Printout

Thursday, May 3, 2018 8:19 AM

10. Multicollinearity

10.1 Nature of Multicollinearity

One of the required assumptions of the classical linear regression model (CLRM) in Chapter 7 is that there is no perfect multicollinearity. In this chapter, we take a critical look at this assumption. To clarify, we firstly begin with the multiple regression model as in Equation 10.1, in general, where $X_1 = 1$ for all observations to enable the intercept term to enter the model.

$$Y_i = \beta_1 X_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \cdots + \beta_k X_{ki} + u_i \tag{10.1}$$

If the independent variables above can be algebraically formed as Equation 10.2:

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \cdots + \lambda_k X_{ki} = 0 \tag{10.2}$$

where $\lambda_1, \lambda_2, \ldots, \lambda_k$ are constant such that not all of them are equal to zero simultaneously.

They are said to have exact linear relationship or perfect multicollinearity. That is, we can acquire the value of any independent variable in the model through the linear combination of other independent variables. For instance, if we want to find the value of X_2, we can apply the addition, subtraction, multiplication and division among other independent variables.

On the other hand, if the formation of independent variables follows Equation 10.3 , rather than Equation 10.2, they are said to have imperfect multicollinearity. Specifically, we cannot obtain any independent variable in the model from the linear combination of other independent variables.

random errors

$$\lambda_1 X_{1i} + \lambda_2 X_{2i} + \lambda_3 X_{3i} + \cdots + \lambda_k X_{ki} + v_i = 0 \tag{10.3}$$

where v_1 is the stochastic disturbance term.

10 MULTICOLLINEARITY ✓
11 HETEROSCEDASTICITY
12 AUTOCORRELATION
} NEXT WEEK: 8, 10 MAY

H-A-M Problem

LAST WEEK: EVIEWS + STATA & WRAP UP!
15, 17 May

STEPS
1. Nature of the problem
2. Sources of the problem
3. Consequences of the problem
4. Detection
5. Remedy measures

To see the difference between perfect and less than perfect multicollinearity, we can rewrite Equation 10.2 and Equation 10.3 as:

According to Equation 10.2 and 10.3, consider Table 10.1 which illustrates the collection of 5 observations for each independent variable (X_2 and X_3). It is obvious that we can multiply X_2 by the constant term to transform it into X_3. In this case, we can establish the relationship, as in Equation 10.2, between these two independent variables by letting $\lambda_1 = -3$ and $\lambda_2 = 1$.

$$-3X_{2i} + X_{3i} = 0$$

Table 10.1: Perfect multicollinearity in explanatory variables

Observation	X_{2i}	X_{3i}	$3X_{2i}$	$v_i = -3X_{2i} + X_{3i}$
1	6	18	18	0
2	12	36	36	0
3	7	21	21	0
4	-5	-15	-15	0
5	2	6	6	0

$$\underbrace{-3}_{\lambda_2} X_{2i} + \underbrace{X_{3i}}_{\lambda_3} = 0$$

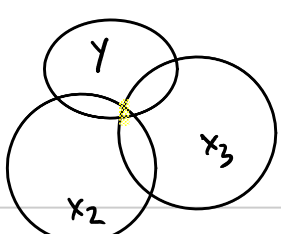
For the case of imperfect multicollinearity, consider Table 10-2. It can be found that we cannot form the relationship as Equation 10.2 due to the difference between independent variables (X_2 and X_3). Even we multiply X_2 by -3, the random disturbance term (v_i) still exists. In Table 10-2, after the forth observation of X_2 is multiplied by 3, it is still different from -12 by 3. Hence, the relationship between these two independent variables can be written as

$$-3X_{2i} + X_{3i} + v_i = 0$$

Table 10-2: Imperfect multicollinearity in explanatory variables

Observation	X_{2i}	X_{3i}	$3X_{2i}$	$v_i = -3X_{2i} + X_{3i}$
1	6	16	18	-2
2	12	45	36	9
3	7	18	21	-3
4	-5	-12	-15	3
5	2	7	6	1

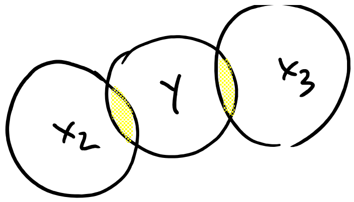
$X_{3i} = 3X_{2i} + v_i$
$16 = 3 \cdot 6 - 2$
$45 = 3 \cdot 12 + 9$
$18 = 3 \cdot 7 - 3$
$-12 = 3 \cdot -5 + 3$
$7 = 3 \cdot 2 + 1$



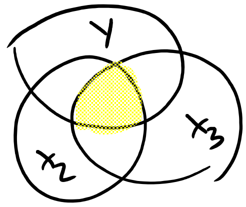
low collinearity between x_2 and x_3!

i.e., $\boxed{3x_{2i} - x_{3i} + v_i = 0}$ where $\lambda_2 = 3$, $\lambda_3 = -1$

No collinearity



No collinearity between x_2 and x_3



high collinearity between x_2 and x_3

Ballentine View of collinearity

$$y = f(x_2, x_3)$$

$$\boxed{y = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i}$$

Ex: $avgscore_i = \beta_0 + \beta_1 \text{expenditure} + \beta_2 \text{avgincome} + u_i$

$i \rightarrow$ provinces in Thailand $(n = 77)$

Read Wooldridge p. 85-87 for the discussion above.

The relationship discussed in this chapter involves only the linear one. Although the independent variable is squared or cubed, as in Equation 10.4, it does not always mean that the model constructed from these variables will suffer the perfect or imperfect multicollinearity. The important factor to consider is whether X_i and X_i^3 can be written in the form of Equation 10.3 or 10.4.

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 X_i^3 + u_i \tag{10.4}$$

In the following sections of this chapter, the consequence of the perfect or imperfect linear relationship of independent variables will be discussed. However, to completely understand the characteristics of multicollinearity, it is essential to know the sources of the problem. In principle, multicollinearity is originated from:

1. *Method of data collection used in the regression model:* sometimes researchers collect the data in the limited amount, causing the sample to concentrate in some group of population rather than to represent the population as a whole.

2. *Restriction imposed in the model:* in the study of relationship between a single dependent variable and multiple independent variables, possibly the linear relationship exists among those independent variables. To illustrate, suppose we study the dependence of the sale of goods Y on the prices of goods X and Y, where X is used to produce Y. With this relationship, when the price of goods X increases, it almost certainly raises the price of goods Y. Hence, there seems to be highly linear relationship between these two variables.

3. *Application of polynomial to the model:* such as Equation 10.4, the X_i^3 is included as another variable. If the data used in the study is restricted within the narrow range, the value of two variable, namely X_i and X_i^3, might not be notably different, resulting in the linear relationship between them.

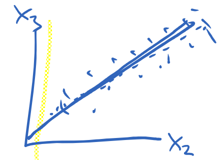
4. *Over-determination of the model:* some models have higher amount of parameters than the amount of observation collected. The evident example of these models is in the medical or human behavior field in which the amount of patients or volunteers is less than the independent variables. Usually, researchers have to discard some variables to make the study possible.

5. *Common trend of independent variables:* the time series data of revenue, expenditure and population seems to move together because, as the time passes, they tend to increase collectively. Thus, the linear relationship among them might occur.

# of observations must be greater than # of explanatory variables. [$n > k$]

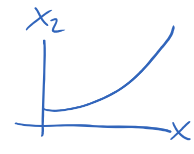
Ex: vote A = $\beta_0 + \beta_1$ expend A + β_2 expend B + β_3 ~~totalexpend~~ + u

percentage of vote for candidate A



Y = consumption expenditure
X_2 = income
X_3 = wealth

	X_3 (wealth) L	H
X_2 (income) L	I	II
H	III	IV



Ex.1 consumption = $\beta_0 + \beta_1$ income + β_2 income^2 + u

x_1 = income
x_2 = income^2 → $\boxed{x_2 = x_1^2}$

→ x_2 is a function of x_1 in non-linear fashion! So, perfect collinearity DOES NOT occur !!!

Ex.2 $\log(\text{consumption}) = \beta_0 + \beta_1 \log(\text{income}) + \beta_2 \log(\text{income}^2) + u$

$x_1 = \log(\text{income})$
$x_2 = \log(\text{income}^2)$

Then $x_2 = 2\log(\text{income}) = 2 \cdot x_1$!!!

$x_2 = 2x_1$ or $2x_1 - x_2 = 0$

where $a_1 = 2$, $a_2 = -1$

PERFECT collinearity emerges.

way out to avoid perfect collinearity → $\log(\text{consumption}) = \beta_0 + \beta_1 \log(\text{income}) + \beta_2 [\log(\text{income})]^2$.

Page 4

Thursday, May 3, 2018 8:19 AM

10.2 Estimation in the Presence of Perfect Multicollinearity

According to Chapter 7, if we have the regression model as shown in Equation 10.1 , we can employ the ordinary least square to estimate β_2 and β_3. By the method of calculus, minimizing the sum of disturbance term squared, we obtain the estimators ($\hat{\beta}_2$ and $\hat{\beta}_3$) as follow:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_{2i} + \hat{\beta}_3 X_{3i} + \hat{u}_i \tag{10.5}$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{3i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \tag{10.6}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum x_{2i}^2) - (\sum y_i x_{2i})(\sum x_{2i} x_{3i})}{(\sum x_{2i}^2)(\sum x_{3i}^2) - (\sum x_{2i} x_{3i})^2} \tag{10.7}$$

where
$y_i = Y_i - \bar{Y}$,
$x_{2i} = X_{2i} - \bar{X}_{2i}$,
$x_{3i} = X_{3i} - \bar{X}_{3i}$

Nonetheless, if the explanatory variables, X_2 and X_3, suffer perfect multicollinearity, namely $X_{3i} = \lambda X_{2i}$ where λ is the constant greater than zero, from the relationship stated, we can substitute $x_{3i} = \lambda x_{2i}$ in equation 10.6 and 10.7 and get

$\hat{\beta}_2 = \frac{0}{0}$ $\hat{\beta}_3 = \frac{0}{0}$

$\hat{\beta}_2$ and $\hat{\beta}_3$ cannot be estimated under perfect collinearity.

We can get the same result for $\hat{\beta}_3$

$$\hat{\beta}_3 = \frac{\lambda^2[(\sum y_i x_{2i})(\sum x_{3i}^2) - (\sum y_i x_{2i})(\sum x_{3i} x_{3i})]}{\lambda^2[(\sum x_{2i}^2)(\sum x_{2i}^2) - (\sum x_{2i})^2]} = \frac{0}{0} \tag{10.8}$$

Page 5

Thursday, May 3, 2018 8:19 AM

On the other hand, consider the case where there is imperfect multicollinearity among regressors. For the model with two regressors, let $X_{3i} = \lambda X_{2i} + v_i$ and substitute $x_{3i} = \lambda x_{2i} + v_i$ into equation 10.6, the estimator of β_2 can be obtained by equation 10.9 which is different from the case with perfect multicollinearity problem. The same is true for both estimators of β_1 and β_3.

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda \sum y_i x_{2i} + \sum y_i v_i)(\lambda \sum x_{2i}^2)}{(\sum x_{2i}^2)(\lambda^2 \sum x_{2i}^2 + \sum v_i^2) - (\lambda \sum x_{2i}^2)^2} \neq \frac{0}{0} \tag{10.9}$$

w/ imperfect collinearity, it is possible to have $\hat{\beta}_2$ and $\hat{\beta}_3$.

Page 6

Thursday, May 3, 2018 8:19 AM

where
$X_{3i} = \lambda X_{2i} + v_i$
$\bar{X}_{3i} = \lambda \bar{X}_{2i}$
$X_{3i} - \bar{X}_{3i} = \lambda(X_{2i} - \bar{X}_{2i}) + v_i$
$x_{3i} = \lambda x_{2i} + v_i$
where $\lambda \neq 0$ and $\sum x_i v_i = 0$

10.3 Practical Consequences of Multicollinearity

In case of near or high multicollinearity, we might encounter the following consequences:

1. Although BLUE, the OLS estimators have large variances and covariances, making precise estimation difficult.

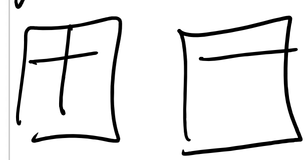
$\hat{\beta}_2 , \hat{\beta}_3$

2. Because of consequence 1, the confidence intervals tend to be much wider, leading to the acceptance of the **zero null hypothesis** (i.e., the true population coefficient is zero) more readily. $\rightarrow \downarrow \hat{t} = \frac{\text{coeff}}{\text{se} \uparrow}$

3. Also because of consequence 1, the t ratio of one or more coefficients tends to be statistically insignificant. low $\hat{t}$

4. Although the t ratio of one or more coefficients is statistically insignificant, R^2, the overall measure of goodness of fit, can be very high. $\Rightarrow$ low $\hat{t}$ but high R^2 } conflicting result occurs

5. The OLS estimators and their standard errors can be sensitive to small changes in the data. see Gujarati



The preceding consequences can be demonstrated as follows.

Large Variance and Covariances of OLS Estimateors

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)} = \frac{\sigma^2}{\sum x_{2i}^2} \cdot VIF \quad \text{where } VIF = \frac{1}{1 - r_{23}^2} \tag{10.10}$$

variance inflating factor

$$Var(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2 (1 - r_{23}^2)} = \frac{\sigma^2}{\sum x_{3i}^2} \cdot VIF \tag{10.11}$$

$$cov(\hat{\beta}_2, \hat{\beta}_3) = \frac{-r_{23}\sigma^2}{(1 - r_{23}^2)\sqrt{\sum x_{2i}^2 x_{3i}^2}} \tag{10.12}$$

where r_{23} is the correlation coefficient between regressors X_2 and X_3 and can be computed by equation 10.13 and the value ranges from -1 to 1.

$$\boxed{r_{23} = \frac{(\sum x_{2i} x_{3i})^2}{\sum x_{2i}^2 \sum x_{3i}^2}} \tag{10.13}$$

The higher the r_{23}, the higher the VIF, the higher $var(\hat{\beta}_2)$:(and $var(\hat{\beta}_3)$:(

According to equation 10.10 and 10.11, it can be seen that the higher the correlation coefficient, the higher the variance of estimators. To ease the analysis, redefine equation 10.10 and 48 by **Variance Inflation Factor (VIF)** by letting,

$$VIF = \frac{1}{1 - r_{23}^2} \tag{10.14}$$

We get

$$Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2} VIF \tag{10.15}$$

$$Var(\hat{\beta}_3) = \frac{\sigma^2}{\sum x_{3i}^2} VIF \tag{10.16}$$

As $r_{23} \to 1$, or the correlation approaches one, VIF $\to \infty$ and the variance will be higher and approaches infinity.
On the contrary, as $r_{23} \to 0$, or the correlation coefficient approaches zero (namely, no linear relationship), VIF $\to 1$ and the variance will be lower. Consider Table 10.3 and Figure 10.1, it can be seen that the higher the correlation , the higher the VIF and the higher the variance of estimators.

$r_{23}\uparrow \to VIF\uparrow \to var(\hat\beta_2)\uparrow \to s.e.(\hat\beta_2)\uparrow \to \hat t = \frac{coeff}{se(\hat\beta_2)}\downarrow$

$var(\hat\beta_3)\uparrow \to s.e(\hat\beta_3)\uparrow \to \hat t = \frac{coeff}{se(\hat\beta_3)}\downarrow$

$\to$ higher chance of accepting Ho when you should actually reject it. (Type II error!)

$\downarrow$

WIDER confidence interval

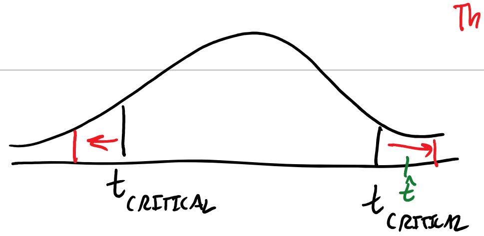
w/o collinearity $\hat\beta_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$

w/ collinearity $\hat\beta_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}\cdot X}$

The higher $r_{23} \to$ The higher X

$\boxed{\begin{array}{l} H_0: \text{INNOCENCE} \\ H_1: \text{GUILTY} \end{array}}$

Release bad guy.



WIDER CI $\to$ Type II error

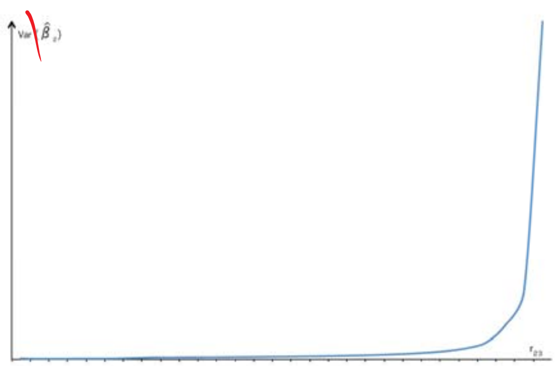
$= \frac{\sigma^2}{\sum x_{2i}^2} \cdot VIF = B \cdot VIF$

Table 10.3: The consequence of an increase in the correlation coefficient on the variance of estimators

$= \frac{\sigma^2}{\sum x_{3i}^2} \cdot VIF = C \cdot VIF$

r_{23}	VIF	$Var(\hat{\beta}_2)$	$Var(\hat{\beta}_3)$
0.00	1	Let [redacted]	[redacted]
0.50	1.33	1.33B	1.33C
0.70	1.96	1.96B	1.96C
0.80	2.78	2.78B	2.78C
0.90	5.76	5.76B	5.76C
0.97	16.92	16.92B	16.92C
0.99	50.25	50.25B	50.25C

Figure 10.1: The consequence of an increase in the correlation coefficient on the variance estimators



When the variance rises due to the level of multicollinearity among independent variables, the standard deviation will certainly rise and at least two negative effects will result. First, the interval estimation will be impaired because the confidence interval will be widen and Table 11.4 (for 95 percent confidence interval and large number of observations). The other negative effect is on hypothesis test since the t-statistic, as in equation 10.17, will be lower and might result in misleading conclusion from hypothesis test.

Page 9

Thursday, May 3, 2018 8:19 AM

Table 10.4: The consequence of an increase in the correlation coefficient on 95 percent confidence interval

r_{23}	95% Confidence interval of β_2
0.00	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}}$
0.50	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}1.33}$
0.70	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}1.96}$
0.80	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}2.78}$
0.90	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}5.76}$
0.97	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}16.92}$
0.99	$\hat{\beta}_2 \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_{2i}^2}50.25}$

$\uparrow r_{23} \rightarrow$ wider CI $\downarrow$ Type II error

$$\hat{t} = \left(\frac{\hat{\beta}_2 - \beta_2}{se(\hat{\beta}_2)\uparrow}\right)\downarrow \tag{10.17}$$

In any models, we might have the high value of coefficient of determination (R^2) and statistically overall significant of the model from F-test. The implication is that the model possesses the explanatory power over the dependent variable. However, it is possible that we might not get the statistically significant result from the test of individual coefficients. That is, some coefficient is not significantly different from zero which means the variable associated with that coefficient lacks explanatory variable because the t-statistic is lower due to multicollinearity problem. This situation is called **conflicting test**, namely the result from t-test contradicts with the one from F-test.

To conclude, if there is perfect multicollinearity among independent variables, we are unable to estimate the parameters in the model. Also, the variance of estimators will approach infinity. Furthermore, if there is imperfect multicollinearity, OLS is still applicable to estimate the parameters. Yet, it has to be aware that the variance of estimators might be so high that some aspects of regression analysis, such as interval estimation and hypothesis test, are negatively influenced.

Page 10

Thursday, May 3, 2018 8:19 AM

10.4 Detection of Multicollinearity

We have already discussed the consequence of both perfect and imperfect multicollinearity among regressors. For the regression analysis, the harmful problem is the situation when there is perfect multicollinearity which will invalidate the estimation of the model in order to explain the true relationship in the population. The case of perfect multicollinearity, thus, can be easily detected.

For the imperfect multicollinearity, if the degree of multicollinearity is not immense, the estimators are still BLUE. Yet, if the degree is huge, the problem will become damaging. Statistically, the extent of multicollinearity can be tested through various approaches. Some of them are discussed here.

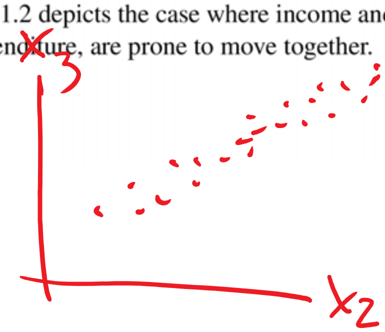
1. *There is conflicting test between t- and F-test*: if we find that the conclusion derived from the two tests are inconsistent, specifically R^2 is high and F-test results in statistical overall significance; whereas, at least, one null hypothesis of some t-tests cannot be rejected, it is reasonable to suspect the multicollinearity problem.

r_{23}

2. *Correlation of regressors is greater than 0.8*: the higher the correlation, the higher the variance of estimators.

3. *Variance inflation factor is greater than 10*: when the regressors face the multicollinearity problem, the value of VIF might be so high that the resulting high variance of estimators adversely affects the regression analysis.

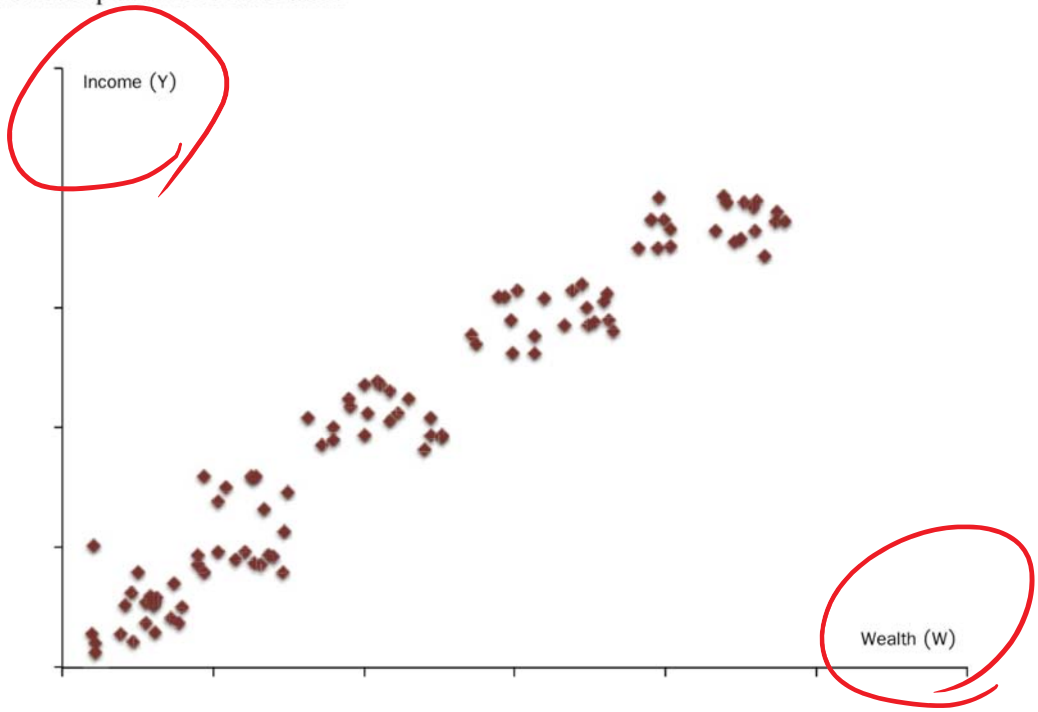
4. *Scatter plot of two regressors is relatively linear*: when we plot the value of on regressor against another and we find that both of them tend to change in the same way, this fact might suggest the existence of multicollinearity. Figure 11.2 depicts the case where income and wealth, which is usually perceived to explain consumption expenditure, are prone to move together.



Page 11

Thursday, May 3, 2018 8:19 AM

Figure 10.2: Scatter plot between two regressors, namely income and wealth, showing the linear relationship between both of them



Page 12

Thursday, May 3, 2018 8:19 AM

10.5 Remedial Measure for Multicollinearity

In principle, the problem of multicollinearity among explanatory variables is not actually serious as we still have BLUE estimators. Notwithstanding, the problem become more severe as the degree of multicollinearity rises and can be alleviated through:

1. *Do nothing*: if the degree of multicollinearity is low, the model is still valid as the BLUE property of estimators is attained.

2. *Apply prior relationship among explanatory variables*: consider the model

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$$

If we know before that the linear relationship between explanatory variables X_2 and X_3 can be written as $\beta_3 = 0.7\beta_2$, we can use this fact to eliminate the problem by

$$\begin{aligned} Y_i &= \beta_1 + \beta_2 X_{2i} + 0.7\beta_2 X_{3i} + u_i \\ &= \beta_1 + \beta_2(X_{2i} + 0.7X_{3i}) + u_i \\ &= \beta_1 + \beta_2 X_i^* + u_i \end{aligned}$$

where $X_i^* = X_{2i} + 0.7X_{3i}$

3. *Discard some explanatory variables*: the removal of the variables could mitigate the problem; but, another problem, namely **specification bias** problem, might occur instead. For example, suppose we want to construct the model where the production is the explained variables; and labor and capital are the explanatory ones. If there is linear relationship between labor and capital, the elimination of one variable might assuage the multicollinearity problem, but might be contrary to economic reasoning. Hence, the decision of which variables will be disposed of should be based on economic theory.

4. *Collect more observation*: this practice will increase $\sum x_i^2$ which is the component of the variances [1]. Accordingly, the variances will be lower despite high correlation among explanatory variables.

$$\boxed{Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum x_{2i}^2 (1 - r_{23}^2)}}$$

5. *Transform the variables*: although there is linear relationship among explanatory variables, it is not necessary that the *first difference* or *ratio transformation* of the variables will have that relationship.

[1] As the data set gets larger, the sample statistic will approach the population parameter. Consequently, we can reasonably state that mean of X is almost stable under the larger data set. In this case, the increase in the size of data set is likely to increase the sum of the square of deviation from the mean

Page 13

Thursday, May 3, 2018 8:19 AM

For the first difference of variables, consider the model in period t and t-1

$$
\begin{aligned}
Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \quad \text{①} \\
Y_{t-1} &= \beta_1 + \beta_2 X_{2,t-1} + \beta_3 X_{3,t-1} + u_{t-1} \quad \text{②} \\
\\
Y_t - Y_{t-1} &= \beta_2(X_{2t} - X_{2,t-1}) + \beta_3(X_{3t} - X_{3,t-1}) + v_t \quad \text{③} = \text{①} - \text{②} \\
\Delta Y_t &= \beta_2 \Delta X_2 + \beta_3 \Delta X_3 + v_t \rightarrow \text{First-difference model.}
\end{aligned}
$$

where

$$
\begin{aligned}
v_t &= u_t - u_{t-1} \\
\Delta Y_t &= Y_t - Y_{t-1} \\
\Delta X_2 &= X_{2t} - X_{2,t-1} \\
\Delta X_3 &= X_{3t} - X_{3,t-1}
\end{aligned}
$$

This transformation perhaps results in no linear relationship among new regressors. Unfortunately, another serious econometric problem might take place which is the **autocorrelation** problem which will be discussed in Chapter 12.

For the ratio transformation of variables, consider the model

$$
\begin{aligned}
Y_t &= \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + u_t \\
\\
\frac{Y_t}{X_{3t}} &= \beta_1 \frac{1}{X_{3t}} + \beta_2 \frac{X_{2t}}{X_{3t}} + \beta_3 \frac{X_{3t}}{X_{3t}} + \frac{u_t}{X_{3t}} \\
\frac{Y_t}{X_{3t}} &= \beta_1 \frac{1}{X_{3t}} + \beta_3 + \beta_2 \frac{X_{2t}}{X_{3t}} + \frac{u_t}{X_{3t}} \\
\boxed{Y_t^* = \beta_1^* + \beta_2 X_{2t}^* + u_t^*}
\end{aligned}
$$

where

$$
\begin{aligned}
Y_t^* &= \frac{Y_t}{X_{3t}} \\
\beta_1^* &= \beta_1 \frac{1}{X_{3t}} + \beta_3 \\
X_{2t}^* &= \frac{X_{2t}}{X_{3t}} \\
u_t^* &= \frac{u_t}{X_{3t}}
\end{aligned}
$$

With this remedial measure, we can reduce the degree of multicollinearity since there is one explanatory variable left in the model. However, when we consider the random disturbance term in this new model, it is possible that the variance of the disturbance term might not be constant, namely **heteroscedasticity**, which will be discussed in the next chapter.