Iranian Economic Review, Vol. 5, No. 5, 2001

Multicolinearity Reconsidered

By:
Hossein Abbasi-Nejad*
&
Shapoor Mohammady

ABSTRACT

In this paper we intend to improve the explanatory power of regressions when the deletion method is used for the remedy of multicolinearity. If one deletes the variable (s) that is (are) responsible for multicolinearity, he loses some information that is not common between the deleted variable (s) and the other remaining variables in the regression. To improve this method, we run the deleted variable (s) on the remaining variable (s) and use its residual as a new regressor in the main regression. Here we also focus on the multicolinearity concept that is related to the population and samples separately. We will also show if one encounters with perfect multicolinearity he can delete one of the variables without any biasedness costs. The procedure that saves some of the variables in the regression and input residuals of the other variable as regressors will give us "net effect regression".

I. Introduction

One of the data problems, which fault classical assumptions in regression analysis, is the presence of highly correlated variables in multiple regression. This problem which is called multicolinearity, reduces the efficiency of the estimation. Various diagnostic tests are provided for the realization of this problem, which are debated in details in various textbooks. Therefore lets not discuss them (the diagnostic tests) here. Many methods have previously been proposed for overcoming the problem.

1. Rigid regression: In this mechanical method due to the prevention

---

*. *Assistant Professor of Faculty of Economics, University of Tehran.*

from vanishing determinant of x′x, it is added by rD. Where r is arbitrary constant and D is a matrix of diagonal elements of x′x. By this method, the estimator will be more efficient but biased (Greene, 1993).

2. Principal component: When multicolinearity rises, one can say those independent variables are not independent from each other. In other words, the number of independent vectors of observations is less than k, where k is the number of parameters that should be estimated. So if we can find principle components that are actually independent, then the efficiency can be increased by the estimation of the model through those variables. But this approach to escape from inefficient estimators causes biasedness and a considerable degree of problematic interpretation of the results.

3. Usage of variables in deviation from the mean or from the polynomials.

4. Using transformations such as logarithms and differences.

5. Omission of variables which caused high multicolinearity.

All of the methods that we have mentioned have their own shortcomings and usually give biased estimators or problems in interpretation. How one can overcome multicolinearity is the question that will be answered in the later sections. But before that, it is useful to consider the nature of multicolinearity. Net effect regression and additional debates constitute the forth section, and a conclusion will appear in the last section.

II. Nature of Multicolinearity

In this section we will focus on multicolinearity. Suppose a sample "s" is chosen from the population "p". Two variables x_1 and x_2 are high correlated. Two events are possible. x_1, x_2 are not independent vectors in population. Then $x_1 . x_2 \cong 0$ and the dimension of our vector space is two instead of three, or vector space $X_1' = (x_1 00) X_2' = (0 x_2 0) Y = (00 y)$ is not the case corresponding to our population. One of the examples that multicolinearity is coming from the population is the estimation of consumption function. The three vector consumption (c), income (y) and wealth (w) are not actually independent. Therefore the dimension of the vector space is less than three and doesn't hold $w \perp y$. This is not the case that has been created in sampling process, because wealth and income are highly correlated (Branson 1989).

The second kind of multicolinearity is the case we called data problem, following textbooks such as (Greene 1993). When three variables are independent, we have:

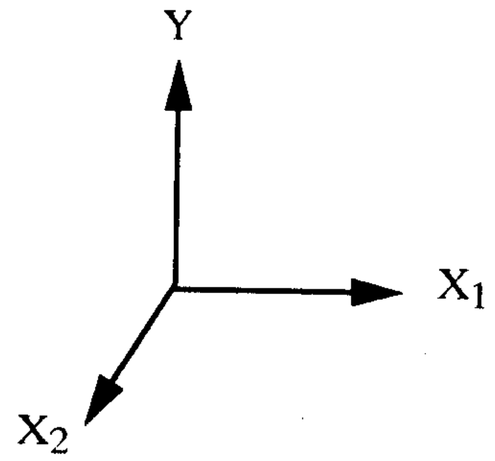


Fig. 1

Such that $X_1 \perp X_2 \perp Y$. However the amounts of x_1 and x_2 in sample are highly correlated. This kind of multicolinearity is related to the functional form of regression in terms of variables, but the first one is related to the specification analysis variable inclusion and exclusion.
For further analysis consider two models:

(A) $\quad Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

(B) $\quad Y = X_2\beta_2 + \varepsilon$

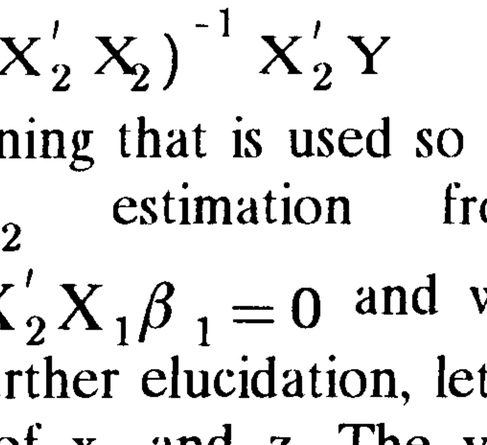
All of classical assumptions hold for them. If A is the correct model and β_2 is estimated from B β'_2 clearly biased, because form

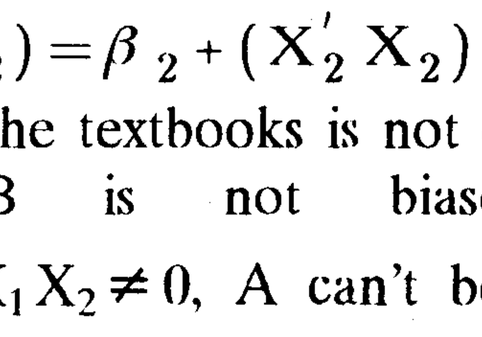
(A), $\beta'_2 = (X'_2 M_1 X_2)^{-1} X'_2 M_1 Y \quad E(\hat{\beta}_2) = \beta_2$

and form B

$$\hat{\beta}_2 = (X'_2 X_2)^{-1} X'_2 Y \qquad E(\hat{\beta}_2) = \beta_2 + (X'_2 X_2)^{-1} X'_2 X_1 \beta_1$$

This reasoning that is used so far in the textbooks is not correct. Since $x_1 x_2 = 0$, $\hat{\beta}_2$ estimation from B is not biased because $(X'_2 X_2)^{-1} X'_2 X_1 \beta_1 = 0$ and when $X_1 X_2 \neq 0$, A can't be the correct model. For further elucidation, let $x_2 = v_1 x_1 + v_2 z$, such that x_2 is the linear combination of x_1 and z. The variable z was deleted from the model, therefore the model is not properly specified. On the other hand, if $v_2 = 0$, then $x_2 = v_1 x_1$, and our vector space which virtually has three vectors $(x_1 x_2 y)$, actually has two vectors. In other words, the dim of V (our vector space) is 2: (Lipchutz 1989, p.69).





Alternatively; first, multicolinearity is not necessarily data problem and second, if regressors are highly correlated, then the model is not correctly specified. "What is the correct specification" is the problem that we answer in the next section which is inturn the remedy of multicolinearity.

III. The Specification and remedy of multicolinearity

Suppose our medel is specified in the following form:

$$^{*}Y = X_1\beta_1 + X_2\beta_2 + \varepsilon \quad \varepsilon \sim N(0,\sigma^2)$$

$$x_2 = v_1 x_1 + v_2 z \qquad \text{with all of the classical assumptions held.}$$

z is nonstochastic because of the assumption of, "x_1, x_2" being nonstochastic.

The Specifications that are assigned by a star are obviously problematic, because some parts of x_1 are included in the x_2. How can one cancel the common effect between x_1 and x_2?

A simple answer which one can imagine, is the OLS,

$$X_2 \sim X_1 \Rightarrow \hat{\nu}_1$$

Compute: $\hat{e}_{x_2} = X_2 - \hat{\nu}_1 X_1$

Make a model such as:

$$Y = X_1\beta_1 + \hat{e}_{x_2}\xi + \xi \qquad \xi \sim \text{iid}\, N(0,\sigma^2)$$

And other classical assumptions held for it. The two characteristics of the recent model are interesting:

i) $\hat{e}_{x_2} \perp X_1$

ii) The remained explanation power of x_2 is included in the model.

The specification, which uses $\hat{e}_{x_2}$, is the best specification. It is easy to show that estimators of β_1, ξ are unbiased and the variance of $\hat{\beta}_1, \hat{\xi}$ is smaller than the case which x_1, x were in the model simultaneously.

$$\hat{\beta}_1 = (X_1' M_{\hat{e}_{x_2}} X_1)^{-1} X_1' M_{e_{x_2}} Y$$

Where $$M_{\hat{e}_{x_2}} = I - \hat{e}_{x_2}(\hat{e}_{x_2}'\hat{e}_{x_2})^{-1} e_{x_2}'$$

and $$M_1 = I - X_1(X_1'X_1)^{-1}X_1'$$

$$\begin{aligned} E(\hat{\beta}_1) &= E[(X_1' M_{\hat{e}_{x_2}} X_1)^{-1} X_1' M_{\hat{e}_{x_2}} Y] \\ &= E[(X_1' M_{\hat{e}_{x_2}} X_1)^{-1} X_1' M_{\hat{e}_{x_2}} (X_1\beta_1 + \hat{e}_{x_2}\xi + \xi)] \\ &= \beta_1 + 0 + E(X_1'\hat{e}_{x_2}X_1)^{-1} X_1' M_{\hat{e}_{x_2}} Y] \\ &= \beta_1 \end{aligned}$$

Since $$(X_1' M_{\hat{e}_{x_2}} X_1)^{-1} X_1' M_{\hat{e}_{x_2}} \hat{e}_{x_2}\xi$$

$$= [I - \hat{e}_{x_2}(\hat{e}_{x_2}'\hat{e}_{x_2})^{-1} e_{x_2}']$$

$$= 0$$

Then the second term in the right hand side of the equation equals to zero and $E(\zeta)=0$ according to assumption 1. for: $\hat{\xi}$

$$E(\hat{\xi}) = E[(\hat{e}'_{x_2} M_1 \hat{e}_{x_2})^{-1} \hat{e}'_{x_2} M_1 (X_1\beta_1 + \hat{e}_{x_2}\xi + \xi)]$$
$$= 0 + \xi + 0$$
$$= 0$$

For verifying that the variance of parameters before and after deletion, are different and using residuals can improve the efficiency of estimation, we can write:

$$\text{var}(\hat{\beta}_1) = \sigma_\varepsilon^2 (X_1' M_2 X_1)^{-1} \geq \sigma'_\varepsilon (X_1' X_1)^{-1}$$
$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

In the later model

$$Y = X_1\beta_1 + \hat{e}_{x_2}\xi + \xi$$
$$\text{var}(\hat{\beta}_1) = \sigma_\varepsilon^2 (X_1' X_1)^{-1}$$

But this is the matter we knew before because of $X_1 \perp \hat{e}_{x_2}$, and which is new is the unbiasedness of estimated parameters through the deletion method.

Alternatively any multicolinearity is a de facto misspecification problem. This is the time for deciding about the deletion of x_1 or x. Two approaches can be suggested for this problem.

i) The partial correlation coefficient method: In this method, researcher auks variables according to their relevance to the dependant variable (Y in our model), then each variable (s) that has (have) high correlation to y should remain in the model. For this purpose it is necessary to calculate the practical correlation. But before that, our data should be transformed to a stationary one. After doing so one can regress Y on all of the independent variables and attain partial correlation coefficients, using the following formula (Greene 1993):

$$P_j = \frac{t_j}{\sqrt{t_j^2 + n - k}}$$

Where j is the index of the variable that partial correlation coefficient should be calculated for it, k is the number of regressors, and t_j is the t-statistic that is obtained from regression.

After this stage a researcher will keep the variable with higher correlation coefficient and delete the other. To avoid biasedness, the method mentioned above can be used.

ii) Correlation with residuals:

The variable that is highly correlated with residuals vector is almost orthogonalto hyperplane[1].

Alternatively, that one which has the smallest correlation coefficient with residuals should be deleted.

The logic behind this claim is shown in the following graph.

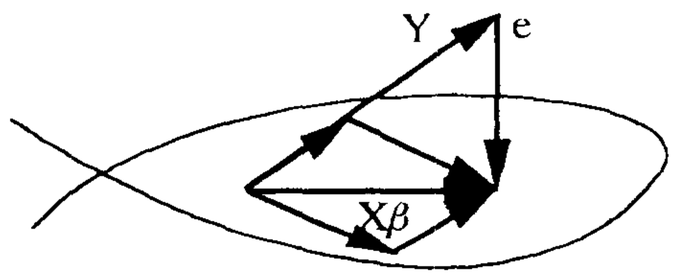
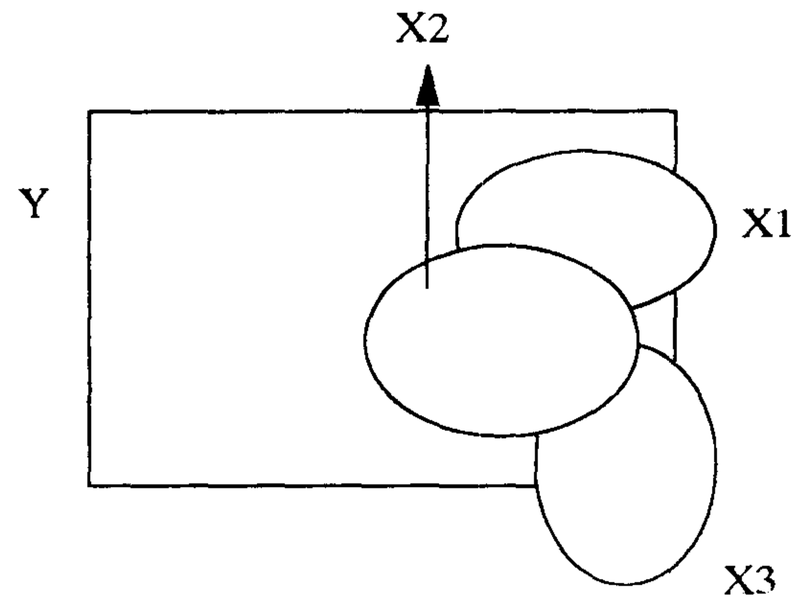


Fig.3

That variable that has the smallest correlation coefficient with e is the most dependent on $X\beta$, and is the main source of multicolinearity. Therefore it should be deleted. The question that may be raised at this stage is that the deleted variable may be the main source of variation in. The answer is if it has high correlation with Y, the other variables will explain the commonalties between them.



let:

$$S_{x_1 Y x_2 Y x_3} \equiv X\beta \qquad \text{hyperplan}$$

Then $r_{x_1 e} \phi r_{x_2 e}, r_{x_3 e} \phi r_{x_2 e}$

Where r is correlation coefficient. The correlation coefficient of x_2 and e is less than both x_2, x_3 and e. It can be seen that the explanation power of x_2 is greater than x_1, x_3, in explaining you in our example. However, the variable that should be deleted according to the previous reasoning is x_2, because of the correlation of x_2 and because it is the greatest in this case.

Another question that should be responded to is, that "What will happen if the residuals of deleted variables $(\hat{e}_{x_2})$ become highly correlated with main regression residuals $(\hat{\xi})$". In this case the entrance of $\hat{e}_{x_2}$ in the main regression, will promote R^2 to one. In order to show this, we write:

$$Y = X_1\beta_1 + \hat{e}'_{x_2}\xi + \xi$$

$$\hat{\beta} = (X'_1 M_{\hat{e}_{x_2}} X_1)^{-1} X'_1 M_{\hat{e}_{x_2}} Y$$

$$\hat{\gamma} = (\hat{e}'_{x_2} M_{x_1} \hat{e}_{x_2})^{-1} \hat{e}'_{x_2} M_{x_1} Y$$

$$\hat{\xi} = Y - X_1(X'_1 M_{\hat{e}_{x_2}} X_1)^{-1} X'_1 M_{\hat{e}_{x_2}} Y - \hat{e}_{x_2}(\hat{e}'_{x_2} M_{x_1} \hat{e}_{x_2})^{-1} \hat{e}_{x_2} M_{x_1} Y$$

$$= Y - X_1(X'_1 M_{\hat{e}_{x_2}} X_1)^{-1} X'_1 [I - \hat{e}_{x_2}(\hat{e}'_{x_2}\hat{e}_{x_2})^{-1}\hat{e}'_{x_2}]Y - \hat{e}_{x_2}(\hat{e}'_{x_2}\hat{e}_{x_2})^{-1}\hat{e}_{x_2}$$

$$= Y - X_1(X'_1 X_1)^{-1} X'_1 Y - \hat{e}_{x_2}(\hat{e}'_{x_2}\hat{e}_{x_2})^{-1}\hat{e}'_{x_2} Y$$

$$= M_{x_1} Y - M_{x_1} Y$$

$$= 0 \Rightarrow R^2 = 1$$

In this case regression will be an identity, but that is not likely to occur. This example is an extreme one that can not exist in the real world. In addition, one can select a variable for deletion that its residual won't be highly correlated with the residual of the main regression.

The other extreme case is the presence of high correlation between x_1, x_2 in the following regression so that the correlation coefficient between is one. Suppose,

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

And all of classical assumptions except full rank of $X'X$ held for it. We showed that by regressing x_2 on x_1 and get residuals, we could run the following regression:

$$Y = X_1\beta_1 + \hat{e}_{x_2}\xi + \xi$$

We know that when correlation coefficient between x_1, x_2 is equal to one, $\hat{e}_{x_2}$ is zero for all of the observations. Then we have:

if corr. $(X_1, X_2) \to 1 \Rightarrow \hat{e}_{x_2} \to 0$

$$\therefore \hat{\xi} = 0 \qquad \text{And}$$

$$Y = X_1\beta_1 + \hat{e}_{x_2}\xi + \xi$$

$$= X_1\beta_1 + \xi$$

$$Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$$

$$E(Y) = X_1\beta_1 + E(\xi)^*$$

$$= X_1\beta_1$$

$$E(Y) = X_1\beta_1 + X_2\beta_2 + E(\varepsilon)^{**}$$

$$= X_1\beta_1 + X_2\beta_2$$

From *and**, we conclude that

$$X_1\beta_1 = X_1\beta_1 + X_1\beta_2$$

When

$$X_2 \neq 0 \Rightarrow \beta_2 = 0$$

This shows that even when $Y = X_1\beta_1 + X_2\beta_2 + \varepsilon$ is the correct model virtually and for the remedy of multicolinearity, one uses the deletion method, estimation will be unbiased. Because in the following expression,

$$E(\beta_1) = \beta_1 + P_{12}\beta_2$$

β_2 Is zero. We saw that in two cases, omission of the variable doesn't cause bias costs when they are orthogonal and when they are linearly dependent.

IV. Generalization and net effect regression

Till now our implicit assumption was that one variable is the main source of multicolinearity and we try to handle it. Now we drop the assumption of the two regressors and suppose more than one variable is responsible for multicolinearity. In this section, models are stated in scalar form.

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon \qquad \varepsilon \sim \text{iidN}(0, \sigma^2)$$

And with validation of all classical assumptions for our model. Let x_2, x_3 be highly correlated, not only with x_1 but also with each other and the omission of one of them doesn't suffice for escaping from multicolinearity. In this case, one can make a regression such as x_3 on x_2 and x_1 and calculate its residuals.

$$X_3 = \gamma_0 + \gamma_1 X_1 + \gamma_2 X_2 + \eta$$

$$R_3 = (X_3 - \hat{\gamma}_0 - \hat{\gamma}_1 X_1 - \hat{\gamma}_2 X_2)$$

Also for $x_2 \sim x_1$

$$R_2 = (X_2 - \hat{\xi}_0 - \hat{\xi}_1 X_1)$$

Then our main regression will be

$$Y = \alpha + \beta_1 X_1 + \theta_3 R_3 + \varepsilon$$

The vector space of the regression above is four-dimensional. For a k-variable model we can write

$$Y = \alpha + \sum_{i=1}^{k} \beta_i x_{it} + \varepsilon$$

Suppose, x_{1t}, ..., x_{st} are independent of each other and the (k-s) remained variables $x_{(s+1)t}$ are the variables that must be served in the regression, then:

$$X_{kt} \sim X_{(k-1)\,t}, X_{(k-2)\,t}, X_{(s+1)\,t} \rightarrow R_{kt}$$
$$X_{k-1} \sim X_{(k-2)\,t}, X_{(k-3)\,t}, X_{(s+1)\,t} \rightarrow R_{(k-1)\,t}$$
$$\vdots$$
$$X_{k-(k-s-2)} \sim X_{(s+1)\,t} \rightarrow R_{(s+2)\,t}$$

Now one can write the main regression in the following form

$$Y = X\beta + \varepsilon$$

or

$$Y = [X\ M\ R] \begin{bmatrix} \beta \\ \Lambda \\ \theta \end{bmatrix} + \varepsilon$$

or

$$Y = X\beta + R\theta + \varepsilon$$

$$X = \begin{bmatrix} 1 & X_{11} & \Lambda & X_{1s} \\ 1 & X_{21} & \Lambda & X_{2s} \\ M & M & M & M \\ 1 & X_{t1} & \Lambda & X_{ts} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \alpha \\ \beta_1 \\ \beta_2 \\ M \\ \beta_s \end{bmatrix}$$

$$R = \begin{bmatrix} R_{1\,(s+2)} & R_{1\,(s+3)} & \Lambda & R_{1\,[s-(s-k)]} \\ R_{2\,(s+2)} & R_{2\,(s+3)} & \Lambda & R_{2\,[s-(s-k)]} \\ M & M & M & M \\ R_{t\,(s+2)} & R_{t\,(s+3)} & \Lambda & R_{t\,[s-(s-k)]} \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_{s+2} \\ \theta_{s+3} \\ M \\ \theta_{s-(s-k)} \end{bmatrix} = \begin{bmatrix} \theta_2 \\ \theta_3 \\ M \\ \theta_k \end{bmatrix}$$

x is the main source of regressions explanatory power and all of the other R variable are net effects of x_{s+2}, ..., X_k on Y.

The estimation process is very easy, because of the orthogonality of variables. $x'x$ is a diagonal matrix and parameters are in the following form.

$$\begin{bmatrix} \hat{a} \\ \hat{\beta}_1 \\ M \\ \hat{\beta}_s \\ \hat{\theta}_{s+1} \\ \hat{\theta}_{s+2} \\ M \\ \hat{\theta}_k \end{bmatrix} = \begin{bmatrix} \overline{Y} - \hat{\beta}_1 \overline{X}_1 - \Lambda - \hat{\beta}_s \overline{X}_s - \hat{\theta}_{s+1} \overline{R}_{s+1} - \Lambda - \hat{\theta}_k \overline{R}_k \\ \dfrac{\Sigma x_{1t} y_t}{\Sigma x_{1t}^2} \\ M \\ M \\ M \\ M \\ M \\ \dfrac{\Sigma R_{kt} y_t}{\Sigma R_{kt}^2} \end{bmatrix}$$

On can show that the estimators are unbiased and efficient.

V. Concluding remarks

In this paper we have covered topics that are related to multicolinearity and got these results.

(1). Multicolinearity is not a data problem, necessarily.

(2). Every multicolinearity problem is a potential specification error.

(3). Net effect regression is the method that can be used effectively for the improvement of explanatory power of the model that the variables are picked up from it.

(4). Partial correlation coefficient and correlation coefficient with residuals are the methods that can be used for the determination of important variables for saving in models and less relevant variables from being deleted.

For future studies, it is proposed that empirical works be done with simultations and proper methods.

References

1. *Amemiya, T. (1985), Advanced Econometric, Harvard university press. USA*
2. *Branson, W, H. (1989). Macroeconomic analysis and policies. International student Edition.*
3. *Greene, W.H. (1993). Econometric analysis, Mc Graw-Hill, New York.*
4. *Gujarati, D. (1995) Basic econometric, international student edition, New York.*
5. *Johnston (1983). Econometrics methods, international student edition, New York.*
6. *Lipschuts, S. (1989). Linear Algebra, Schaum's outline series, Mc Grow-Hill, New York.*