Measurement and Data Driven Modelling

Lab 4: Model Order Selection

Warre De Winne, Arno De Haseleer

05/05/2022

3.4.3. Plot together the evolution of VLS, Vval, VAIC and compare the estimate of the optimal model order (the minima of each curve) between the methods. Compare and interpret the three curves.

In the following graph (figure 1) one can see the comparison between the two different estimation methods and the result of the validation dataset.

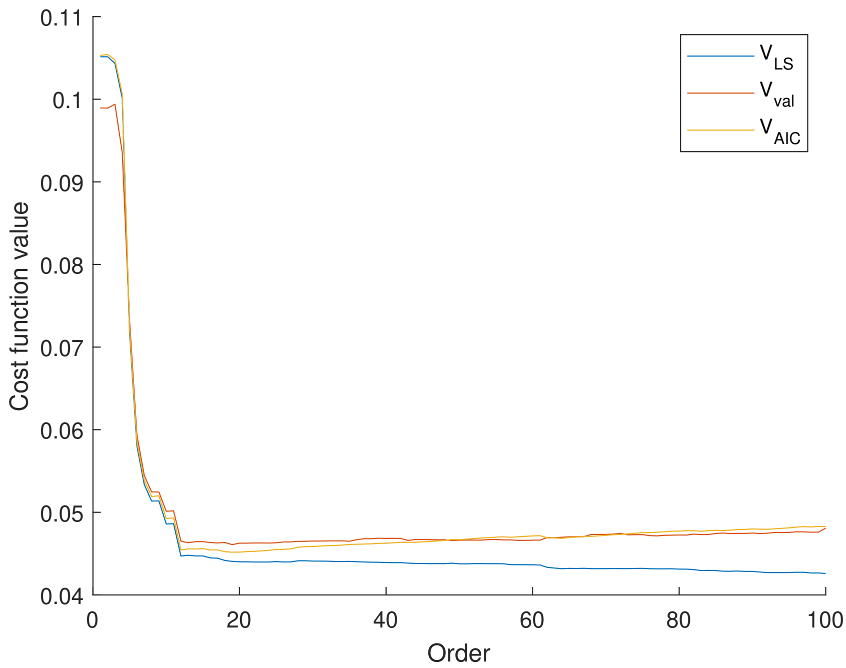


Figure 1: Comparison of the different cost functions

The following orders minimize the estimations:

V_{LS}	100
V_{val}	19
V_{AIC}	19

The model order that corresponds to the minimum of each curve is the estimate of the optimal model order using that specific method. Both V_{val} and V_{AIC} are minimal at 19, this means that these criteria have identified 19 as the optimal model order. This suggests that a model of order 19 provides the best balance between model complexity and how well it fits the system.

V_{LS} is minimal at 100, which is the maximum order value. One can see that the graph of V_{LS} follows the other two closely until $n = 19$, and then keeps decreasing further. When increasing the model order, more parameters are added to the model. It allows to better fit the data. However, adding too many parameters to the model can lead to overfitting, where the model becomes too complex and begins to fit the noise in the data instead of the underlying patterns. This is why it is important to use model selection criteria, like V_{AIC}, to help identify the optimal model order.

3.5.1. Increase the SNR by 20 dB. Repeat the estimation and the model order selection and interpret the results. What are the differences and why do they occur?

For a SNR = 26dB, the obtained values are seen in figure 2. Compared to the measurements with a SNR of 6 dB, all cost functions keep decreasing in value past $n = 19$. This makes sense, since increasing SNR means that there is less noise on the data. Therefore, there will be less risk of overfitting the model. Increasing the order will therefore always slightly improve the model of the system.

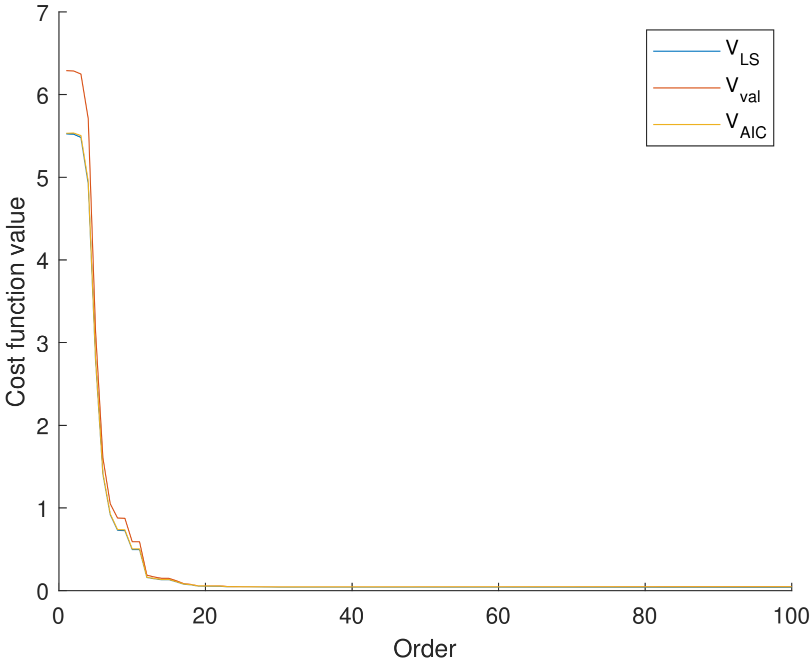


Figure 2: Comparison of the different cost functions with SNR = 26dB

3.5.2. Decrease the bandwidth of the original system. What influence does that have on the optimal model order? Explain.

The bandwidth of the passband filter is decreased from $\omega = [0.3, 0.6]$ to $\omega = [0.4, 0.5]$. The comparison can be seen in figure 3. One can see that for all criteria, the cost function values are a bit lower, and that the minimum for V_{val} and V_{AIC} lies at $n = 21$. The lower overall values of the criteria suggest that the model is now able to capture more of the important features of the data, even at higher order models. The slightly higher minimum value at $n = 21$ suggests that a slightly more complex model can capture all of the relevant information of the system, without overfitting.

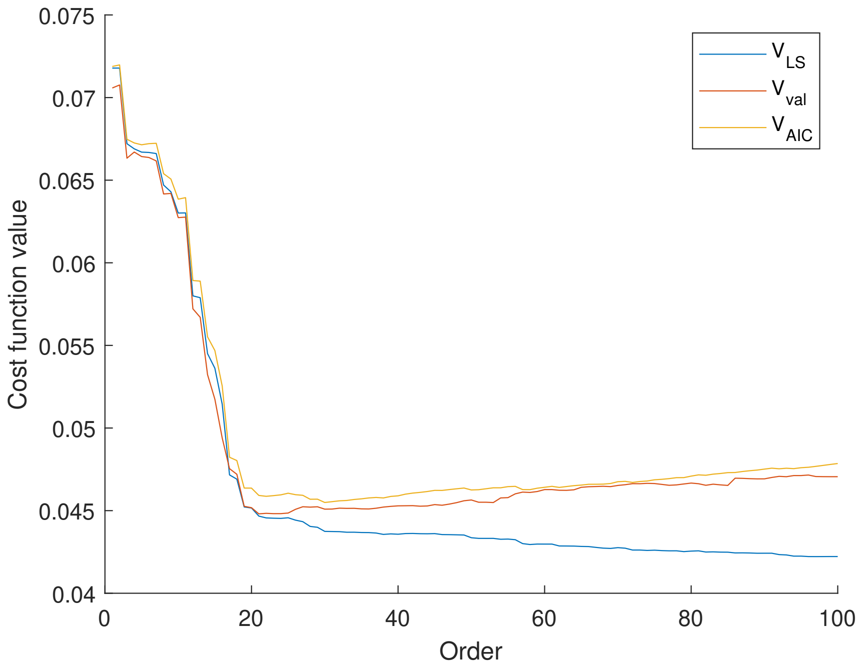
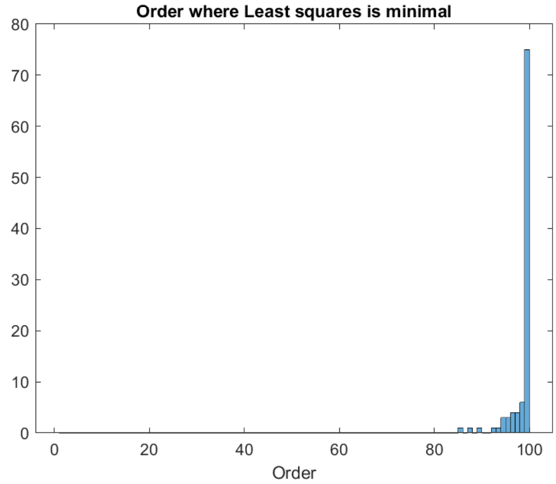


Figure 3: Comparison of the different cost functions with BW: $\omega = [0.4, 0.5]$

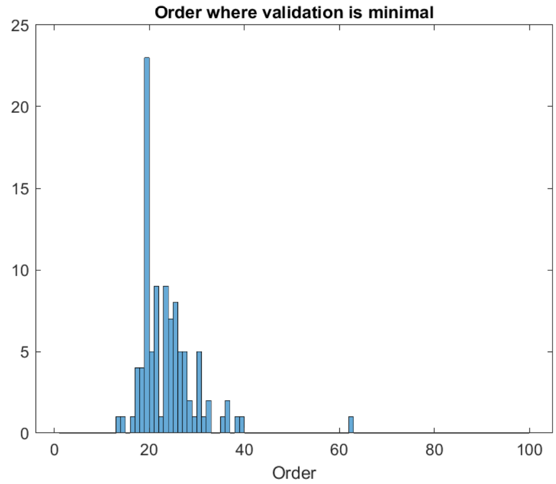
3.6. Histograms

The histograms for all three estimations for 100 iterations can be seen in figure 4. Like expected, the histogram of V_{val} and V_{AIC} have the most entries at the bins of $n = 19$. This clear peak shows that this order is consistently selected as the optimal model order, showing a high degree of confidence in the selection. The peak is slightly narrower for V_{AIC}, which means that it has the most confidence in the model order selection.

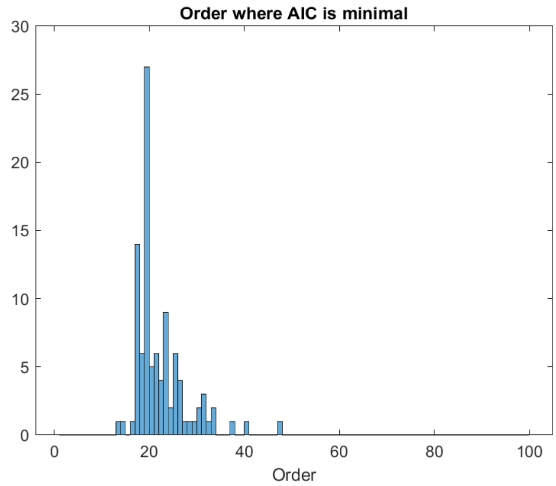
The Least Squares approach has a very narrow peak at $n = 100$, which is the maximum value. Here again the Least Squares thinks it is creating a better model, while it is in fact overfitting.



(a)



(b)



(c)

Figure 4: Histograms for all model estimations