What I did is some sort like a robust check of the main model.

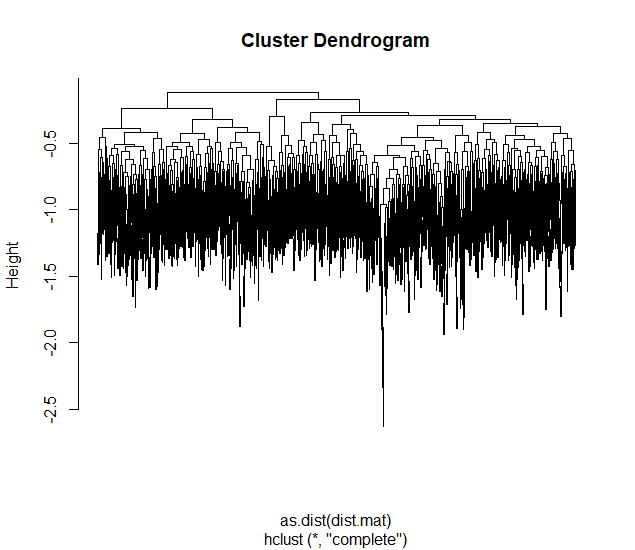
As a robustness check, my basic idea is almost the same or to say, follows the route of the main model. I first clustered the items (pid, size) combinations into different categories. Then I performed regression analysis within each cluster (multivariate regression and generalized linear model are used). I also used ordinal classification.

I used stepwise variable selection for multivariate regression.

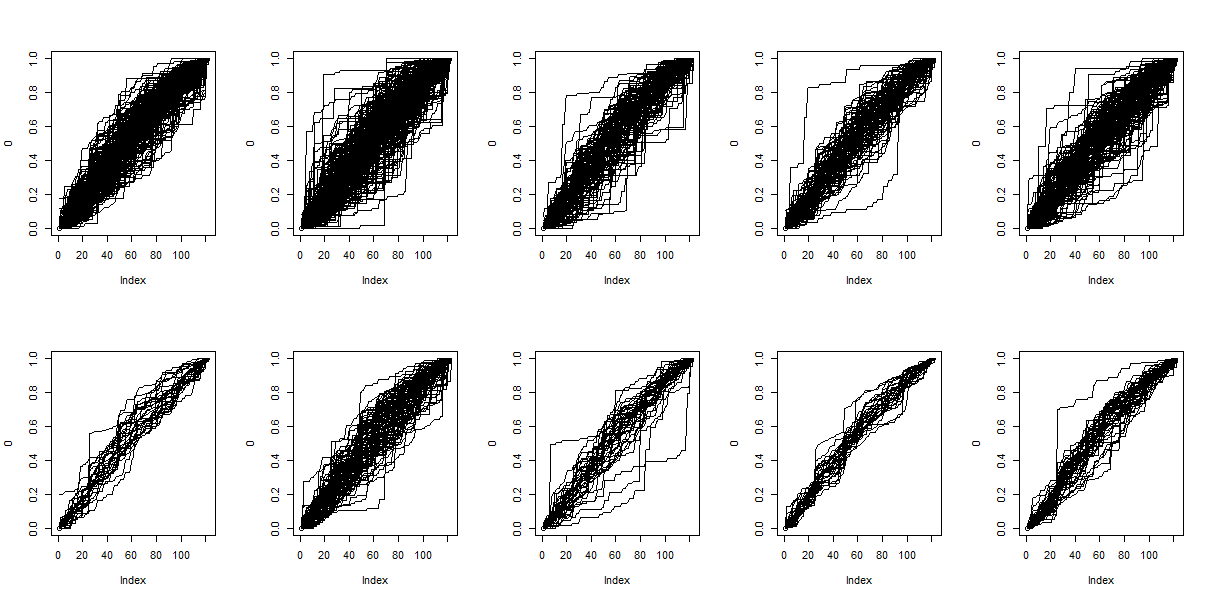
To test the goodness of the models I used, I subset the full data set by date and used the real sales record of January 2018 as the test rule.

The results are showing below (as an example, a subsample of items whose cumulative is selling numbers are more than 5 is used, the total number of subsamples used is 1166):

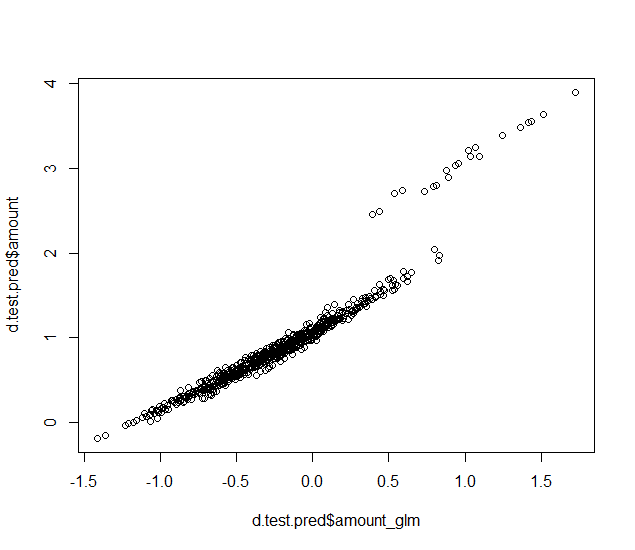
Step1: clustering based on the items whose cumulative selling number are >5



Step2: based on the clustering result, I used 10 branches and chose branch 3 as an example for the following model fitting.

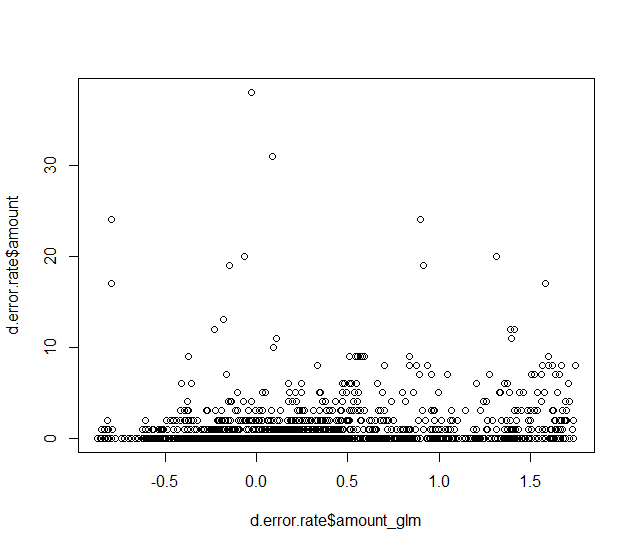
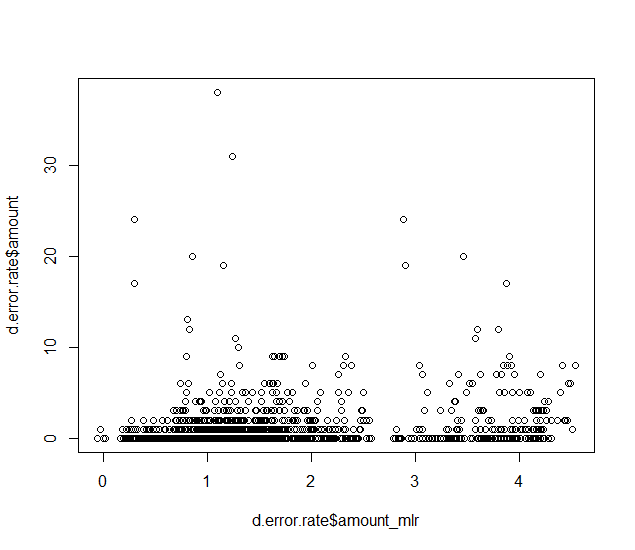


Step3: multivariate regression prediction. Multivariate regression is used for prediction of daily sales. A stepwise variable selection is preformed and 5 explanatory variables were selected. Using the same set of explanatory variables, a generalized linear model is used and the prediction for sales in February is plotted.



Ordinal clustering is also tried, but performed poorly.[[1]](#footnote-1)

Stpe4: from the above plot, the two models are consistent enough. So the data were divided into two parts for testing. The data from 2017 were used to train the model and the data in 2018 Jan were used for prediction and evaluation.



However, these models perform not well. As shown above, the left plot gives the mlr estimate vs. the actual selling records; the right plot shows the glm predict vs. the actual selling record.

The work till now suggests that other models are needed for prediction.

1. rpartScore function was used. For DMC result, I’ll try different packages. [↑](#footnote-ref-1)