[Frequently Asked Questions]

Prediction in Linear Models and Multicollinearity

Subject:	Business Economics
Course:	B. A. (Hons.), 5th Semester, Undergraduate
Paper No. & Title:	Paper – 531 Elective Paper Q1 – Advanced Econometrics
Unit No. & Title:	Unit – 1 Relaxing the Assumptions of The Classical Linear Model
Lecture No. & Title:	Lecture – 3 Prediction in Linear Models and Multicollinearity

Frequently Asked Questions

Q1. What is prediction in linear models?

A1. It is the estimated value of Y for given value of x .In k variables model, it is the estimated value of Y when all explanatory variables have some fixed or known values.

Q2. What is the prediction error?

A2. Prediction error is the difference between actual value of Y and estimated value of Y. (Thus prediction error = $e_0 = Y_0 - \hat{Y}_0$)

Q3. What is the mean value prediction?

A3. Many times we may be interested to know about $E(Y_0) = E(Y|X_0)$ rather than actual value of Y given X_0. This is called mean value prediction.

Here prediction error = $E(Y_0) - \hat{Y}_0$.

Q4. What is meant by the term multicollinearity?

A4. When two or more explanatory variables are between themselves linearly related, multicollinearity arises. If such a relation is prefect than it is perfect multicollinearity, otherwise it is called less than or imperfect multicollinearity.

Q5. What will happen due to perfect multicollinearity?

A5. In the case of perfect multicollinearity, the regression coefficient are indeterminate and also their standard errors will be infinite hence indeterminate.

Q6. What is the effect of imperfect multicollinearity?

A6. In the case of less than (or imperfect) multicollinearity, the regression coefficients are obtainable but their standard errors will be large.
This results in conclusions from the fitted model which are not reliable.

Q7. If R^2 is high but no regression coefficients are statistically significant, what would you say?
A7. We suspect multicollinearity. It could be severe also.

Q8. If $|X'X| \approx 0$ ($i.e.$ Determinant of $X'X$ is very near to zero what would you suspect?
A8. We suspect multicollinearity. It would be severe also.

Q9. How can you detect multicollinearity on the basis of smallest and largest Eigen value of $X'X$?
A9. We can find condition Index CI= $\sqrt{\frac{\lambda_{max}}{\lambda_{min}}}$. If CI lies between 10 to 30, there is moderate to strong multicollinearity. It is severe if CI exceeds 30.

Q10. What is auxilliary regression?
A10. To detect multicollinearity, we can run the regression of X_i on all the other $X's$.
This is called auxilliary regression.
We compute R_i^2 for such a regression and compute the ratio

$$F_i = \frac{R_i^2 / (k-2)}{(1-R_i^2) / (n-k+1)}.$$

Which is F statistic with degrees of freedom $(k-1)$ and

$(n-k+1)$. If F_i is significant, it shows that there is multicollinearity, as X_i is linearly related with other $X's$.

Q11. What is the simplest way to overcome the problem of multicollinearity ?

A11. If we observe that some variable X_2 is linearly related with X_3 $(or\, X_4)$ we can drop X_2 as it is superfluous causing the problem of multicollinearity.

Q12. Is Ridge regression a suitable method for overcoming the problem of Multicollinearity?

A12. Perhaps yes. This approach is in fact related to augmenting the data matrix X by adding some more observations.

Ridge regression estimator is given by $\underline{\hat{\beta}}_R = (X'X + CI_n)^{-1} X'\underline{y}$ With

$V\left(\underline{\hat{\beta}}_R\right) = \sigma^2[(X'X + CI)^{-1} X'X(X'X + CI)^{-1}]$

Here C is a positive constant chosen in such a way that

$E\left[\left(\underline{\hat{\beta}}_R - \underline{\beta}\right)' (\underline{\hat{\beta}}_R - \underline{\beta})\right]$ is minimum.

Ridge Regression estimator is a biased estimator.