7.2: Principal Components and Quantile Regression

Dr. Bean - Stat 5100

Suppose you are a dairy farmer trying to determine if a change in feed will lead to a significant increase in milk production, after accounting for the weight and heredity of the dairy cows. You notice that the model residuals are heteroskedastic across weight, though you determine that you can eliminate the heteroskedasticity with a log-transformation.

For this scenario, provide at least one argument in favor of OLS regression with the transformed data, vs quantile regression with the un-transformed data.

Recall the form of the check loss functions. Why would a check loss function be more robust to outlier values than a squared loss function?

What would have to be true of the model residuals if the mean predictions (OLS) were significantly different from the median predictions of quantile regression?

What would be the issue with trying to estimate the quantile regression model associated with the 95th percentile with a small sample size? Would you have the same problem trying to estimate the model associated with the median?