

LINEAR REGRESSION ANALYSIS

MODULE – IX

Lecture - 30

Multicollinearity

Dr. Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

Remedies for multicollinearity

Various techniques have been proposed to deal with the problems resulting from the presence of multicollinearity in the data.

1. Obtain more data

The harmful multicollinearity arises essentially because rank of $X'X$ falls below k and then $|X'X| = 0$ which clearly suggests the presence of linear dependencies in the columns of X . It is the case when $|X'X|$ is close to zero which needs attention. Additional data may help in reducing the sampling variance of the estimates. The data need to be collected such that it helps in breaking up the multicollinearity in the data.

It is always not possible to collect additional data to various reasons as follows.

- The experiment and process have finished and no longer available.
- The economic constraints may also not allow to collect the additional data.
- The additional data may not match with the earlier collected data and may be unusual.
- If the data is in time series, then longer time series may force to take data that is too far in the past.
- If multicollinearity is due to any identity or exact relationship, then increasing the sample size will not help.
- Sometimes, it is not advisable to use the data even if it is available. For example, if the data on consumption pattern is available for the years 1950-2010, then one may not like to use it as the consumption pattern usually does not remain same for such a long period.

2. Drop some variables that are collinear

If possible, identify the variables which seems to causing multicollinearity. These collinear variables can be dropped so as to match the condition of full rank of X -matrix. The process of omitting the variables may be carried out on the basis of some kind of ordering of explanatory variables, e.g., those variables can be deleted first which have smaller value of t -ratio.

In another example, suppose the experimenter is not interested in all the parameters. In such cases, one can get the estimators of the parameters of interest which have smaller mean squared errors than the variance of OLSE of full vector by dropping some variables. If some variables are eliminated, then this may reduce the predictive power of the model. Sometimes there is no assurance that how the model will exhibit less multicollinearity.

3. Use some relevant prior information

One may search for some relevant prior information about the regression coefficients. This may lead to specification of estimates of some coefficients. More general situation includes the specification of some exact linear restrictions and stochastic linear restrictions. The procedures like restricted regression and mixed regression can be used for this purpose.

The relevance and correctness of information plays an important role in such analysis but it is difficult to ensure it in practice. For example, the estimates derived in U.K. may not be valid in India.

4. Employ generalized inverse

If $\text{rank}(X'X) < k$, then the generalized inverse can be used to find the inverse of $X'X$. Then β can be estimated by $\hat{\beta} = (X'X)^- X'y$.

In such case, the estimates will not be unique except in the case of use of Moore-Penrose inverse of $(X'X)$. Different methods of finding generalized inverse may give different results. So applied workers will get different results. Moreover, it is also not known that which method of finding generalized inverse is optimum.

5. Use of principal component regression

The principal component regression is based on the technique of principal component analysis. The explanatory variables are transformed into a new set of orthogonal variables called as principal components. Usually this technique is used for reducing the dimensionality of data by retaining some levels of variability of explanatory variables which is expressed by the variability in study variable. The principal components involves the determination of a set of linear combinations of explanatory variables such that they retain the total variability of the system and these linear combinations are mutually independent of each other. Such obtained principal components are ranked in the order of their importance. The importance being judged in terms of variability explained by a principal component relative to the total variability in the system. The procedure then involves eliminating some of the principal components which contribute in explaining relatively less variation. After elimination of the least important principal components, the set up of multiple regression is used by replacing the explanatory variables with principal components.

Then study variable is regressed against the set of selected principal components using ordinary least squares method. Since all the principal components are orthogonal, they are mutually independent and so OLS is used without any problem. Once the estimates of regression coefficients for the reduced set of orthogonal variables (principal components) have been obtained, they are mathematically transformed into a new set of estimated regression coefficients that correspond to the original correlated set of variables. These new estimated coefficients are the principal components estimators of regression coefficients.

Suppose there are k explanatory variables X_1, X_2, \dots, X_k . Consider the linear function of X_1, X_2, \dots, X_k like

$$Z_1 = \sum_{i=1}^k a_i X_i$$

$$Z_2 = \sum_{i=1}^k b_i X_i \text{ etc.}$$

The constants a_1, a_2, \dots, a_k are determined such that the variance of Z_1 is maximized subject to the normalizing condition that $\sum_{i=1}^k a_i^2 = 1$. The constant b_1, b_2, \dots, b_k are determined such that the variance of Z_2 is maximized subject to the normality condition that $\sum_{i=1}^k b_i^2 = 1$ and is independent of the first principal component.

We continue with such process and obtain k such linear combinations such that they are orthogonal to their preceding linear combinations and satisfy the normality condition. Then we obtain their variances. Suppose such linear combinations are Z_1, Z_2, \dots, Z_k and for them, $Var(Z_1) > Var(Z_2) > \dots > Var(Z_k)$. The linear combination having the largest variance is the first principal component. The linear combination having the second largest variance is the second largest principal component and so on. These principal components have the property that $\sum_{i=1}^k Var(Z_i) = \sum_{i=1}^k Var(X_i)$. Also, the X_1, X_2, \dots, X_k are correlated but Z_1, Z_2, \dots, Z_k are orthogonal or uncorrelated. So there will be zero multicollinearity among Z_1, Z_2, \dots, Z_k .

The problem of multicollinearity arises because X_1, X_2, \dots, X_k are not independent. Since the principal components based on X_1, X_2, \dots, X_k are mutually independent, so they can be used as explanatory variables and such regression will combat the multicollinearity.

Let $\lambda_1, \lambda_2, \dots, \lambda_k$ be the eigenvalues of $X'X$, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_k)$ is $k \times k$ diagonal matrix, V is a $k \times k$ orthogonal matrix whose columns are the eigenvectors associated with $\lambda_1, \lambda_2, \dots, \lambda_k$.

Consider the canonical form of the linear model

$$\begin{aligned} y &= X\beta + \varepsilon \\ &= XVV'\beta + \varepsilon \\ &= Z\alpha + \varepsilon \end{aligned}$$

where

$$Z = XV, \alpha = V'\beta, V'X'XV = Z'Z = \Lambda$$

Columns of $Z = (Z_1, Z_2, \dots, Z_k)$ define a new set of explanatory variables which are called as **principal component**.

The OLSE of α is

$$\begin{aligned}\hat{\alpha} &= (Z'Z)^{-1}Z'y \\ &= \Lambda^{-1}Z'y\end{aligned}$$

and its covariance matrix is

$$\begin{aligned}V(\hat{\alpha}) &= \sigma^2(Z'Z)^{-1} \\ &= \sigma^2\Lambda^{-1} \\ &= \sigma^2 \text{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_k}\right).\end{aligned}$$

Note that λ_j is the variance of j^{th} principal component and $Z'Z = \sum_{i=1}^k \sum_{j=1}^k Z_i Z_j = \Lambda$.

A small eigenvalue of $X'X$ means that the linear relationship between the original explanatory variable exist and the variance of corresponding orthogonal regression coefficient is large which indicates that the multicollinearity exists. If one or more λ_j are small, then it indicates that multicollinearity is present.

Retainment of principal components

The new set of variables, i.e., principal components are orthogonal, and they retain the same magnitude of variance as of original set. If multicollinearity is severe, then there will be at least one small value of eigenvalue. The elimination of one or more principal components associated with smallest eigenvalues will reduce the total variance in the model. Moreover, the principal components responsible for creating multicollinearity will be removed and the resulting model will be appreciably improved.

The principal component matrix $Z = [Z_1, Z_2, \dots, Z_k]$ with Z_1, Z_2, \dots, Z_k contains exactly the same information as the original data in X in the sense that the total variability in X and Z is same. The difference between them is that the original data are arranged into a set of new variables which are uncorrelated with each other and can be ranked with respect to the magnitude of their eigenvalues. The j^{th} column vector Z_j corresponding to the largest λ_j accounts for the largest proportion of the variation in the original data. Thus the Z_j 's are indexed so that $\lambda_1 > \lambda_2 > \dots > \lambda_k > 0$ and λ_j is the variance of Z_j .

A strategy of elimination of principal components is to begin by discarding the component associated with the smallest eigenvalue. The idea behind to do this is that the principal component with smallest eigenvalue is contributing the least variance and so is least informative.

Using this procedure, principal components are eliminated until the remaining components explain some preselected variance in terms of percentage of total variance. For example, if 90% of total variance is needed, and suppose r principal components are eliminated which means that $(k - r)$ principal components contribute 90% of the total variation, then r is selected to

$$\frac{\sum_{i=1}^{k-r} \lambda_i}{\sum_{i=1}^k \lambda_i} > 0.90.$$

Various strategies to choose required number of principal components are also available in the literature.

Suppose after using such a rule, the r principal components are eliminated. Now only $(k - r)$ components will be used for regression. So Z matrix is partitioned as

$$Z = (Z_r \quad Z_{k-r}) = X(V_r \quad V_{k-r})$$

where submatrix Z_r is of order $n \times r$ and contains the principal components to be eliminated. The submatrix Z_{k-r} is of order $n \times (k - r)$ and contains the principal components to be retained.

The reduced model obtained after the elimination of r principal components can be expressed as

$$y = Z_{k-r} \alpha_{k-r} + \varepsilon^*.$$

The random error component is represented as ε^* just to distinguish with ε . The reduced coefficients contain the coefficients associated with retained Z_j 's. So

$$Z_{k-r} = (Z_1, Z_2, \dots, Z_{k-r})$$

$$\alpha_{k-r} = (\alpha_1, \alpha_2, \dots, \alpha_{k-r})$$

$$V_{k-r} = (V_1, V_2, \dots, V_{k-r}).$$

Using OLS on the model with retained principal components, the OLSE of α_{k-r} is

$$\hat{\alpha}_{k-r} = (Z'_{k-r} Z_{k-r})^{-1} Z'_{k-r} y.$$

Now it is transformed back to original explanatory variables as follows:

$$\alpha = V' \beta$$

$$\alpha_{k-r} = V'_{k-r} \beta$$

$$\Rightarrow \hat{\beta}_{pc} = V_{k-r} \hat{\alpha}_{k-r}$$

which is the **principal component regression estimator** of β .

This method improves the efficiency as well as combats multicollinearity.