Problem 4.4

(a) is variable represents the noise level, therefore is the normal distributed noise on the data since is an independent and identically distributed random variable. We normalize the function so that noise level is calibrated to the function level.

The orthogonality of Legendre polynomials:

To normalize function and obtain :

Applying the orthogonality of the Legendre polynomials:

Because is a standard normal with and

Therefore the normalization factor is .

(b)

First we have the formula for and in the orthogonal Legendre polynomials form

With training data

We calculated the value of Legendre polynomial from order 0 to 10 at each . The training data hence becomes for

Then is a three dimensional linear regression.

Similarly is an eleven dimensional linear regression.

(c)

The actual target function is and the hypothesis is

Therefore we have the for as below:

Because we have normal distributed with mean and variance

We have

(d) (The graphs are attached at the end of the answer.)

The over fit measure is significantly positive when the noise level is high, while the number of data points and target function complexity are fixed. Because of the noise, the data deviates from the actual target function. Yet, has more power to fit the data, which results in more significant out of sample error because also tries to fit the noise.

The over fit measure is also significantly positive when the target function complexity is high, while the number of data points is low (). The deterministic noise is dependent on target function complexity. Similarly has more power to fit the data including the noise, resulting a significant out of sample error. However, when the number of data points become bigger, the higher model complexity leads to lower error measure.

Additionally, when the number of data points is small, serious overfitting also occurs. In other words, as the number of data points grows larger, the error measure grows smaller. The reason is that with more data points, the trained model is more closed to the actual target.

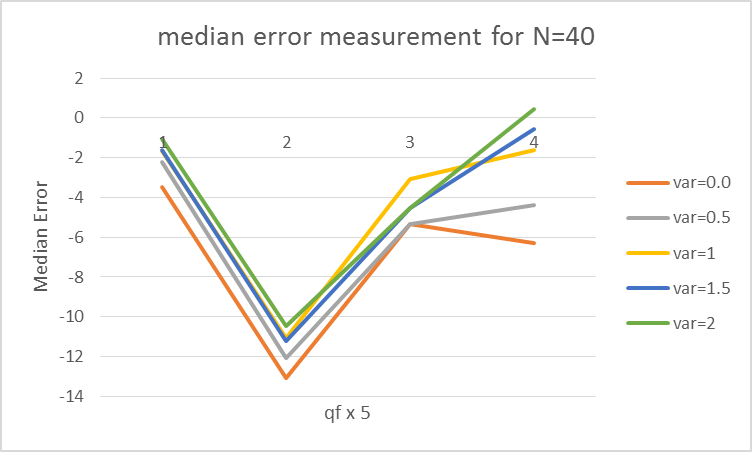
All in all, the over fit measure is negative when the noise is low, the target function complexity is close to and the number of data points is large. In that case, has a more power to fit the actual correct data (with minimal noise). Therefore the out of sample error of is small and hereby the over fit measure is small.

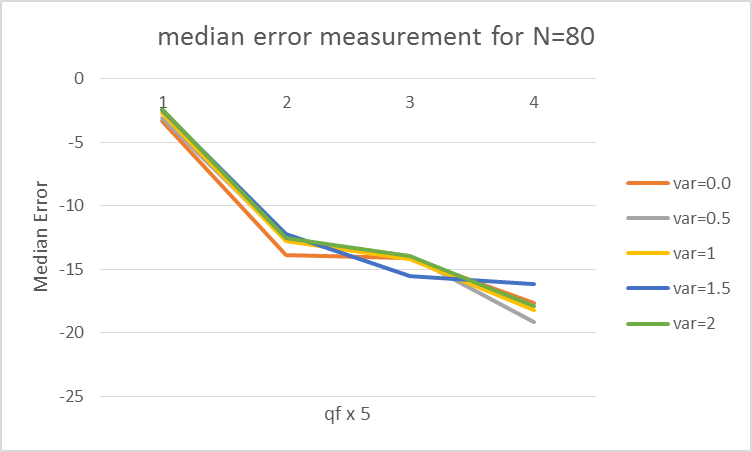
The mean and median are in general very close to each other. However, when the mean overfit measure is high, the difference between the mean and median tends to become larger. This may be that when mean is high, most values are high and it’s likely that certain estimate produce a terrible extreme value and hence influence the mean, resulting in a larger difference between the mean and median. On the contrary, median error measurements remains relatively consistent since it is not heavily influenced by extreme values.

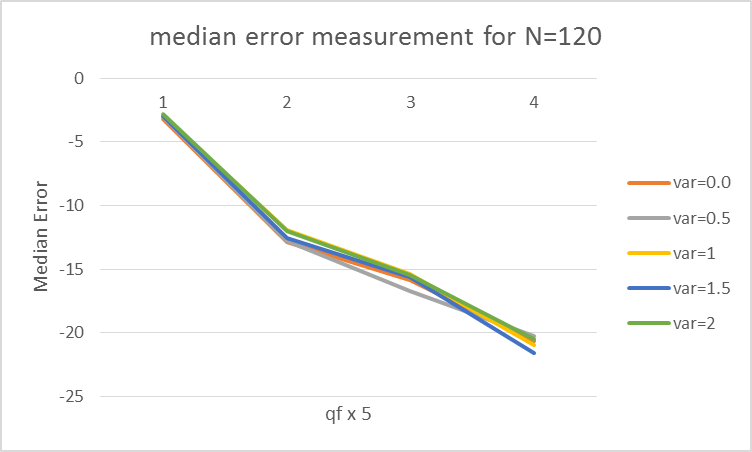
Below is the table for all error measurements for all variances, QFs and sample sizes.

The entries on the left side should be multiplied by the number after var=x. For example, the first entry should be . The right side entries need not to be multiplied.









1. In each graph, we can see that model complexity (qf) have different effects based on the number of data points. When the number of data points are small, high complexity leads to severe overfitting and when the number of data points are big, high complexity leads to smaller median error.
2. And we can see that as goes from 40 to 120, the line shift down in the median error measure across the graphs.
3. The variance have small effects on the overall error measurements. Yet overall the less the variance, the smaller error measurements.