R_i^2 is greater than R^2 from another model that we regress Y on these X_i and other X.

4. VIF and TOL

Again, we follow the rule of thumb that VIF should not exceed 10, which will happen when $r_{23}^2 = 0.8$, while TOL should be closer to 1 rather than 0.

(3) Detecting multicollinearity

Example: Using the same weight model of, but add a few more regressors

› $wei_i = \beta_1 + \beta_2 sex_i + \beta_3 hei_i + \beta_4 hein_i + \beta_5 ss_i + \beta_6 exc_i + u_i$

where ss_i is shoe size and exc_i is how many exercising days in a week.

We can ask for a report of VIF and TOL after a regression.

```
------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -2.051504   3.830285    -0.54   0.594    -9.708134    5.605127
         hei |   .7849816   1.206179     0.65   0.518    -1.626135    3.196099
        hein |  -.6364788   1.187903    -0.54   0.594    -3.011063    1.738105
          ss |   2.448109   1.143415     2.14   0.036     .1624558    4.733763
         exc |  -.2271317   .6111166    -0.37   0.711    -1.448737     .994473
       _cons |  -61.16189   38.61168    -1.58   0.118    -138.3455    16.02175
------------------------------------------------------------------------------

    Variable |       VIF       1/VIF
-------------+----------------------
       1.sex |      2.90    0.344756
         hei |     77.56    0.012894
        hein |     73.11    0.013677
          ss |      5.21    0.192038
         exc |      1.10    0.906274
-------------+----------------------
    Mean VIF |     31.98
```

(3) Detecting multicollinearity

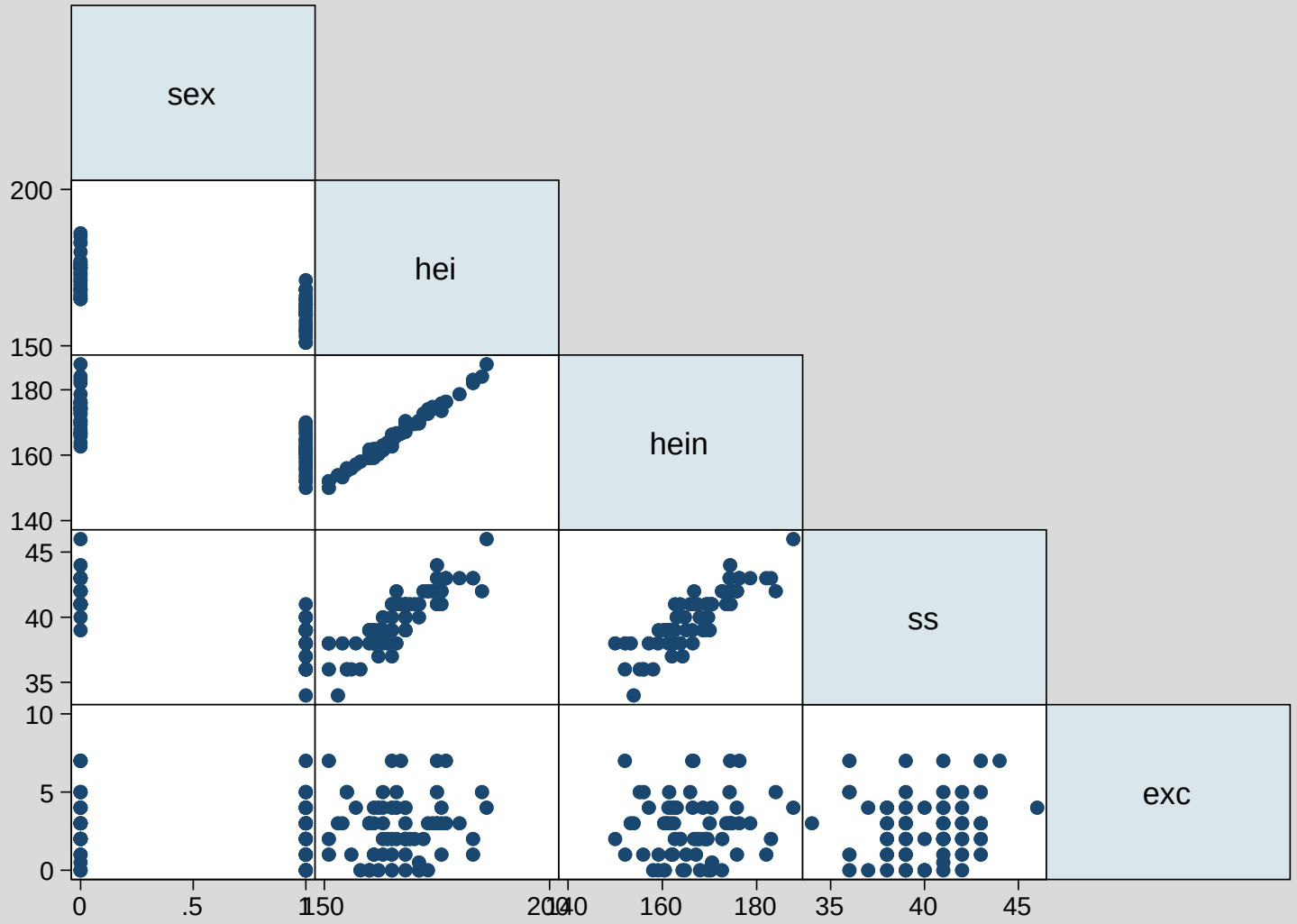
5. Scatter plot

Scatter plot is a good practice revealing linear relationship between two variables, see example below here.

Using the same weight model, we get the correlation matrix here.

	sex	hei	hein	ss	exc
sex	1.0000				
hei	-0.7396	1.0000			
hein	-0.7346	0.9928	1.0000		
ss	-0.7939	0.8708	0.8634	1.0000	
exc	-0.1970	0.0599	0.0831	0.1018	1.0000

(3) Detecting multicollinearity



(4) Remedial measures

1. Do nothing

If and only if when we do not any other choice than using deficient data set. Also, we may not be able to draw any meaningful insight from the regression.

2. Priori information

Supposed we have an income- consumption model as such,

› $Y_i = \beta_1 + \beta_2 income_i + \beta_3 wealth_i + u_i$

we know that income and wealth are highly colinear. If we know that from previous empirical work

› $\beta_3 = 0.1\beta_2$ then

› $Y_i = \beta_1 + \beta_2 income_i + 0.1\beta_2 wealth_i + u_i$ and

› $Y_i = \beta_1 + \beta_2 X_{2i} + u_i$

where $X_{2i} = income_i + 0.1 wealth_i$, we can eliminate one of the variables.

(4) Remedial measures

3. Combining cross-sectional and time-series data

The most completed data would be panel data, repeated samples over time. If that is not possible to obtain, we may use '**pooled-data**', combining multiple waves of data into one large data set.

4. Dropping a variable(s) and specification bias

This is the easiest method, and probably the best if possible. If we are sure which variable should be dropped, according to wrong specification of a model, dropping one of them is very easy and efficient.

Consider dropping $hein_i$ from the show size model, our results would be as follows.

(4) Remedial measures

› First model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(5, 62)        =      8.97
       Model |  3825.50185         5   765.10037   Prob > F        =    0.0000
    Residual |  5289.56936        62  85.3156348   R-squared       =    0.4197
-------------+----------------------------------   Adj R-squared   =    0.3729
       Total |  9115.07121        67  136.045839   Root MSE        =    9.2366

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -2.051504   3.830285    -0.54   0.594    -9.708134    5.605127
         hei |   .7849816   1.206179     0.65   0.518    -1.626135    3.196099
        hein |  -.6364788   1.187903    -0.54   0.594    -3.011063    1.738105
          ss |   2.448109   1.143415     2.14   0.036     .1624558    4.733763
         exc |  -.2271317   .6111166    -0.37   0.711    -1.448737     .994473
       _cons |  -61.16189   38.61168    -1.58   0.118    -138.3455    16.02175
------------------------------------------------------------------------------
```

› Second model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(4, 63)        =     11.27
       Model |  3801.00927         4  950.252317   Prob > F        =    0.0000
    Residual |  5314.06194        63  84.3501895   R-squared       =    0.4170
-------------+----------------------------------   Adj R-squared   =    0.3800
       Total |  9115.07121        67  136.045839   Root MSE        =    9.1842

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -2.110053   3.807001    -0.55   0.581    -9.717738    5.497631
         hei |   .1568177   .2819089     0.56   0.580    -.4065322    .7201677
          ss |   2.462676   1.136605     2.17   0.034     .1913511    4.734001
         exc |  -.2933697    .595086    -0.49   0.624    -1.482554    .8958147
       _cons |  -62.84932   38.26466    -1.64   0.105    -139.3152    13.61651
------------------------------------------------------------------------------
```

(4) Remedial measures

› Third model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(3, 64)        =     12.72
       Model |   3405.0221         3  1135.00737   Prob > F        =    0.0000
    Residual |  5710.04911        64  89.2195173   R-squared       =    0.3736
-------------+----------------------------------   Adj R-squared   =    0.3442
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4456

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       1.sex |  -5.769422   3.508961    -1.64   0.105    -12.77938    1.240533
         hei |   .5782139   .2098837     2.75   0.008     .1589229    .9975048
         exc |  -.2961036     .61202    -0.48   0.630    -1.518754     .926547
       _cons |  -32.88176   36.69289    -0.90   0.374    -106.1842    40.42072
------------------------------------------------------------------------------
```

(4) Remedial measures

5) Adding more observations

When possible, more observations lead to more variability in X and therefore, may lead to reduction of severity of multicollinearity problem.

6) Variable transformation

Transforming variables can be complicated, we can either perform

› **Ratio transformation** which will lead us to another problem of heteroscedasticity or

› **First difference form** of variable which is popular in time-series data.

Therefore, transforming is not very much recommended.

(1) The nature

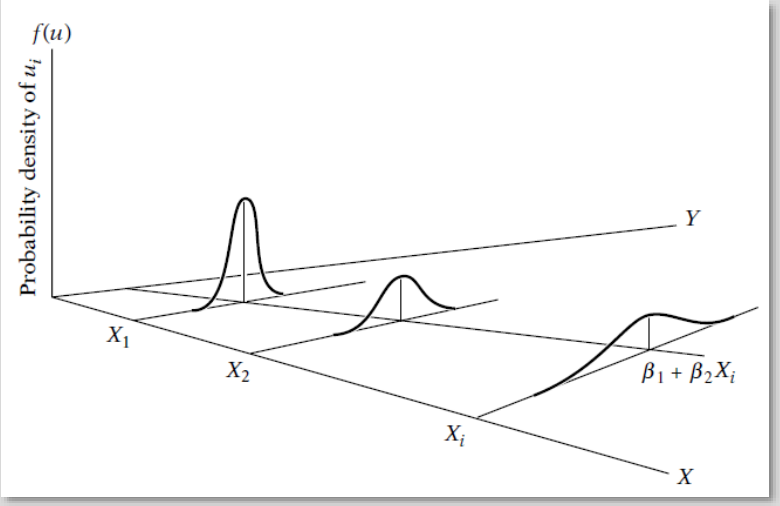
Recall the assumption of homoscedasticity or

› $E(u_i^2|X_i) = \sigma^2$

when this assumption is relaxed, we have

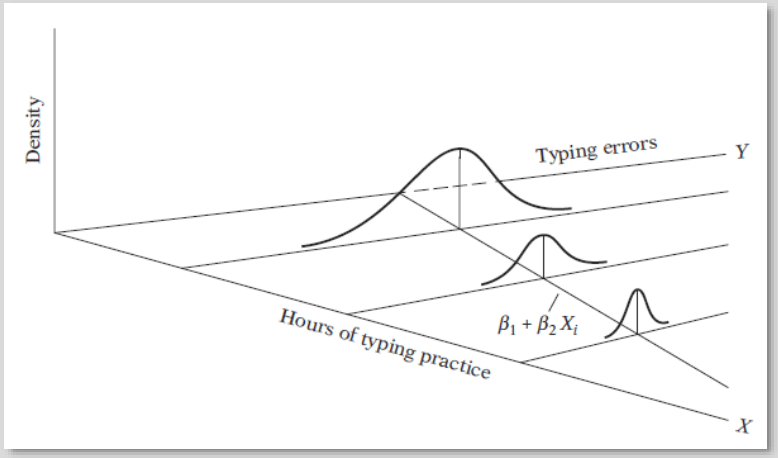
› $E(u_i^2|X_i) = \sigma_i^2$

which means that we allow the error term scattered around each X_i to be different. Two classic examples are income-saving model and error-learning model as follows.



As people get richer, they have more choices over their consumption-saving, leading to a larger variance of saving (Y_i) on larger income (X_i).

(1) The nature



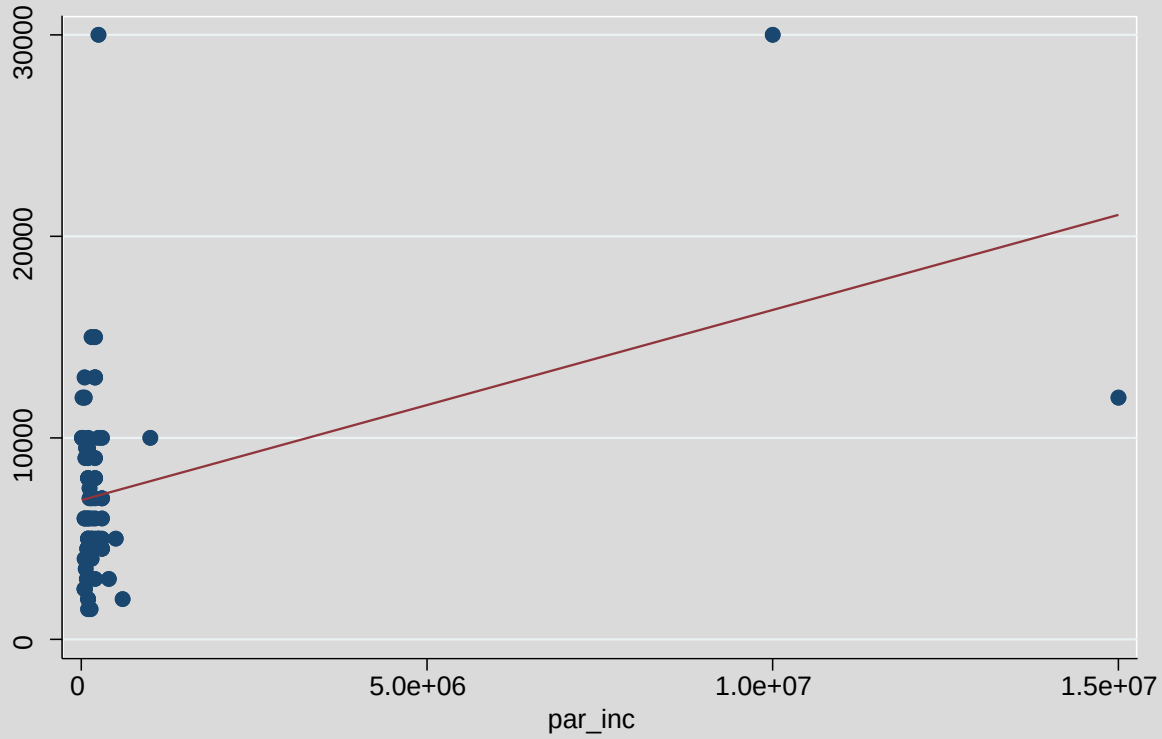
Similarly, people practicing more hours on typing leads to lower typing errors. However, some people maybe pretty good at typing at first and some can be pretty bad due to their familiarity to keyboard layout or language. There are larger difference between people when start practicing but those difference will be minimized as they keep practicing.

Apart from the nature of data, there are some other causes for heteroscedasticity.

(1) The nature

1. Presence of outliers

An outlier is an observation that is much different from the rest, either little or largely different. Inclusion and exclusion of an outlier can alter the result of a regression substantially.



(1) The nature

2. Specification error

For example, a price demand model can be heteroscedastic if we do not include price of complementary or competing commodities. Other commodities' price may be the source of scattered quantity demanded at some level.

3. Skewness of distribution

For example, plotting wealth and education level can show this problem because there a fewer people with high wealth, leading to lower variance in education level, while larger groups of population with lower wealth.

4. Incorrect data transformation and functional form.

Details for this topic will be skipped on this point.

Heteroscedasticity is more likely to be found in cross-sectional data rather than time-series since they deal with members of a population at a given point of time.

(2) Effects on estimation

We begin our examination on simple linear regression, recall that with homoscedasticity assumption yields the variance of the estimator as

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_i^2}$

Relaxing the assumption, the variance becomes

› $\text{var}(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{\left(\sum x_i^2\right)^2}$

$\hat{\beta}_2$ is not be BLUE, with heteroscedasticity. It is still linear and unbiased but it is not efficient anymore, comparing to deriving estimator from another method called **generalized least squares** (GLS).

(Note that an estimator not being efficient meaning that $\text{var}(\hat{\beta}_i)$ is not the lowest.)

(2) Effects on estimation

To estimate with GLS, we start from the basic model of

› $Y_i = \beta_1 + \beta_2 X_i + u_i$

Then we transform these variables by dividing by σ_i, if this standard error is known.

› $\frac{Y_i}{\sigma_i} = \frac{\beta_1}{\sigma_i} + \frac{\beta_2 X_i}{\sigma_i} + \frac{u_i}{\sigma_i}$

Then figure out the variance, we get

› $\text{var}\left(\frac{u_i}{\sigma_i}\right) = E\left(\frac{u_i}{\sigma_i}\right)^2 = \frac{1}{\sigma_i^2} E(u_i^2)$ and if σ_i is known

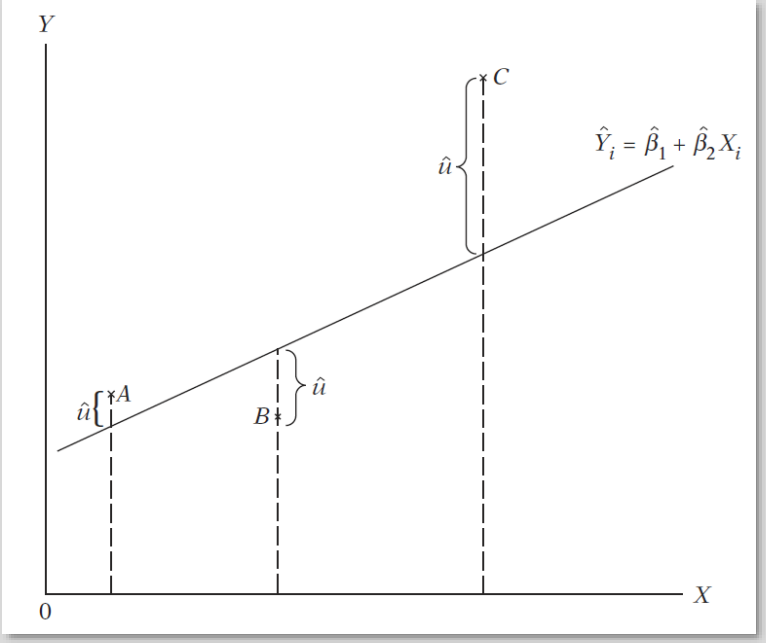
› $\text{var}\left(\frac{u_i}{\sigma_i}\right) = \frac{1}{\sigma_i^2}\left(\sigma_i^2\right) = 1$

which is actually homoscedastic. To retrieve the estimators, we follow least squares method as usual

(2) Effects on estimation

› $\min_{\hat{\beta}_1,\hat{\beta}_2} \sum \left(\frac{u_i}{\sigma_i}\right)^2 = \sum \left(\frac{Y_i}{\sigma_i} - \frac{\hat{\beta}_1}{\sigma_i} - \frac{\hat{\beta}_2 X_i}{\sigma_i}\right)^2$

This method is specifically called **weight least squares (WLS)**, weighing each term especially the error term with the error term itself. Estimators derived from this estimation are called **WLS estimators**. WLS is a class of GLS.



(2) Effects on estimation

WLS estimators derived will have less variance compared to ordinary OLS. Therefore, estimators from OLS is not with the least variance anymore, losing the quality of being efficient.

Consequences of relying on OLS are as follows.

1) OLS estimation allowing heteroscedasticity

› Using OLS while assuming σ_i^2 are known, $\text{var}(\hat{\beta}_2)$ is larger compared to the variance from WLS.

› Larger CI and t value is small, leading to insignificant conclusion.

2) OLS estimation disregarding heteroscedasticity

› Using OLS while assuming homoscedasticity (σ^2) when heteroscedasticity is present, $\text{var}(\hat{\beta}_2)$ will be biased.

› We do not know whether the bias is positive (overestimate: actual variance is lower) or negative (underestimate: actual variance is higher), depending on the relationship between σ_i^2 and X_i.

› Conclusion or inference drawn from hypothesis tests may be misleading.

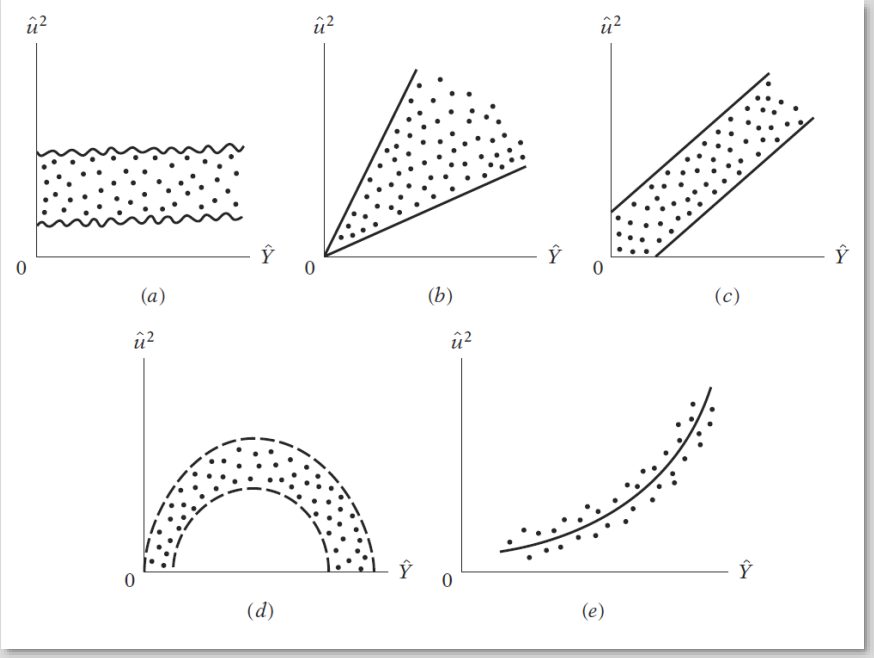
(3) Detecting heteroscedasticity

1. Graphical method

› **Step 1:** Estimate coefficients with OLS.

› **Step 2:** Retrieve $\hat{u}_i^2$. (In STATA, look for predicting residual)

› **Step 3:** Plot $\hat{u}_i^2$ with X_i or $\hat{Y}_i$. There might be multiple X_i in our regression function, so we can rely on $\hat{Y}_i$ as well.



Which of these plots that heteroscedasticity is present?

(3) Detecting heteroscedasticity

2. Park Test

The intuition of Park test assumes that when heteroscedasticity is present, σ_i^2 is a kind of function of X_i. Steps are as follows.

› **Step 1**: Estimate coefficients with OLS and retrieve $\ln \hat{u}_i^2$

› **Step 2**: Estimate $\ln \hat{u}_i^2 = \alpha + \beta \ln X_i + \nu_i$

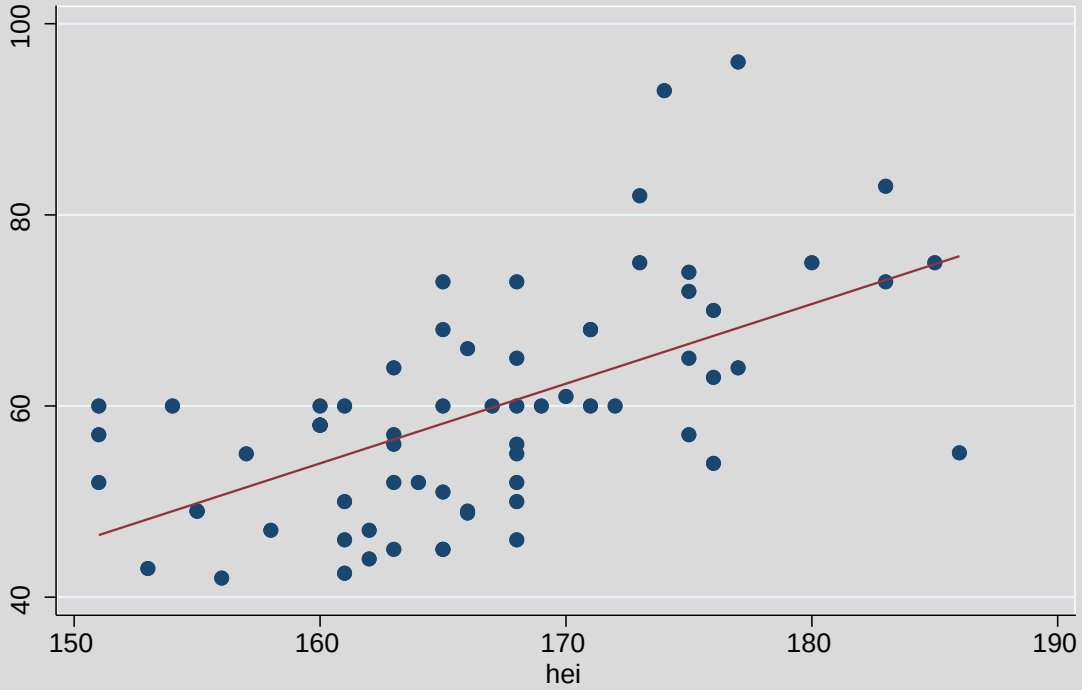
› **Step 3**: Test the significance from zero of β. If it is, it would suggest that heteroscedasticity is present.

Example: Revisit the weight model with only one explanatory variable or height as follows.

› $wei_i = \beta_1 + \beta_2 hei_i + u_i$

We first plot the relationship between weight and height.

(3) Detecting heteroscedasticity



Notice that we do not see significant shift of the variance of $\hat{u}_i$ for each level of X_i.

(3) Detecting heteroscedasticity

› The model

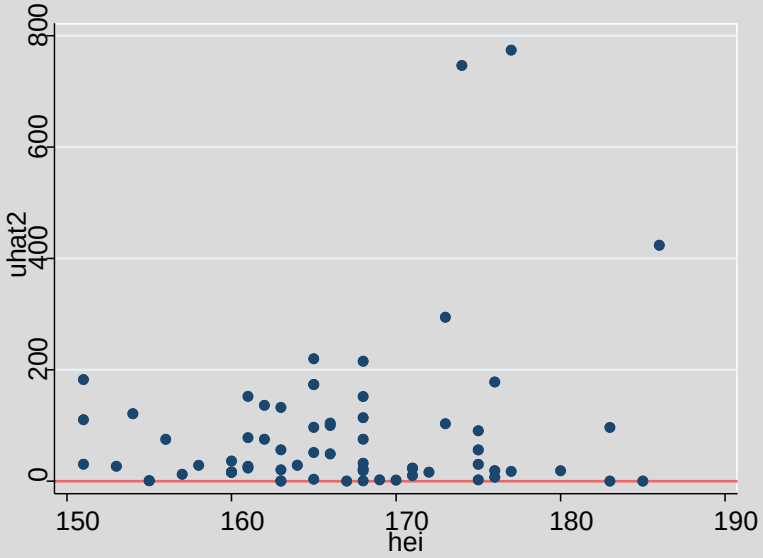
```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =     35.07
       Model |  3162.68606         1  3162.68606   Prob > F        =    0.0000
    Residual |  5952.38514        66  90.1876537   R-squared       =    0.3470
-------------+----------------------------------   Adj R-squared   =    0.3371
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4967

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1408188     5.92   0.000     .5527483    1.115056
       _cons |   -79.4222   23.49114    -3.38   0.001    -126.3238   -32.52063
------------------------------------------------------------------------------
```

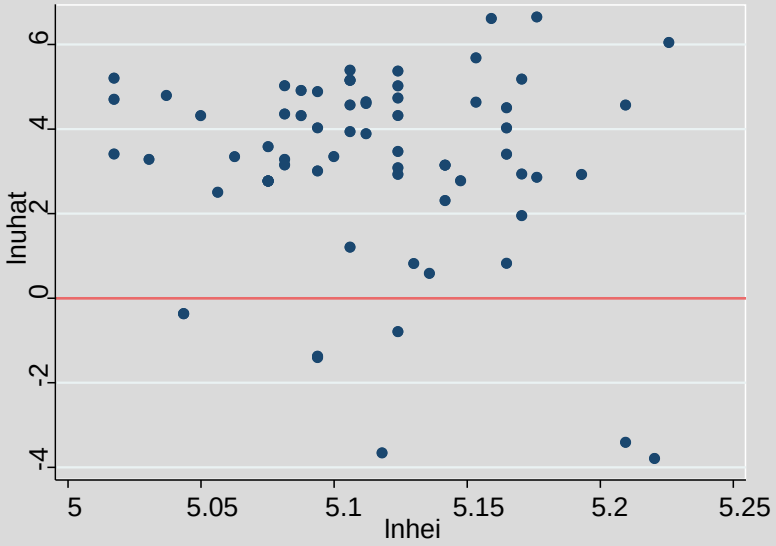
› Park test

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =      0.46
       Model |  2.40426654         1  2.40426654   Prob > F        =    0.5023
    Residual |  348.746061        66  5.28403123   R-squared       =    0.0068
-------------+----------------------------------   Adj R-squared   =   -0.0082
       Total |  351.150328        67  5.24104967   Root MSE        =    2.2987

------------------------------------------------------------------------------
      lnuhat |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       lnhei |  -3.842844   5.696973    -0.67   0.502    -15.21722    7.531529
       _cons |   22.80914   29.13852     0.78   0.437    -35.36778    80.98606
------------------------------------------------------------------------------
```

(3) Detecting heteroscedasticity

We can see that when we plot $\hat{u}_i^2$ with height, and $\ln(\hat{u}_i^2)$ with height, there is no correlation with each other.





(3) Detecting heteroscedasticity

3. Breusch-Pagan (BP) Test

This is a general case for other tests that follow Breusch and Pagan's idea (such as the Breusch-Pagan-Godfrey: BPG test in the textbook). This test is taken from Wooldridge page 270.

› **Step 1**: Estimate coefficients with OLS and retrieve $\hat{u}_i^2$.

› **Step 2**: Regress $\hat{u}_i^2 = \delta_1 + \delta_2 X_{2i} + \cdots + \delta_k X_{ki} + \nu_i$ and retrieve $R^2_{\hat{u}_i^2}$.

› **Step 3**: Calculate F-stat by

$$F_{cal} = \frac{R^2_{\hat{u}_i^2}/(k)}{(1 - R^2_{\hat{u}_i^2})/(n-k-1)}$$

› **Step 4**: Test the null hypothesis of homoscedasticity. If we can reject the null hypothesis ($F_{cal} > F_{cri}$) at the selected significant level, heteroscedasticity is present in our model.

(3) Detecting heteroscedasticity

Here is the result of BP-test. ($\hat{u}_i^2 = \delta_1 + \delta_2 hei_i + \nu_i$)

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =      3.45
       Model |  67079.8689         1  67079.8689   Prob > F        =    0.0676
    Residual |  1281878.37        66  19422.3995   R-squared       =    0.0497
-------------+----------------------------------   Adj R-squared   =    0.0353
       Total |  1348958.24        67   20133.705   Root MSE        =    139.36

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    3.840458   2.066514     1.86   0.068    -.2854704    7.966387
       _cons |   -552.3531   344.7323    -1.60   0.114    -1240.633    135.9271
------------------------------------------------------------------------------
```

Now let's try calculate F,

› $F_{cal} = \frac{R^2_{\hat{u}_i^2}/(k)}{(1-R^2_{\hat{u}_i^2})/(n-k-1)} =$

And find the $F_{cri} =$

(3) Detecting heteroscedasticity

4. White's test

White's test is also very similar to Breusch-Pagan. The differences are the residual model and test statistics.

The steps here applies for 2 explanatory variables, but it is extendable. White's test has an advantage over BP test as it is not sensitive to the assumption of normality.

› **Step 1:** Estimate coefficients with OLS and retrieve $\hat{u}_i^2$.

› **Step 2:** Regress

$$\hat{u}_i^2 = \delta_1 + \delta_2 X_{2i} + \delta_3 X_{3i} + \delta_4 X_{2i}^2 + \delta_5 X_{3i}^2 + +\delta_6 X_{2i} X_{3i} + \nu_i$$

and retrieve $R^2_{\hat{u}_i^2}$.

Additions of higher power and cross product imply that the error variance is functionally related to regressors, their squares, and their cross product.

(3) Detecting heteroscedasticity

› **Step 3**: Calculate the test stat, which in this case, we use **Lagrange Multiplier** (LM) stat

$$LM_{cal} = n \cdot R^2_{\hat{u}_i^2} \sim \chi^2_{k-1.}$$

LM is very useful in many cases due to its basic calculation and **asymptotically** distributed as Chi-square with k d.f.

› **Step 4**: Test the null hypothesis of homoscedasticity. If we can reject the null hypothesis ($LM_{cal} > \chi^2_{k-1.}$) at the selected significant level, heteroscedasticity is present in our model.

(3) Detecting heteroscedasticity

Here is the result of White's test. ($\hat{u}_i^2 = \delta_1 + \delta_2 hei_i + \delta_2 hei_i^2 + \nu_i$)

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(2, 65)        =      1.96
       Model |  76814.6367         2  38407.3183   Prob > F        =    0.1488
    Residual |   1272143.6        65    19571.44   R-squared       =    0.0569
-------------+----------------------------------   Adj R-squared   =    0.0279
       Total |  1348958.24        67   20133.705   Root MSE        =     139.9

------------------------------------------------------------------------------
       uhat2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |   -42.8244   66.19907    -0.65   0.520    -175.0331    89.38427
        hei2 |   .1392366   .1974249     0.71   0.483    -.2550482    .5335214
       _cons |   3348.114   5541.327     0.60   0.548    -7718.679    14414.91
------------------------------------------------------------------------------
```

Now let's try calculate LM stat

› $LM_{cal} = n \cdot R^2_{\hat{u}_i^2} =$

And find the critical value of $\chi^2_{k-1.} =$

(3) Detecting heteroscedasticity

Note that in STATA, the procedures are far way simpler.

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =     35.07
       Model |  3162.68606         1  3162.68606   Prob > F        =    0.0000
    Residual |  5952.38514        66  90.1876537   R-squared       =    0.3470
-------------+----------------------------------   Adj R-squared   =    0.3371
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4967

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1408188     5.92   0.000      .5527483    1.115056
       _cons |   -79.4222   23.49114    -3.38   0.001     -126.3238   -32.52063
------------------------------------------------------------------------------
```

```
. estat hettest
Breusch-Pagan / Cook-Weisberg test
for heteroskedasticity

Ho: Constant variance
Variables: fitted values of wei

         chi2(1)      =     4.38
         Prob > chi2  =   0.0364
```

```
. estat imtest, white
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(2)      =     3.87
         Prob > chi2  =   0.1443

Cameron & Trivedi's decomposition of IM-test
---------------------------------------------------
              Source |       chi2     df      p
---------------------+-----------------------------
  Heteroskedasticity |       3.87      2    0.1443
            Skewness |       3.10      1    0.0782
            Kurtosis |       0.84      1    0.3608
---------------------+-----------------------------
               Total |       7.81      4    0.0988
---------------------------------------------------
```

(4) Remedial measures

1. Weighted Least Squares (WLS)

It can be estimated when σ_i^2 is known.

2. Data transformation: selecting a class of GLS

Advantage of this method is that we do not need asymptotic property. However, a major drawback is we need to speculate relationship between σ_i^2 and X_i.

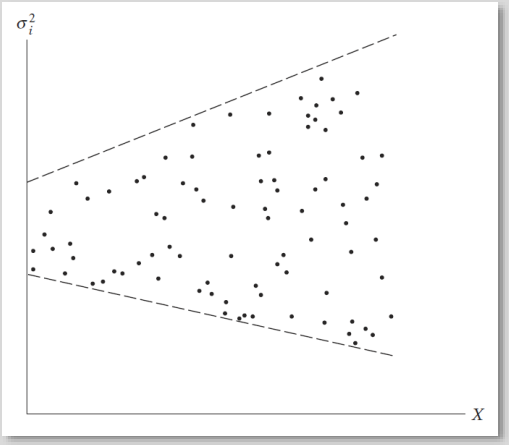
For example, if we assume that σ_i^2 is proportional to X_i so that

› $E(u_i^2) = \sigma^2 X_i$ then $\sigma^2 = \frac{E(u_i^2)}{X_i} = \frac{E(u_i)}{\sqrt{X_i}}$

we can transform our model into

› $\frac{Y_i}{\sqrt{X_i}} = \frac{\beta_1}{\sqrt{X_i}} + \beta_2\sqrt{X_i} + \frac{u_i}{\sqrt{X_i}}$ where X_i must be > 0

$E\left(\frac{u_i}{\sqrt{X_i}}\right) = \sigma^2$ or homoscedastic.



(4) Remedial measures

However, we can see another drawback of this method is that the interpretation of coefficients are now different.

Moreover, there are multiple forms of transformation, due to relationship between σ_i^2 and X_i.

3. White's robust standard errors

This method assumes asymptotic property of our data, which is quite common in national cross-sectional data.

We do not cover how it is derived, even in the book Gujarati does not as well, but we compared the result of normal estimation with using White's robust standard errors.

(4) Remedial measures

› OLS model

```
      Source |       SS           df       MS      Number of obs   =        68
-------------+----------------------------------   F(1, 66)        =     35.07
       Model |  3162.68606         1  3162.68606   Prob > F        =    0.0000
    Residual |  5952.38514        66  90.1876537   R-squared       =    0.3470
-------------+----------------------------------   Adj R-squared   =    0.3371
       Total |  9115.07121        67  136.045839   Root MSE        =    9.4967

------------------------------------------------------------------------------
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1408188     5.92   0.000     .5527483    1.115056
       _cons |   -79.4222   23.49114    -3.38   0.001    -126.3238   -32.52063
------------------------------------------------------------------------------
```

› OLS model with robust standard error

```
Linear regression                               Number of obs     =        68
                                                F(1, 66)          =     27.84
                                                Prob > F          =    0.0000
                                                R-squared         =    0.3470
                                                Root MSE          =    9.4967

------------------------------------------------------------------------------
             |               Robust
         wei |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         hei |    .833902   .1580391     5.28   0.000     .5183667    1.149437
       _cons |   -79.4222   25.98683    -3.06   0.003    -131.3066   -27.53781
------------------------------------------------------------------------------
```

(1) The nature

Recall the assumption of no autocorrelation or serial correlation

› $cov(u_i, u_j | x_i, x_j) = E(u_i u_j) = 0$ where $i \neq j$

Put simply, the term means "correlation between members of series of observations ordered in time (as in time series data) or space (as in cross-sectional data)." If the problem exists,

› $E(u_i u_j) \neq 0$ where $i \neq j$

There can be multiple sources of autocorrelation.

1. Inertia

For example, when an economy is in recovery, there is a 'momentum' built into policies driving economic outcome from previous period.

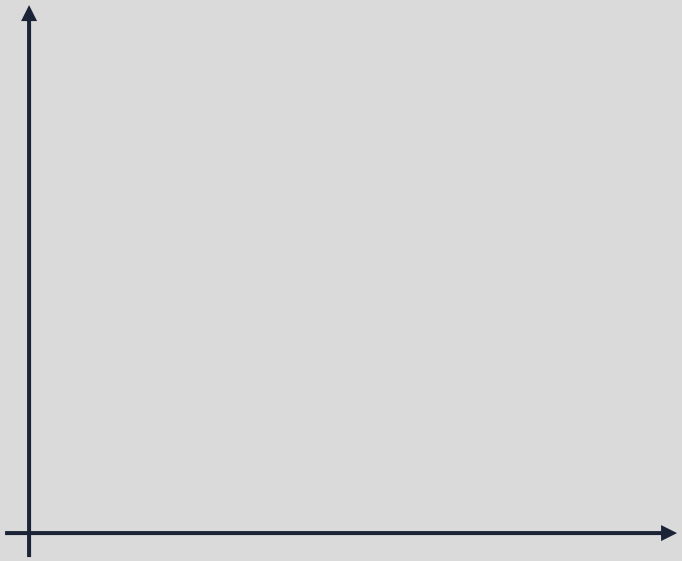
(1) The nature

2. Lags

For instance, a consumption-income model usually includes previous consumption expenditure

› $cons_t = \beta_1 + \beta_2 income_t + \beta_3 cons_{t-1} + u_i$

This is because either psychologically, technologically, or institutionally, people do not change consumption pattern rapidly across periods.



3. Cobweb phenomenon

Most likely to occur with agricultural products, their supply cannot adjust with price instantaneously.

(1) The nature

4. Nonstationarity

Stationarity refers to time-invariant characteristics such as mean, variance, covariance, which is quite rare in time-series data. For example, GDP of an economy is always growing with non-systematic shocks that makes the characteristics time-variant.

5. Specification bias: incorrect functional form

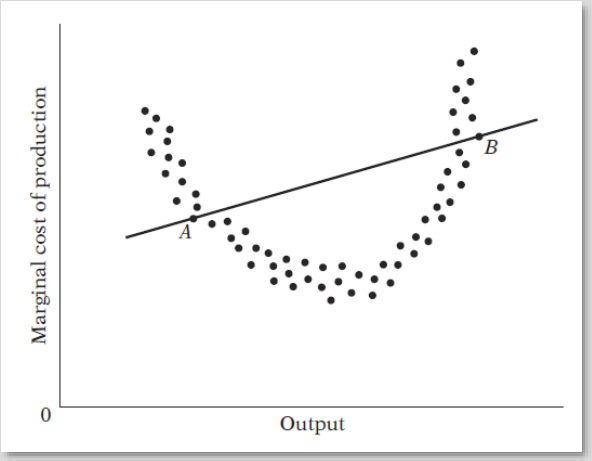
For example, a curved marginal cost which the 'true' model is

› $MC_i = \beta_1 + \beta_2 Q_i + \beta_3 Q_i^2 + u_i$

but if we try to fit our data with a linear model instead

› $MC_i = \beta_1 + \beta_2 Q_i + u_i$

(1) The nature



Between point A and B, linear model underestimates marginal cost while before point A and beyond point B, the model overestimates marginal cost.

If we correlate u_i, u_j we can see clearly that there is a pattern following the curve, hence autocorrelation is present in the linear model.

However, this problem does not surface when we fit the data with polynomial model.

(2) Effects on estimation

From this point on, we will focus on a time-series model, varying Y_t and X_t through time, not across groups of observation in the same period, the model becomes

› $Y_t = \beta_1 + \beta_2 X_t + u_t$

If the error terms are correlated, assumed linearly

› $u_t = \rho u_{t-1} + \varepsilon_t$

This equation is called **first-order autoregressive scheme** or shortly **AR(1)**. If the error term t also correlates with two period back, the model is called **AR(2)**, and so on.

› $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \varepsilon_t$

(2) Effects on estimation

Normally, OLS with no autocorrelation will yield the variance of an estimator as

› $\text{var}(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_t^2}$

If u_t follows AR(1), the variance becomes

› $\text{var}(\hat{\beta}_2)_{AR(1)} = \frac{\sigma^2}{\sum x_t^2}[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum x_t^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum x_t^2} + \ldots + 2\rho^{n-1}\frac{\sum x_t x_n}{\sum x_t^2}]$

We cannot say for sure whether $\text{var}(\hat{\beta}_2)$ is more or less than $\text{var}(\hat{\beta}_2)_{AR(1)}$.

Though $\hat{\beta}_2$ is still linear and unbiased, variance is not minimum, or not being efficient. See this proof of GLS on page 422.

(2) Effects on estimation

If we run the regressing, disregarding autocorrelation, we will find that

› $\hat{\sigma}^2 = \frac{\sum \hat{u}_t^2}{(n-2)}$ is likely to underestimate the true σ^2.

› Overestimation of R^2.

› If σ^2 is not underestimated, $\text{var}(\hat{\beta}_2)$ may still underestimate $\text{var}(\hat{\beta}_2)_{AR(1)}$, and the latter is still inefficient compared to $\text{var}(\hat{\beta}_2)_{GLS}$.

› Both t and F tests are no longer valid.

(3) Detecting autocorrelation

1. Graphical method

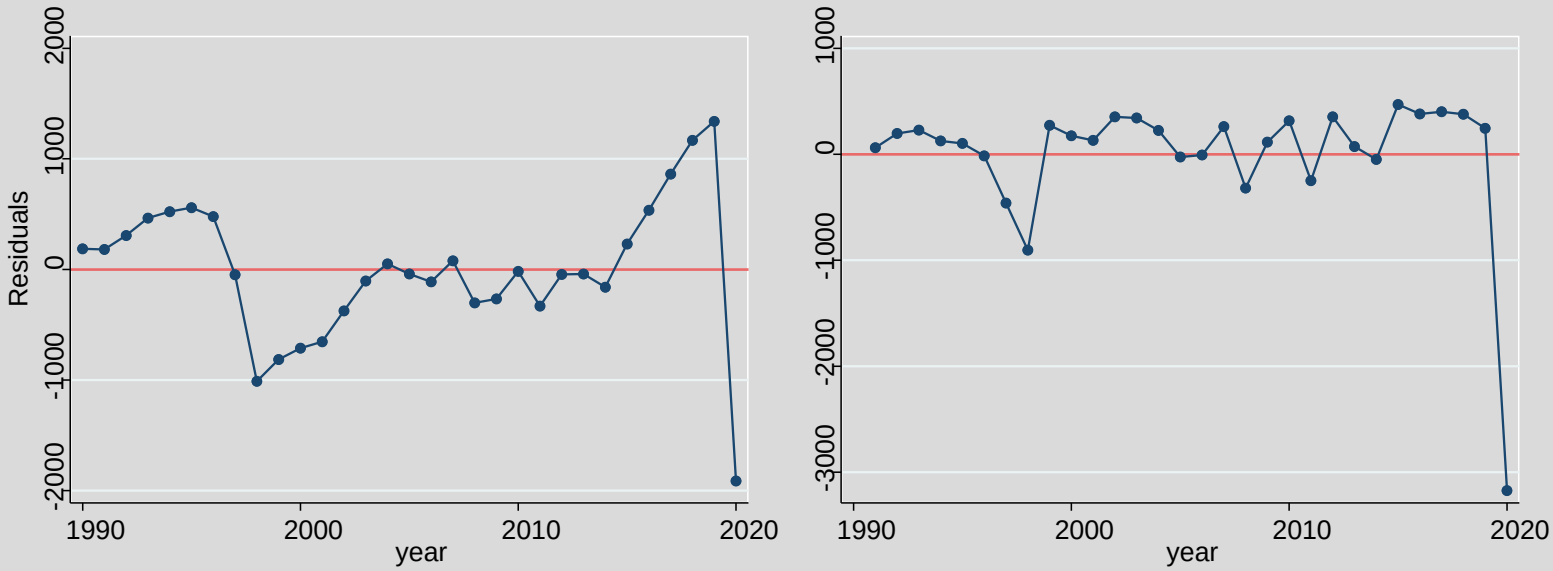
The first one, once again, is not informal but visually telling. If we plot the residuals with time period, we might see interconnection between time period.

Example: Data of GDP (Y_t), in CVM, and headline CPI (X_t) in Thailand from 1990 to 2020 are taken from the Bank of Thailand statistics page. We model it as usual.

› $Y_t = \beta_1 + \beta_2 X_t + u_t$

Now see the result of the estimation on the right-hand side. After that, we can predict $\hat{u}_t$ and plot them over time.

(3) Detecting autocorrelation



On the left-hand side, this is a plot from the model residual. On the other hand, the right-hand side shows a plot from another model that autocorrelation is already resolved.

If autocorrelation is not present, residuals should be randomly distributed around 0 and has no obvious interconnection with the previous period.

(3) Detecting autocorrelation

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------
```

We can also see that R^2, t and F are very high, leading to a very significant coefficient, but this is due autocorrelation.

(3) Detecting autocorrelation

2. Durbin-Watson d Test

Durbin-Watson ***d*** statistics is defined as

› $d = \frac{\sum_{t=2}^{n}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{n}\hat{u}_t^2}$

implies sum of squared differences to the RSS.

For Durbin-Watson test, we assume

› Regression model includes intercept term.

› Regressors X are stochastic or fixed in repeated sampling.

› u_t are generated by AR(1) and normally distributed.

› The model does not include lagged variable(s) of the dependent variable, such as

$$Y_t = \beta_1 + \beta_2 X_t + \gamma Y_{t-1} + u_t$$

› No missing observations in the data.

(3) Detecting autocorrelation

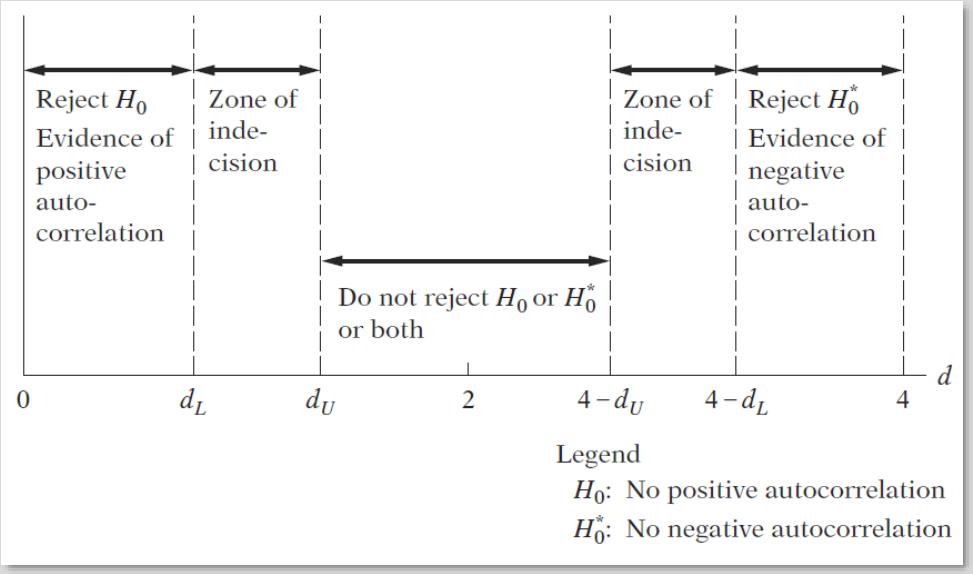
We can also define 'approximate' version of the d stat as

› $d \approx 2(1 - \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2})$ now let's define

› $\hat{\rho} = \frac{\sum \hat{u}_t \hat{u}_{t-1}}{\sum \hat{u}_t^2}$ then

› $d \approx 2(1 - \hat{\rho})$

Since $-1 \leq \hat{\rho} \leq 1$, therefore $0 \leq d \leq 4$



(3) Detecting autocorrelation

Example: using the same dataset, we follow the steps here.

› **Step 1**: State the hypothesis

Since we are testing for general autocorrelation, without any speculation of positive or negative correlation, we are going to set hypothesis for both.

› H_0: No autocorrelation $(d_U < d < 4 - d_U)$

› H_a: Positive autocorrelation $(0 < d < d_L)$ or

negative autocorrelation $(4 - d_L < d < 4)$

Do not forget that we can also reach the inconclusive answer for this test.

› **Step 2:** Run the OLS regression and obtain the predicted residuals.

(3) Detecting autocorrelation

› **Step 3:** Find the critical values. Two 'attributes' that are important for finding d_L and d_U are **number of coefficients** and **number of observation.**

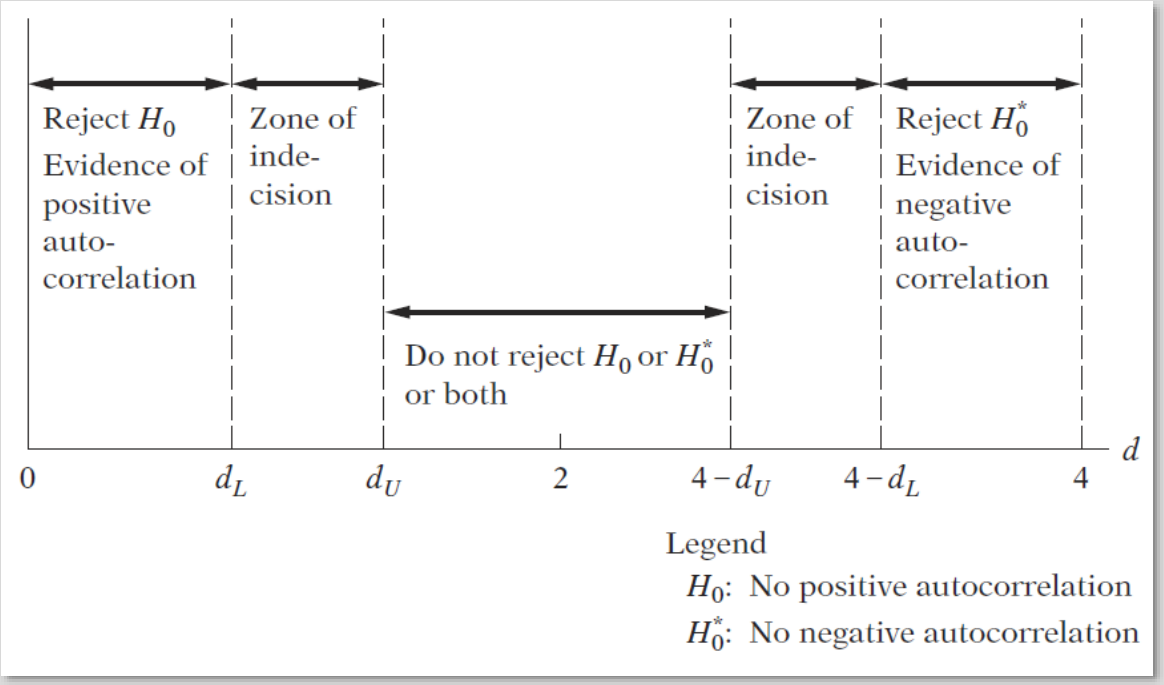
› $d_L =$ and $4 - d_L =$

› $d_U =$ and $4 - d_U =$

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------

. estat dwatson

Durbin-Watson d-statistic(  2,    31) =  1.065525
```

(3) Detecting autocorrelation

› **Step 4:** Conclude the test.



(3) Detecting autocorrelation

3. General test of autocorrelation: The Breusch-Godfrey test (BG)

Assume a model of

› $Y_t = \beta_1 + \beta_2 X_t + u_t$

Now if the error term u_t follows the pth-order autoregressive AR(P) as follows

› $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \cdots + \rho_p u_{t-p} + \varepsilon_t$

The concept is to test these coefficients simultaneously by following the steps.

› **Step 1:** State the hypothesis

› H_0: No autocorrelation ($\rho_1 = \rho_2 = \cdots = \rho_p = 0$)

› H_a: Autocorrelation (ρ are not simultaneously zero)

Step 2: Run the OLS regression and obtain the residuals.

(3) Detecting autocorrelation

› **Step 3:** Estimate this equation, also including regressor(s) into this one.

› $\hat{u}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{u}_{t-1} + \hat{\rho}_2 \hat{u}_{t-2} + \cdots + \hat{\rho}_p \hat{u}_{t-p} + \varepsilon_t$

Then obtain the R^2 from this model

› **Step 4:** Calculate LM-statistics, if the sample is large, BG have shown that

› $\text{LM} = (n-p)R^2 \sim \chi^2_p$

› **Step 5:** Look for the critical value in chi-square table and reject the null hypothesis if the LM exceeds the critical value.

Note that, in STATA, the default lag is 1 but there is an option to include more lags.

(3) Detecting autocorrelation

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------

. estat bgodfrey

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       1     |          4.597               1                   0.0320
---------------------------------------------------------------------------
                     H0: no serial correlation
```

(4) Remedial measures

First of all, make sure that the model is correctly specified (we are going to deal with autocorrelation only). Most of the time we do not know the relationship between u_t and u_{t-1}. In other words, we do not know the value of ρ.

1. First-difference method

The first difference equation takes the form of

› $Y_t - Y_{t-1} = \beta_2(X_t - X_{t-1}) + (u_t - u_{t-1})$ or

› $\Delta Y_t = \beta_2 \Delta X_t + \varepsilon_t$ where $\varepsilon_t = \Delta u_t$

The rule of thumb is that we can use this equation to estimate when $d < R^2$. Note that this model has no intercept term. If included, the intercept is interpreted as **time trend**.

(4) Remedial measures

As we can see from the result, when we use the first-difference model, we cannot reject the null hypothesis of BG test any longer because ε_t is a stationary white-noise error term. However, note that we lose 1 observation due to using difference term.

```
      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =      1.99
       Model |  911983.341         1  911983.341   Prob > F        =    0.1695
    Residual |    12843973        28  458713.323   R-squared       =    0.0663
-------------+----------------------------------   Adj R-squared   =    0.0330
       Total |  13755956.4        29  474343.323   Root MSE        =    677.28
------------------------------------------------------------------------------
      D.cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       hcpi |
        D1. |   114.3877   81.12535     1.41   0.170    -51.79004    280.5654
            |
      _cons |  -76.02649    196.934    -0.39   0.702    -479.4275    327.3745
------------------------------------------------------------------------------

. estat bgodfrey
Breusch-Godfrey LM test for autocorrelation
------------------------------------------------------------------------------
    lags(p)  |          chi2               df                 Prob > chi2
-------------+----------------------------------------------------------------
       1     |          0.015               1                   0.9013
------------------------------------------------------------------------------
                       H0: no serial correlation
```

(4) Remedial measures

Now we can try excluding the constant term.

```
      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 29)        =      3.22
       Model |  1432375.26         1  1432375.26   Prob > F        =    0.0833
    Residual |  12912337.4        29  445253.014   R-squared       =    0.0999
-------------+----------------------------------   Adj R-squared   =    0.0688
       Total |  14344712.7        30  478157.089   Root MSE        =    667.27

------------------------------------------------------------------------------
       D.cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |
         D1. |   90.01267   50.18555     1.79   0.083     -12.6283    192.6536
------------------------------------------------------------------------------
```

(4) Remedial measures

2. Estimating ρ

The reason why we want to know the value of ρ is that we can use this value to transform variables, then use the transformed ones in the GLS estimation.

Assume that the error term follows AR(1) scheme,

› $u_t = \rho u_{t-1} + \varepsilon_t$ where $1 < \rho < 1$

we transform $t - 1$ equation by multiply ρ

› $\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho u_{t-1}$

Then, subtract the t with the equation above to remove the coexistence of the element that are the same between time

› $(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \varepsilon_t$ or

› $Y_t^* = \beta_1^* + \beta_2^* X_t^* + \varepsilon_t$

There are several methods that we can estimate this value of ρ.

(4) Remedial measures

2.1 ρ based on Durbin-Watson d statistics.

From the Durbin-Watson test, we can derive that

› $\hat{\rho} \approx 1 - \frac{d}{2}$

2.2 ρ estimated from the residuals

We can also estimate another model postestimation,

› $\hat{u}_t = \rho \hat{u}_{t-1} + \nu_t$

2.3 Cochrane-Orcutt iterative procedure

Iterative method estimates ρ by starting at some value, mostly 0, then successively approximate multiple times until the value of ρ is stable. Then, ρ can be put into the transformation.

(4) Remedial measures

2.4 Prais-Winsten transformation

Using the same concept of iterative ρ, Prais-Winsten fixed losing 1 observation from the first-difference method because of no antecedent by adding

› $Y_1\sqrt{1-\rho^2}$ and $X_1\sqrt{1-\rho^2}$

Note that Prais-Winsten transformation will retain original number of observation, which might be very important especially for a small sample dataset.

Compare the results from the original model to the third and the fourth method in the next page.

(4) Remedial measures

› **Original model**

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =    312.82
       Model |   131519253         1   131519253   Prob > F        =    0.0000
    Residual |  12192677.5        29  420437.155   R-squared       =    0.9152
-------------+----------------------------------   Adj R-squared   =    0.9122
       Total |   143711930        30  4790397.67   Root MSE        =    648.41

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   6.286548    17.69   0.000     98.33017     124.045
       _cons |   -1827.06   507.7518    -3.60   0.001    -2865.529   -788.5911
------------------------------------------------------------------------------
```

(4) Remedial measures

› Cochrane-Orcutt iterative procedure

```
      Source |       SS           df       MS      Number of obs   =        30
-------------+----------------------------------   F(1, 28)        =     80.97
       Model |  29962164.9         1  29962164.9   Prob > F        =    0.0000
    Residual |  10360869.6        28  370031.057   R-squared       =    0.7431
-------------+----------------------------------   Adj R-squared   =    0.7339
       Total |  40323034.5        29  1390449.47   Root MSE        =     608.3

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   108.1161   12.01497     9.00   0.000     83.50453    132.7276
       _cons |  -1644.838   999.9944    -1.64   0.111    -3693.234     403.558
-------------+----------------------------------------------------------------
         rho |   .4680855
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    1.065525
Durbin-Watson statistic (transformed) 1.298594
```

(4) Remedial measures

› Prais-Winsten transformation

```
      Source |       SS           df       MS      Number of obs   =        31
-------------+----------------------------------   F(1, 29)        =     84.10
       Model |  30077374.3         1  30077374.3   Prob > F        =    0.0000
    Residual |  10371295.2        29  357630.871   R-squared       =    0.7436
-------------+----------------------------------   Adj R-squared   =    0.7348
       Total |  40448669.5        30  1348288.98   Root MSE        =    598.02

------------------------------------------------------------------------------
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   107.0018   10.16464    10.53   0.000     86.21275    127.7908
       _cons |  -1546.283   820.4299    -1.88   0.070     -3224.25     131.685
-------------+----------------------------------------------------------------
         rho |   .4720733
------------------------------------------------------------------------------
Durbin-Watson statistic (original)    1.065525
Durbin-Watson statistic (transformed) 1.297453
```

(4) Remedial measures

3. Newey-West method

This method does not deal with autocorrelation directly, instead, it is very much like White's robust standard error.

The corrected standard errors are known as **HAC standard error.** (heteroscedasticity and autocorrelation).

Newey-West approach is strictly speaking valid in large samples.

```
Regression with Newey-West standard errors      Number of obs     =         31
maximum lag: 1                                  F(  1,        29) =     208.18
                                                Prob > F          =     0.0000

------------------------------------------------------------------------------
             |             Newey-West
         cvm |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        hcpi |   111.1876   7.706168    14.43   0.000     95.42672    126.9485
       _cons |   -1827.06   576.8485    -3.17   0.004    -3006.848   -647.2726
------------------------------------------------------------------------------
```