Text Version

Expand Print

Consequences of Multicollinearity for Polynomial Regression

Let's talk about what happens when multicollinearity is present in a polynomial regression model. Remember that, if the assumptions for GLMs are satisfied, the ordinary least square estimators of the regression estimators are best linear unbiased estimates (a property that is known as BLUE). Even if high multicollinearity is present, the OLS estimators retain the property of BLUE. Then why is multicollinearity a problem?

Multicollinearity affects the results of regression in the following ways: It affects the parameter estimates because they depend on the correlated independent variables included in the regression model. In extreme cases, multicollinearity can cause the parameter estimates to have the wrong sign. It can also cloud the practical meaning of the true regression coefficients (Bowerman, O’Connell, and Dickey 1986). Multicollinearity can inflate the standard errors of the parameter estimates and the predictions. Note that, to test the null hypothesis that a parameter estimate is not significantly different from zero, you use the t-test. The t-test formula uses the standard error of the parameter estimate in the denominator, which affects the calculated value of the t-statistic and hinders your ability to determine the significance of the independent variables.

The overall fit of the equation is largely unaffected by multicollinearity. Because multicollinearity has little effect on the overall fit of the equation, it also has little effect on the use of that equation for prediction, as long as the independent variables maintain the same pattern of multicollinearity in the future period that they demonstrated in the sample (Bowerman, O’Connell, and Dickey 1986). Multicollinearity can also lead to serious rounding errors when you compute the point estimates of the parameters due to a nearly dependent X matrix.

Close