MULTICOLLINEARITY: DOES IT REALLY MATTER?

By: Ibnu Widiyanto

Instructor of Faculty of Economics-Universitas Diponegoro

ABSTRAKSI

Tulisan ini dirancang untuk menjelaskan multikolinearitas (M/C). Hal ini disebabkan keberadaan multikolinearitas bisa merusak model. Keberadaannya menjadi penyebab kegagalan dalam memperkirakan estimasi least squares karena koefisien regresi dan varians menjadi tidak menentu. Secara harafiah, M/C menunjukkan hubungan linear antar variabel independen. Meskipun banyak cara guna mendeteksi keberadaan M/C, tidak ada satu teknik pun yang mampu menunjukkan penyebab M/C. Namun, salah satu cara untuk mengobati M/C adalah dengan menggunakan teknik ridge regression. Teknik ini dirancang untuk menghasilkan varians minimum dalam regresi. Akhirnya, jika keberadaan multikolinearitas tidak mengurangi kebaikan model, keberadaannya M/C bisa diabaikan.

Kata kunci : Multikolinearitas (M/C). Estimasi Least Squares, Koefisien Regresi dan Varians Tidak Menentu.

I. INTRODUCTION

This paper is a fairly tedious and replicative paper, and deliberately so. This paper provides no new theoretical implications, offers no new interpretations, and even draws no new implications of Multicollinearity (M/C) discussion. Instead, this paper is provoked by some inquiries about why multicollinearity has to be tackled in a regression model. The answer is usually as follows: "when developing a classical model, it is unusual that our data estimation conforms just exactly to the theory underlying the model. Many problems often arise when dealing with a model construction, data estimation and even with interpretation. One of the most prevalent snags is the presence of multicollinearity. Hence, it is important to conceive the presence of multicollinearity so an imprecise deduction can be avoided."

This paper henceforth concentrates its energies on providing a careful discussion of the presence of multicollinearity in a system of a regression model. Another is to offer a careful analysis of how to detect and to solve the multicollinearity problem. To this extent, this paper is organized in the following manner. Following this introductory Section One, Section Two deals with an overview of Multicollinearity. Section Three identifies a degree of multicollinearity often appearing in a regression. How to detect multicollinearity is the subject of Section Four. Section Five commences the remedies to rectify the problem. It discusses in detail the ridge regression as a means of solving the multicollinearity problem. Finally, the conclusion of the paper is presented in Section Six.

II. WHAT IS MULTICOLLINEARITY?

One of the assumptions of the classical regression model is that the matrix of independent variables, X, has full rank so $(X'X)^{-1}$ exists. It is therefore there is no exact linear relationship between any of the independent variables in the model. This linear relationship, if happens, actually is adopted by some econometricians to reflect model or data problems. This is called multicollinearity.[1] If there is an exact linear relationship, it implies that such independent variables are exactly collinear. This means the correlation coefficient for these variables is equal to unity ($rx_ix_j =1$)[2], so variances go to infinity.[3] Hence the parameters become indeterminate because it is impossible to obtain numerical values for each parameter separately and the method of least squares breaks down.

Now please consider a simple example, that is, total monthly sales being detected to be influenced by, at least, three independent variables such as price of products in thousand dollars (X_1), weekly sales promotion frequency (X_2) and monthly sales promotion frequency (X_3). Variables X_2 and X_3 are extremely collinear because variable X_3 is exactly four times variable X_2. Each parameter makes perfect sense if only one of the collinear variables appears in the model. However, when these two variables appear simultaneously, then we face, at least, one difficult situation, *viz*., in interpreting the results. The coefficient of the X_2 variable is known as a partial regression coefficient which can be interpreted as the change in Y correlated with a unit change in X_2, *ceteris paribus*. But, this *ceteris paribus* assumption is impossible to apply. We cannot keep all other variables constant because it looks false in such a model.

---

[1] Multicollinearity comes from three different words, multi- *many*, co - *together* and linearity - *the quality of being in lines*. Hence multicollinearity literally means linear relationships among many, as more than one, (independent) variables.

[2] On the contrary, if the independent variables are not intercorrelated at all ($rx_ix_j = 0$), the variables are called to be orthogonal. So, covariance ($\sum X_iX_j$) is zero. Therefore, there is no problem with collinearity.

[3] Consider variance in the case of two independent variables, viz., $Var(\beta_k) = \sigma^2 / \{(1-r_{12}^2) S_{kk}\}$.

It is obvious that once there is a change in X_2, so does X_3. As a result, we cannot compute the least squares parameter estimates.

It is clear that multicollinearity, as in above example, might be easy to discover. It can be seen *apriorily* from the model. But, in practice, we are almost always facing the more difficult situation regarding a high (and low) degree of multicollinearity. Multicollinearity arises when two or more variables (or combinations of variables) are highly (but not perfectly) correlated with each other (Ghosh, 1991, Maddala, 1992, Pindyck and Rubinfeld, 1998, and Cooper and Schindler, 2006). The columns of the regressor matrix X are not orthogonal. This implies that the columns of X are linearly dependent. Further, Schmidt (1976) defines multicollinearity is said to exist if the rank of the regressor matrix X is less than K, the number of regressor. If this happens, the least square estimator, $\beta = (X'X)^{-1} X' Y$, does not exist.[4] When this β estimator does not exist, multicollinearity exists. But, seemingly, multicollinearity is not a condition that either exists or does not exist in economic functions, but rather a phenomenon inherent in most relationships due to the nature of economic magnitudes. However, the consequences of multicollinearity are that the sampling distributions of the β estimators may have large variances. When the variances are large , the estimators will be imprecise. The estimators will be unstable as well as the sample value will be far away from the true value. Therefore, the β estimators may be unreliable to apply.

Similarly, Judge *et al* (1985, p. 899) states:

"Extreme multicollinearity exists when there is at least one linear dependency among the columns of X, and this means that the X matrix is less than full column rank".

This can be happened for several reasons, namely the existence of a physical constraint upon the explanatory variables, the poor implicit design as well as the lack of data observations. This situation may potentially cause the existence of multicollinearity in our econometric model.

[4] The proof is shown by Schmidt (1976) as follows: X'X is of dimension K x K and has the same rank as X. Therefore, if the rank of X is less than K, X'X is singular.

For more detailed discussion, consider the following hypothetical matrices.[5]

	X'X	$(X'X)^{-1}$	$\lvert X'X \rvert$
case # 1	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$	$\lvert 1 \rvert$
case # 2	$\begin{bmatrix} 1 & 0.999 \\ 0.999 & 1 \end{bmatrix}$	$\begin{bmatrix} 500.25 & -499.75 \\ -499.75 & 500.25 \end{bmatrix}$	$\lvert 0.001999 \rvert$
case # 3	$\begin{bmatrix} 1 & 0.9999 \\ 0.9999 & 1 \end{bmatrix}$	$\begin{bmatrix} 5000.25 & -4999.75 \\ -4999.75 & 5000.25 \end{bmatrix}$	$\lvert 0.00019999 \rvert$

From above matrices, we see that case # 1 has orthogonal variables. In this sense, the explanatory variables (Xs) will be linearly independent. Hence the β estimators will be the same as those given by simple regressions of Y on each X. This actually shows that we don't have any problem with multicollinearity. But, in practice, it is uncommon to have orthogonal variables in our collected data though we can set up our experimental designs.

In case # 2 and case # 3, we see that there are increasing correlations between the explanatory variables. It is evident that the covariances[6] and determinants in case # 2 and case # 3 change dramatically. This situation therefore describes a linear relationship (multicollinearity condition) between the explanatory variables (Johnston, 1984, Ghosh, 1991, and Maddala, 1992). When multicollinearity appears, it becomes very difficult to precisely identify and interpret the magnitudes of the explanatory variables. In some cases, our estimated value may differ significantly from true value. Therefore the existence of multicollinearity, no doubt, will reduce our sensitivity in estimating the economic phenomenon.

Finally, we conclude that multicollineriaty is literally linear relationships among independent variables. This happens because the rank of the regressor matrix X is less than the number of regressors, K. So this implies that one of the regressors can have a linear (or perhaps "almost" linear) combination of the others. The presence of multicollinearity, in some cases,

[5] We reproduce the matrices that were firstly proposed by Johnston (1984).

[6] Covariance is shown at the off-diagonal matrix.

disturbs our estimation and interpretation the model (Cooper and Schindler, 2006, Hair *et.al*, 1998, Hair *et. al*, 2006, and Malhotra, 2004).

III. PERFECT AND NEAR MULTICOLLINEARITY

We have discussed the existence of multicollinearity which may harm our model. From above discussion, it is easy to detect extremely collinearity appearing in our model. It is almost like *"Black and White"*. It is apparent. But what about the *"Grey"* collinearity (not extremely collinearity) which is perhaps the most difficult problem to detect. To understand this collinearity, we should be able to distinguish a degree of multicollinearity that might exist in our model.

In general there are two major degrees of multicollinearity, namely, perfect multicollinearity and near multicollinearity.[7] These degrees cause whether or not data problems can be easily detected. In some cases, they may produce whether or not our estimation and interpretation results are specifically defective.

As mentioned earlier, the matrix of independent variables, X, should have full column rank in order to avoid exact linear relationships among independent variables. If not, the variables are extremely correlated and the variances become infinite. When the rank of matrix X is less than the number of regression, K, this implies that one of the regressors is a linear combination. If this is the case, this actually reveals the presence of perfect multicollinearity.

Greene (2005) and Cooper and Schindler (2006) denote that the case of perfect multicollinearity is a serious failure of the assumption of the model, not of the data. For an empirical discussion, we may consider our previous sales-promotion example.

$$Y = \beta_1 + \beta_2 X_1 + \beta_3 X_2 + \beta_4 X_3 + \varepsilon \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

where:

- Y : total monthly sales in thousand dollars
- X_1 : price of product per unit in dollar
- X_2 : weekly sales promotion frequency in number of calls
- X_3 : monthly sales promotion frequency in number of calls
- ε : error disturbances

From the equation (1), it is impossible to catch individual effects of weekly sales promotion frequency and monthly sales promotion frequency. It is obvious that monthly frequency of sales promotion is the aggregation of weekly

[7] Some authors called exact multicollinearity instead of perfect multicollinearity.

frequency *s.t.* $X_3 = 4 \otimes (X_2)$. This is actually a badly specified model, and it reveals that its failure has nothing to do with the quality of data involved. In this analysis, Greene (2005) concludes that no matter how much data are added in the model, we cannot estimate the parameters because of modeling and identification problem. In our case, parameters of our sales-promotion model are said to be unidentified. Therefore, it is worth emphasizing that perfect multicollinearity, undoubtedly, deals with the construction of a regression model.

Schmidt (1976, p. 40-48) even further reveals that if the rank of the regressor matrix X is less than the number of regressors K, perfect multicollinearity exists. This situation means that one of the regressor is a linear combination. If this situation exists, it is impossible to solve the set of least square normal equation s.t. $X'X\,\beta = X'\,Y$. Thereby there does not exist a β estimator which uniquely minimizes Sum of Squared Errors (SSE).[8] When there is no unique solution for the least square estimators, the solutions for X'Y will be infinite (Johnston, 1984, p. 241). It is therefore very time consuming and useless to estimate the least estimators when perfect multicollinearity exists.

The common situation in econometric experiments, particularly with data problems, is one of highly but not perfectly collinearity. It is known as near multicollinearity or "almost" collinearity. This situation happens when the correlation coefficient, r, becomes so high (close to 1, but not equal to 1) in absolute value. In other words, one of the regressors is "almost" a linear combination of the others. Unlike perfect multicollinearity which is easily detected from the model, near multicollinearity is relatively quite difficult to identify.

In the case of near multicollinearity, we can still compute the least square estimators because of the existence of non singular matrix. It is important to know that the matrix X'X is almost singular, but it is still not singular. This can be happened if linear relationships among independent variables are not equal to unity in absolute value ($r < 1$).[9] It is much more common for independent variables to be correlated with $r \rightarrow 1$. This becomes a statistical problem rather than a modeling problem because the difficulty in estimation is not on identification but on precision. It is apparent that the higher the correlation between the regressors, the greater the variances, then the less precise our estimates will be.

Technically, Schmidt (1976) explains the existence of near multicollinearity using the inverse of matrix X'X as follows:

[8] See Schmidt (1976, p. 41) for the proof.

[9] See the formula for variance, in footnote 3.

$$(X'X)^{-1}_{jj} = 1/\{x'_j\ [\,I - X^*_j\ (X^{*'}_j\ X^*_j)^{-1} X^*_j\,]\ x_j\} \quad \text{........................(2)}$$

The right hand side denominator is sum of squared errors when x_j is regressed on X^*_j. If x_j is "almost" a linear combination of columns in X^*_j, the sum of squared errors will be very small. Accordingly, the inverse matrix $(X'X)^{-1}_{jj}$ becomes so large, therefore the least squares estimators have large variances. Finally, it is also clear that the β estimators will not be quite precise to use for estimating the model.[10]

It is clear from equation (2) that not all coefficients will be influenced similarly by collinearity. Suppose $x'_j\ [\,I - X^*_j\ (X^{*'}_j\ X^*_j)^{-1} X^*_j\,]\ x_j = W$, so the sampling variance of the least squares estimates is Var (β_j) = σ^2/ W. It is obvious that W is the sum of squared errors from multiple regression of the ***j***-th independent variables on the other k-1 independent variables. Therefore sum of squared errors decrease with increasing collinearity between ***j***- th independent variables and the remaining independent variables so that the sampling variances tend to rise.

Finally, not only does the presence of multicollinearity produce bad impact either on analyzing the model or on estimating the coefficients of the parameters, but also in some cases, it is not easily identified because there are, at least, two kinds of multicollinearity, *viz*. perfect and near multicollinearity. Perfect multicollinearity is a relationship when a linear combination ($r = 1$) exists. Meanwhile, if one of the regressors is "almost" a linear combination ($r < 1$), this is the case of near multicollinearity.

To some extent, we can distinguish these multicollinearities by observing the model, identifying data and employing some manipulation techniques. We will discuss how to detect multicollinearity in the next section. But, multicollinearity, indeed, is not the source of problem otherwise it reduces standard errors substantially when dropping or more independent variables from the equation. However, no doubt, the presence of multicollinearity may cause our estimation becomes imprecise or even failed. Therefore it is important to realize this multicollinearity problem, since it is still possible to detect and cure the problems (Malhotra, 2004, and Cooper and Schindler, 2006).

[10] Although the estimators are biased, their variances are sufficiently smaller than those of the least square estimates. If this is case, we face the "ridge regression" estimators. We discuss this later.

IV. MULTICOLLINEARITY DETECTION

It is fully understood that the presence of multicollinearity (M/C) may cause a failure in estimating the least squares estimates due to indeterminate coefficient estimates and infinite variances. Thereby, the seriousness of the effects of perfect or near multicollinearity seems to rely on the degree of inter-correlation as well as on the overall correlation coefficient. These effects, consequently, may misguide or mislead our decision in specifying and estimating the model. Thus it is a benevolent step to detect the presence of multicollinearity before postulating the model.

There are several ways to detect the presence of multicollinearity. Koutsoyiannis (1984) proposes some methods to test whether or not multicollinearity exists, that is, by employing Frisch's Confluence Analysis[11] or by applying the experimental work of Farrar-Glauber. In the Frisch's Analysis, the procedure is to regress the dependent variable on each one of the independent variables separately. Hence, we obtain all the elementary regressions and then examine the results based on ***a priori*** and statistical criterion. In this examination, we analyze the effects on the individual coefficients, standard errors and coefficient of determination (R^2) when inserting additional variables. Further, in general cases, Greene (1990) underlines this Frisch's Analysis by observing the changes of the parameter estimates, the significance of the estimated coefficients, and the signs of the coefficients which reflect the magnitudes among independent variables. This observation is useful to distinguish whether or not the highly correlated regressors exist. Thus it is helpful to note that the Firsch's Confluence Analysis and the observation proposed by Greene is quite invariant for multicollinearity tests.

Finally, the Firsch's Analysis concludes the method for multicollinearity tests by classifying a new variable as useful, superfluous, and detrimental.[12] This classification indeed signals whether or not multicollinearity presents.

Another method for multicollinearity tests proposed by Koutsoyiannis is the statistical work of Farrar-Glauber. In this approach, Farrar- Glauber assume that there is only perfect multicollinearity in a function. Consequently,

---

[11] It is also mentioned as Bunch-Map Analysis.

[12] In this method, Frisch notes that a variable is considered to be ***useful*** if it improves R^2 without showing the individual coefficients unacceptable (wrong) on a priori consideration; it is ***superfluous*** if the variable does not improve R^2 and does not influence to any considerable extent on the values of the individual coefficients; finally, it is considered as ***detrimental*** if the new variable affects considerably the sign or the values of the estimated coefficients. Fore more details, see Koutsoyiannis (1984, p. 239)

it is not surprising that Farrar-Glauber employ three statistical considerations related with the function's observation for such multicollinearity tests. These statistics are a *Chi Square* test for detecting the presence of multicollinearity in a function, an *F* test for locating which variables are multicollinear, and a *t* test for finding out the pattern of multicollinearity. It can be concluded that the Farrar-Glauber method for detecting multicollinearity is based only on the correlation coefficients of the independent variables. This method obviously ignores measuring the strength of dependence of Y on the Xs. It is a fact, however, that the effects of multicollinearity depend partly on the overall coefficient of determination ($R^2_{Y.X_1...X_k}$). Therefore, it seems that the Farrar-Glauber method may contain some weaknesses.

Because some methods for multicollinearity tests may possess some weaknesses, for example the Farrar-Glauber method, it is not astonishing that Judge *et al* (1985) advocate different steps to detect the presence of multicollinearity. In their suggestion, Judge *et al* propose to analyze the characteristic roots and vectors of X'X for revealing the presence and the nature of multicollinearity in a sample with poor design. In their proposals, Judge et al feel more confident to employ the spectral decomposition of X'X for multicollinearity tests rather than to employ the Farrar- Glauber method.

In particular, the spectral decomposition of X'X is as follows:

$$X'X = \sum_{i=1}^{k} \lambda_i \, p_i \, p_i' \quad \text{...(3)}$$

This spectral decomposition will be more intellectually accurate to detect multicollinearity because it cannot only isolate which variables are interrelated but it also deals with the issues of what constitutes a "small" characteristic root whether collinearity is harmful (Judge, 1985, p. 902). To some extent, Belsey, Kuh and Welsch (1980) also advocate a similar analysis called the singular value decomposition.[13]

Because computer programs often print out $|X'X|$, we can apply to the computation of the eigenvalues of X'X s.t $|X'X| = \lambda_1 \lambda_2 ... \lambda_k$. In this sense, a small determinant means that some (or many) of the eigenvalues will be small (Johnston, 1984). Accordingly, this implies that $X\mathbf{p}_i \cong \mathbf{0}$. Therefore, we can hopefully identify the "true" linear dependencies between the columns of matrix X. But this eigenvalues computation should not end up until this step. This computation must be followed by another analysis such as variances

[13] See Johnston (1984, p. 249).

calculation since the eigenvalues only give a little help in assessing effects on the individual coefficients.

Additionally, Johnston (1984) and Greene (2005) suggest that the most useful single diagnostic way for multicollinearity tests is the application of coefficient of determination (R_i^2). In the multiple regression, total sum of squared deviations for X_i (TSS_i) determines the minimum sampling variances that might be achieved for $\boldsymbol{b}_i$. The variance of $\boldsymbol{b}_i$ is as follows:

$$\text{Var}[\boldsymbol{b}_i] = \sigma^2 / [(1- R_i^2)\, TSS_i] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots(4)$$

where R_i^2 is the R^2 in the regression of x_1 on the other independent variables in the regression. Any linear relationship in the sample data will increase all sampling variances, but the magnifications for different coefficients will be indicated by a comparison of the R_i^2 s.

It is evident that the role of these R_i^2 s is relatively essential. In particular cases, the simple correlations among the variables may not give an adequate indication of the problem. In this case, in spite of having low correlation of the variables[14], the indication of collinear problems can still be determined by the corresponding R^2 in the regression. If the corresponding R^2 is found to be small[15], this implies that further analysis is required before estimating the model. In a practical sense, this is handled by measuring Tolerance (1- R_i^2), which should be greater than 0.1 and Variance Inflation Factors (VIF) as (1/Tolerance), which should be less than 10 to indicate non-existence of multicollinearity (Hair *et. al*, 1998, Hair *et.al*, 2006, Malhotra, 2004, Sekaran, 2003).

An alternative test for multicollinearity has been developed by Belsley, Kuh and Welsch (1980). In this test, Belsley *et al* advocate the combined application of two diagnostic tools, namely, the condition number of the X matrix and the regression coefficient variance decomposition. These tools obviously represent the two-step procedures recommended by Belsley *et al* for detecting the most collinear affected coefficients.

The condition number of the X matrix is the square root of the ratio of the maximum to the minimum eigenvalues (characteristic roots) s.t.

---

[14] Some authors, for example Greene (1990), denote $r<0.5$ is considered to be low.

[15] In this example, Greene (1990) remarks that we should be concerned about multicollinearity if the overall R^2 in the regression is less than any of the individual R_i^2 we have been considering.

$$\kappa(X) = \sqrt{\frac{\lambda_{max}}{\lambda_{min}}} \quad \text{..(5)}$$

where λ_{max} and λ_{min} denote the maximum and minimum characteristic roots of X'X respectively. The condition number of the X matrix, $\kappa(X)$, is indeed a measure of the sensitivity (elasticity) of the **b** estimators to changes in X'Y or X'X. Further, Greene (2005) remarks that because the eigenvalues rely on the scaling of the data, we should standardize the X matrix by dividing each column of X by $(\boldsymbol{x}'_1 \boldsymbol{x}_i)^{½}$. If the columns of X are orthogonal[16], the condition number will be unity. But, this number, $\kappa(X)$, will be greater than one if there is collinearity between the columns. Hence, we can conclude that the greater the intercorrelation among the variables is, the higher $\kappa(X)$ will be. Finally, Belsley et al (1980) intellectually advise that collinearity problems may appear when $\kappa(X)$ is greater than 20.[17]

The second procedure recommended by Belsley *et al* is the regression coefficient variance decomposition. In this computation, we should inspect the proportions of the sampling variance of each $\boldsymbol{b}_i$ associated with those characteristic roots. The variance of this single coefficient $\boldsymbol{b}_i$ is as follows:

$$\text{Var}[\boldsymbol{b}_i] = \sigma^2 \sum_{i=1}^{k} \frac{p_{ki}^2}{\lambda_i} \quad \text{..(6)}$$

As a result, we may compute the proportions of var $[\boldsymbol{b}_i]$ associated with any single characteristic roots (λ_i). This proportion, in turn, reflects the presence of multicollinearity. Following Judge *et al* (1985), the proportions are

$$\varphi_{ki} = \frac{p_{ki}^2 / \lambda_i}{\sum_{i=1}^{k} p_{ki}^2 / \lambda_i} \quad \text{..(7)}$$

The presence of two or more large values of φ_{ki} indeed indicates that multicollinearity may influence the precision of estimation of the related coefficients. The values of φ_{ki} in excess of 0.50 are likely to be large for this purpose (Belsley *et al*, 1980, p. 100-108) so that we can confirm which variables are seemingly interrelated. Therefore, this analysis is useful to

---

[16] It is important to recall that when the columns are orthogonal, the individual R_i^2 becomes zero.

[17] This is also called as a "danger" level.

indicate which coefficients' variances are adversely affected by the multicollinearity.[18]

The other way to detect is by employing a Theil's measure (m). The formula of a Theil's measure is as follows:

$$m = R^2 - \sum_{i=1}^{k} \left(R^2 - R_{-i}^2 \right)$$

where R^2 is coefficient of determination which can be obtained by taking a regression of dependent variable and its explanatories, while R_{-i}^2 is a squared multiple correlation of taking a regression between dependent variable and its explanatory variables excluding X_i. A Theil's measure closes to zero indicates no multicollinearity (Ghosh, 1991, Maddala, 1992, and Pindyck and Rubinfeld, 1998).

In short, the rules of thumb to detect M/C are that we may have high R^2 as well as significant F statitistic value but few (if not none) significant t ratios also high pairwise correlation among regressors (Malhotra, 2004, and Widiyanto, 2004).

However, all detection techniques here are rather useless as they only indicate the presence of multicollinearity (Gosh, 1991, and Maddala, 1992). This is just a complaining detection technique without offering a solution how to rectify the problem. Even, these techniques cannot locate systematically which explanatory variables are the cause for multicollinearity. What should we do then becomes the next issue.

V. SOME SOLUTIONS TO MULTICOLLINEARITY PROBLEM

Several methods have been proposed for rectifying this multicollinearity problem. Of course, one of the proposed remedies is to correct the model. As in equation (1), we can easily drop variables suspected of causing the problem from the regression because a decision on adding more data which are simply the same[19] is no help in multicollinearity remedy. But, in doing so, we should be careful because we may come upon the problems of specification such as the

---

[18] Although this analysis is quite practical, some authors find some flaws in this method, for example, small eigenvalue doesnot always mean that our least squares estimator may have the same imprecision for each of the parameters. It is possible that the ***k*** th coefficient is not affected by a small eigenvalue as long as the elements of the ***i*** th row of the ***k x k*** matrix that diagonalized X'X, p_{ki}, is small. Hence, the collinearity relationship may not be happened in the ***k*** th variable. For more details, see, for example, Judge et al (1985, p. 903-904).

[19] Also, it is very common that there are difficulties to get better data. But, the joint use of time-series and cross-section data sometimes will be helpful. For an example, people do this combination data in the analysis of demand functions. See Johnston (1984, p. 250-251).

biased estimators.[20] Hence, we should make some considerations about our dilemma.

Due to the flaws in omission of (relevant) variables, a number of other approaches have been suggested. Judge *et al* (1985) classify the purely mechanical approaches to solve multicollinearity into two types, namely, traditional and non traditional solutions. In the traditional solutions, at least two approaches are offered by econometricians to handle this multicollinearity problem such as the ridge estimation and principal components.[21] In addition, the works of Stein and its extensions are good examples for non traditional solutions to the multicollinearity problem. In this paper, however, the primary discussion is on traditional solutions because not only does this paper become very long but also only few econometricians deal with non traditional solutions to multicollinearity problem.[22]

Ridge regression was originally employed to investigate the sensitivity of least square estimates based on particular data presumably indicating near-extreme collinearity.[23] But, this ridge regression recently is conceived to be a "rectified estimator" probably with sufficiently smaller variances than those of the least squares estimates.

In discussions of ridge regression estimator, we consider the linear regression model $Y = X\beta + \varepsilon$ as well as assume that X is standardized s.t X'X is the matrix of simple correlations among the independent variables. The generalized ridge regression estimator of β is as follows:[24]

$$\tilde{b} = (X'X + k\,I)^{-1} X'Y \qquad \ldots\ldots(8)$$

where $k > 0$ is the scalar chosen arbitrarily.[25] One property of the ridge regression is that although it is a biased estimator, its variance is less than the least squares estimator's variance. In this analysis, we can view easily that expected ridge estimator ($E[\tilde{b}]$), after substituting Y into $X\beta + \varepsilon$ and assuming $E[\varepsilon] = 0$, is as follows:

$$E[\tilde{b}] = (X'X + k\,I)^{-1} X'X\beta \qquad \ldots\ldots (9)$$

Thereby, its variance becomes

$$Var[\tilde{b}] = \sigma^2 (X'X + k\,I)^{-1} (X'X) (X'X + k\,I)^{-1} \qquad \ldots\ldots (10)$$

[20] For proof, see Greene (1990, p. 259-261).
[21] See also (Greene (1990), Johnston (1984) and Schmidt (1976).
[22] To some extent, if readers want to apply this non traditional solutions, extensive discussions in Judge et al (1985, p. 922-927) will be good example.
[23] The term of "Ridge Regression" was firstly proposed by Hoerl and Kennard. See, for example,
Hoerl and Kennard (1970) and Judge et al (1985)
[24] See, for example, Schmidt (1976), Johnston (1984), Judge et al (1985) and Greene (1990).
[25] If k = 0, the estimator becomes the Least Squares estimator.

This simply means that the ridge regression estimator's variance is smaller than that of the ordinary least squares estimator.[26] Indeed, this raises the possibility that its mean square error may be less than the OLS estimator's mean square error (Schmidt, 1976). It is, however, difficult to determine the value of k. To some extent, we can try various values of k to stabilize the $\tilde{b}$ vector. For example, we may choose, say k = 0.01, and then try successively larger values until the coefficients stabilize.[27] Further, Judge et al (1985) give an excellent discussion for selecting values of k for use in $\tilde{b}$.[28] In this k values selection, Schmidt (1976) comments that the larger k, the larger is the bias of $\tilde{b}$, but the smaller is its variance, accordingly, the ridge regression mean error square error becomes smaller. Therefore, it produces an estimator superior to least square.[29] Unfortunately, Greene (2005) and Judge *et al* (1985) note that the mean square error (MSE) is problematic to be a beneficial measure because it is a function of the unknown parameters that we are undertaking to estimate. Consequently, as a practical matter, we may not take an inference about β without any MSE improvement.

Johnston (1984), to some extent, advocates an MSE improvement by dropping one or more explanatory variables. In sodoing, we construct a three-variable model, s.t.

$$y = \beta_2 \boldsymbol{x}_2 + \beta_3 \boldsymbol{x}_3 + \mu \quad \text{.....(11)}$$

where the variables are expressed in deviation form.[30] From equation (11), we can verify that variance of b_{12} is smaller than it of $b_{12.3}$. [31] Thereby, we can make a trade off between biasedness and variance. Because the mean square error is the sum of sampling variance and square bias, we can compare MSE

---

[26] As $\mathrm{Var}[\beta] = \sigma^2 (X'X)^{-1}$ and k > 0, $\{\mathrm{Var}[\beta] - \mathrm{Var}[\tilde{b}]\}$ is certainly positive semidefinite.

[27] In determining the values of k, Judge et al finally conclude the works of Hoerl and Kennard (1970) and Theobald (1974), namely, $k < \sigma^2/\theta^2_{max}$ (where θ^2_{max} is the largest element of the vector $\boldsymbol{\theta} = \mathbf{P}'\boldsymbol{\beta}$ and P is the matrix whose columns are orthonormal characteristic vectors of X'X s.t PP' = I) or $k < 2\sigma^2/\beta'\beta$ respectively.

[28] Judge et al discuss extensively proposals for determining values of k, namely, adaptive ordinary ridge estimators (by employing, for example the ridge trace) and adaptive generalized ridge estimator (by applying, for example an operational version of the optimal choice). They denote the ordinary ridge estimator to be $b^*(k)$ and the generalized ridge estimator firstly introduced by Hoerl and Kennard to be $b^*(D)$ where D is a diagonal matrix of constants s.t $d_i \geq 0$ where for any arbitrary values, $d_i = k$.

[29] Judge et al call this estimator as an adaptive ridge estimator.

[30] Deviation form is $\mu - \bar{\mu}$. But, this disturbance deviation has no effect on derivations so we may ignore it.

[31] Recall that b_{12} and $b_{12.3}$ are OLS estimate of β_2. But, the first is a slope of simple regression y on $\boldsymbol{x}_2$; while the latter is a slope of multiple regression y on $\boldsymbol{x}_2$ in the presence of $\boldsymbol{x}_3$. Both are not necessarily the same.

(b_{12}) and MSE ($b_{12.3}$) to come across which one has smaller MSE. Once we know the smallest MSE, we can drop such variable(s) causing unstable coefficients and accomplish a regression to get the minimum variance. In our example, Johnston (1984) compares both MSEs so that

$$\frac{\text{MSE}(b_{12})}{\text{MSE}(b_{12.3})} = 1 + r_{23}^2 \left(\psi^2 - 1\right) \quad \text{...} (12)$$

where the ψ^2 statistic is the ratio of the true (but unknown) β_3 to the true variance (not the estimated) variance of $b_{13.2}$. In this computation, the ψ^2 statistic is as follows:

$$\psi^2 = \frac{\beta_3^2}{\sigma^2 / \sum x_3^2 \left(1 - r_{23}^2\right)} = \frac{\beta_3^2}{\text{Var}(b_{13.2})} \quad \text{……………………...}(13)$$

Consequently, from equation (12), it is obvious that MSE(b_{12}) will be less than MSE($b_{12.3}$) if ψ^2 is less than unity.

Moreover, Johnston (1984) adds that if we wanted to get an estimate of β_2 as precise as possible and had a strong belief that ψ^2 was less than one, it might be reasonable to drop $\boldsymbol{x}_3$ from the regression and might easily accomplish a regression of $\boldsymbol{y}$ on $\boldsymbol{x}_2$. But, of course, it is very difficult to determine ψ^2. It is unknown. Hence, as his exhaustive study of the three- variable model, Johnston concludes that the best procedure is to carry out the direct OLS regression of $\boldsymbol{y}$ on $\boldsymbol{x}_2$ and $\boldsymbol{x}_3$ unless we believed that $\psi^2 < 1$, we might omit the variable.

This approach will be quite useful in improving MSE for only three-variable case. It is, however, not only unsuitable when dealing with the ***m***-variable case but also dubious while omitting variable(s).[32] It is not surprising if some authors suggest to impose a set of linear restrictions on the coefficients for improving MSE. In doing so, Johnston writes the usual linear model s.t. $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\mu}$ with the set of q ($\leq$ m) restrictions[33] organized in $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$. The estimator following these restrictions is as follows:

$$b_{JOH} = b + (X'X)^{-1} R' [R (X'X)^{-1} R']^{-1} (r - Rb) \quad \text{……………………..} (14)$$

where b is the unrestricted OLS estimator. From this model, we can conclude that b_{JOH} has better MSE than b if any quadratic form in MSE(b_{JOH}) is less than or equal to the quadratic form in MSE(b) s.t.

$$k' \text{MSE}(b_{JOH}) k \leq k' \text{MSE}(b) k \quad \text{……………………………...} (15)$$

for any nonnull m-element vector k; and the sum of the MSEs of the restricted estimators is less than or equal to it of the MSEs of the unrestricted estimators s.t.

[32] See Greene (1990) for an example flaw in omitting variable(s).

[33] Note that m is the number of variables.

tr MSE(b_{JOH}) ≤ tr MSE(b) ....(16)

Both conditions, equation (15) and (16), being the strong and weaker criterions respectively are already investigated by many authors, as in Johnston. To some extent, Johnston (1984) recites that the strong criterion, equation (15), will be satisfied if $F = \frac{\hat{\lambda}}{q}$ where

$$\lambda = \frac{(\mathbf{r}-\mathbf{R}\beta)' \left[R \quad (X'X)^{-1} \quad R'\right]^{-1} (\mathbf{r}-\mathbf{R}\beta)}{2\sigma^2} \leq ? \quad \text{........(17)}$$

Hence, it is simply to deduce that the F statistic will be a useful value to distinguish which estimators are better in MSE.

Explicitly, Johnston (1984) summarizes the practical procedure for employing this F statistic as follows:[34]

1. Compute the usual F statistic, based on the difference in the sum of squared errors from the restricted and unrestricted regressions.
2. If $F > F(q, n\text{-}m)_{0.95}$, say, in the Wallace and Toro-Vizcarrondo table, reject the hypothesis that the restricted estimators are better in MSE. If $F < F(q,n\text{-}m)_{0.95}$, use the restricted estimators. (Please see also Cooper and Schindler, 2006)

Thus the test for the improvement in MSE is to compare the sample F statistic with a critical value tabulated by Wallace and Toro-Vizcarrondo (1969).

Fortunately, some authors find that the restricted estimators own smaller variances than the unrestricted OLS estimators. But, it is worth noting that the restricted estimator may probably be biased if the restrictions are not correct. Despite the fact that we might have a biased estimator, we finally come across the trade off between bias and variance as being our previous dilemma. Therefore, hopefully, we are able to find out the smallest MSE in our model to get the ridge estimators for dealing with multicollinearity.

An alternative approach coping with multicollinearity is principal components. The primary usefulness of principal components analysis lies in its function as an exploratory tool. It can mitigate the effects of multicollinearities in the data and advocate useful data transformations (Judge *et al*, 1985, p. 912). Unlike the ridge estimation which is purely mechanical, the appeal of principal components is more intuitive. [To some extent, a discussion

[34] This is for the strong MSE criterion. For the weaker strong criterion, equation (16) will be satisfied if $\lambda \leq q/2$. The F critical values utilize the table tabulated by Goodnight and Wallace. See Johnston (1984, p.258) and Goodnight and Wallace (1972).

of principal components in this paper is primarily summarized from extensive studies of Judge *et al* (1985) and Greene (2005).]

Judge *et al* (1985) compose that principal components regression is basically a method of inspecting the sample data or design matrix for directions of variability and using this information to reduce the dimensionality of the estimation problem. This reduction can be attained by imposing exact linear constraints which are sample specific, but have certain maximum variance properties making their use attractive.

Please consider the transformation model from our usual linear model as follows:

$$y = XPP'\beta + \mu = XP\varphi + \mu = Z\varphi + \mu \quad \text{.................................} \quad (18)$$

where P is a (K x K) matrix whose columns ($\mathbf{p_i}$) are orthogonal characteristic vectors of X'X associated with λ_1 (the largest characteristic root) and Z is the (T x K) matrix of principal components s.t. $\mathbf{z_i} = X\,\mathbf{p_i}$. The principal components estimator of β is obtained by deleting one or more of the variables $\mathbf{z_i}$. As we apply OLS and make a transformation to the original parameter space, we can partition Z into two parts Z_1 (to be retained) and Z_2 (to be deleted) imposing an identical partitioning on P. Accordingly, after some manipulations, the principal components estimator can be obtained by an inverse linear transformation. Since we delete the components in Z_2 (so $P_2\varphi_2 = 0$), the principal components estimator of β is $\beta^* = P_1\varphi_1 = P\varphi^*$. This estimator is actually known to have smaller variances than the least squares estimator, b, but it is biased unless the restrictions $P_2'\boldsymbol{\beta} = 0$ are true. Hence, it seems that the principal components estimator is likely the same as the ridge estimator although either has different approach.

The next procedure of principal components regression deals with how to select deleted components and to understand the consequences of such deletion. In doing this procedure, Judge *et al* (1985) and Greene (2005) suggest to compute and find out "small" characteristic roots from Z_2 (this is to preserve the variability in the sample data when reducing the dimensionality of the estimation problem) and to employ test of hypotheses of the sample specific restrictions $P_2'\boldsymbol{\beta} = 0$ using classical or MSE tests (this is to test whether or not linear dependencies "almost" hold for population variables).

Finally, Greene (2005) concludes that the use of principal components is an attempt to extract from the X matrix a small number of variables that account for most or all of the variation in X. But he also notes that there are problems with using this estimator such as producing the failure of all ad hoc data search procedure as principal components are not chosen on the basis of any relationship of the regressors to y and ambiguous interpretation of the results.

VI. CONCLUSION

Multicollinearity is an econometric condition when the measured variables are highly intercorrelated. This condition makes us difficult to analyze the individual effects of such measured variables. As discussed earlier, multicollinearity becomes a latent problem in estimating and deducing the results. It causes either identification or precision problem. Thus it looks that the discussion of multicollinearity is not very comforting.

There are two degrees of multicollinearity, namely perfect and near multicollinearity. Perfect multicollinearity seems to be modeling dilemma while near multicollinearity presumably deals with data problems. Indeed, it is not easy to identify multicollinearity. Fortunately, some econometricians find some approaches in detecting the presence of these multicollinearities. For an example, Belsley *et al* develops an excellent technique to measure this multicollinearity by computing the condition number of matrix and the regression coefficient variance decomposition, while Theil proposes a Theil's measure. However, all proposed techniques seem to be of no use as they cannot further show how to solve multicollinearity.

Nevertheless, after detecting the presence of multicollinearity, we can solve this problem by employing some solution procedures which are not related with the suggested detection technique. Though, there are at least two solutions to rectify the problem such as traditional and non traditional solutions, this paper only focuses on discussing traditional solutions. In the traditional solutions, we conclude that both the ridge regression and the principal components regression are a sound technique to solve the multicollinearity problem. Although both techniques are adequate, the ridge regression is mechanically better than the principal components regression. As the ridge regression deals with minimum variance in the regression, econometricians suggest that if the presence of multicollinearity does not reduce model performance (goodness of fit), they might ignore the presence of multicollinearity in a proposed regression model.

REFERENCES

Belsley, D., E. Kuh and R.E. Welsh (1980), *Regression Diagnostic, Identifying Influential Data and Sources of Collinearity*, (New York: Wiley & Sons)

Cooper, Donald R., and Pamela S. Schindler (2006), *Business Research Method*,

(New York: McGraw-Hill Co. Inc.)

Goldberger, Arthur S. (1991), *A Course in Econometrics*, (Cambridge: the President and Fellows of Harvard College)

Ghosh, Sukesh K., 1991, *Econometrics : Theory and Applications*, (Englewood Cliffs:

Prentice Hall, Inc)

Greene, William H. (2005), *Econometric Analysis*, Fifth edition, (New York: Macmillan Publishing Company)

Johnston, J. (1984), *Econometric Method*, Third edition, (New York: McGraw-Hill Book,Inc.)

Judge, George G., W.E Griffiths, R. Carter Hill, Helmut Lütkepohl and Tsoung- Chao Lee (1985), *The Theory and Practice of Econometrics*, Second edition, (New York: John Wiley & Sons)

Hair, Joseph F., Rolph E. Anderson, Ronald L. Tatham, and William C. Black, 1998, *Multivariate Data Analysis*, (Englewood Cliffs: Prentice Hall, Inc.).

-------, Robert P. Bush, and David J. Ortinau, 2006, *Marketing Research* , (New York: Mc Graw –Hill Book, Inc.)

Hoerl, A.E. and R.W. Kennard (1970), "Ridge Regression: Biased Estimation of Nonorthogonal Problems", *Technometrics*, 12, 55-67

Klein, Lawrence R. (1974), *A textbook of Econometrics*, Second edition, (Englewood Cliffs: Prentice- Hall, Inc.)

Koutsoyiannis, A. (1984), *Theory of Econometrics*, Second edition, (Hong Kong: Mcmillan Publishers Ltd.)

Maddala, G.S., 1992, *Introduction to Econometrics*, (Toronto: MacMillan Publishing Company.)

Malhotra, Naresh K., 2004, *Marketing Research*, (Englewood Cliffs: Prentice Hall, Inc.)

Malinvaud, E. (1980), *Statitiscal Methods of Econometrics*, Translated by MRS. A. Silvey, Third revised edition, (Amsterdam: North Holland Publishing Company)

Pindyck, Robert S. and Daniel L. Rubinfeld (1998), *Econometric Models and Econometric Forecasts,* Third edition, (New York: McGraw- Hill, Inc.)

Savin, N.E (1984), "Multiple Hypothesis Testing", *Handbook of Econometrics Volume 2* Edited by Zvi Griliches and Michael D. Intriligator, (Amsterdam: Elsevier Science Publishers B.V.)

Schmidt, Peter (1976), *Econometrics*, First edition, (New York: Marcel Dekker, Inc.)

Sekaran, Uma, 2003, *Research Methods For Business: A Skill Building Approach*, New York: John Wiley & Sons, Inc.)