CV scores for flu data

Minh Tam Hoang

7/12/2019

Time-series cross validation procedure:

In this approach, the minimum size of training set is 1526 and each training set contains one more observation than the previous one. The size of each test set relies on the forecast horizon of interest. The forecast accuracy measures which are RMSE in this approach are calculated and averaged across all test sets. Since we start with 156 observations for each series to produce a h-step forecast, the procedure works as follows:

1) For test set, select observations at time 156+i,…,156+i+h-1 (i = 1 and h is forecast horizon) and use the observations at time 1,2,3,….156 + i - 1 to estimate the model and compute the forecast values. Compute the h-step errors on forecast from time 156 to 156+i+h-1.

2) Repeat the above step for i = 2,3,4,….,252 - 156-h + 1 (252 is the total observations for each series)

3) Calculate RMSE based on the errors obtained for each test set and average them across all test sets.

I apply the procedure for h-step forecasts(h = 1,2,3,4) described above for flu data, with VARMA(and VAR) model fitted for training sets. Each VARMA model is selected and estimated using MTS package and produce h-step forecasts. Then, the errors are obtained and are used to calculate RSME.

Based on the results from table of cross-correlation matrices, VARMA(3) and VARMA(2,1) are recommended for only seasonally- differenced data.

Log transformation and power transformation are also applied to the original data to stablize the variance and seasonal diffrences is performed after that. Possible choices for VARMA model include: VAR(2), VAR(3), VAR(1,1).

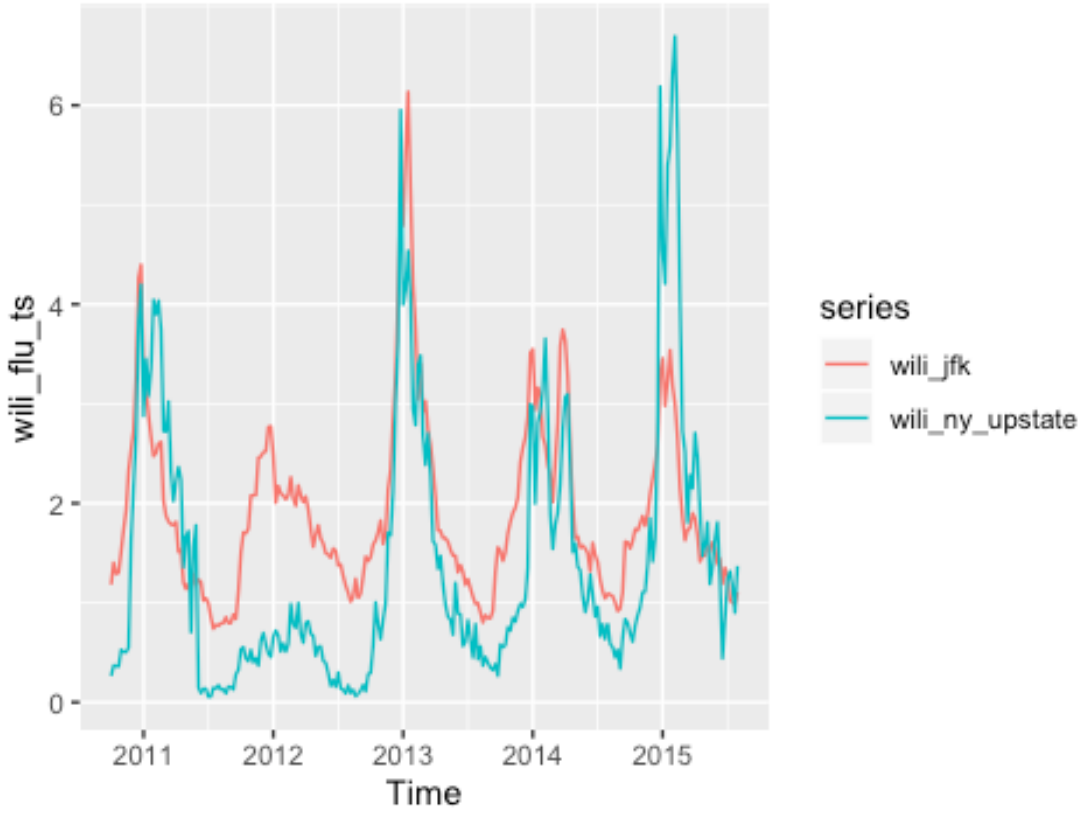
According to RMSE produced by the CV procedure, VAR(3) with log transformation gives the most suitable model for forecasting flu in JFK with the lowest CV scores in JFK. Meanwhile,VAR(3) with power transformation gives the lowest CV scores in NY upstate.

Then, I apply the procedure for four forecast horizons(h = 1,2,3,4) described above for flu data which only include wili data and repeat all necessary computations to compute RMSE.

In comparison with CV RMSE obtained from data include wiki data, CV scores for wili-jfk and wili ny upstate are slightly higher, which means that the values of forecast accuracy measures are worse than they are with wiki data. Hence, the involvement of wiki data in

the data set may be useful in predicting the behaviour of flu in these two regions in a short run.

CV-scores for data inlude wili-jfk and wili-ny-upstate only



Non-transformed data VAR(3) model for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate
## 1 0.9413326        1.107091
## 2 0.9511843        1.105222
## 3 0.9664420        1.105636
## 4 0.9727373        1.104355
```

VAR(4) model for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate
## 1 0.9080135        1.108047
## 2 0.9117012        1.106116
## 3 0.9046250        1.111780
## 4 0.9198523        1.110687
```

Log-transformed data

VAR(3) model for seasonally-differenced data

VAR(4) Log-transformed data

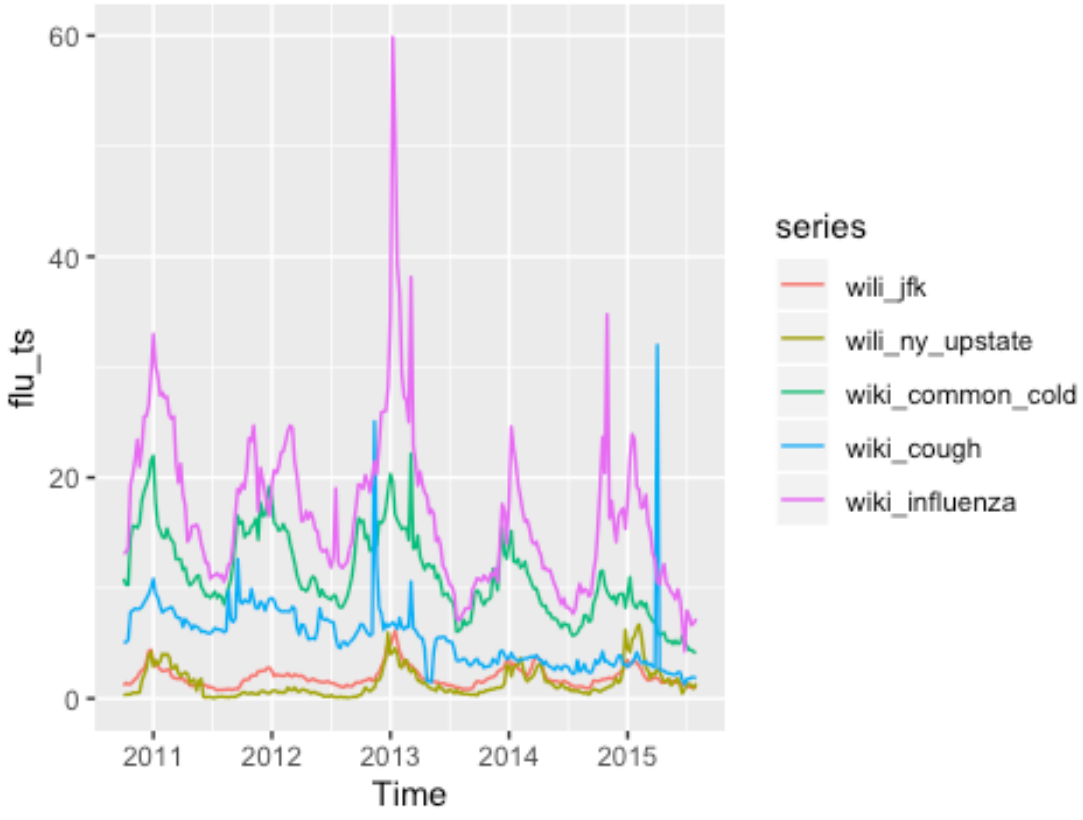
VAR(3) power-transformed data

```
##    wili_jfk wili_ny_upstate
## 1 0.9984137        1.754816
## 2 0.9279547        1.668754
## 3 0.9653866        1.585679
## 4 0.9632120        1.575508
```

VAR(2) power-transformed data

```
##    wili_jfk wili_ny_upstate
## 1 0.9765779        1.668794
## 2 1.0074822        1.274186
## 3 1.0009197        1.366790
## 4 1.0025131        1.325235
```

CV-scores for data that include both ili and wiki data



Non-transformed data VAR(3) model for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate wiki_common_cold wiki_cough wiki_influenza
## 1 0.8840797        1.152892         2.459151   4.087361       7.914744
## 2 0.9105101        1.153363         2.229822   4.104393       8.196742
## 3 1.0024295        1.109682         2.120139   4.070933       8.382637
## 4 1.0655660        1.103476         2.041605   4.059564       8.557252
```

VAR(2) model for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate wiki_common_cold wiki_cough wiki_influenza
## 1 0.9329955        1.151876         2.507113   4.092844       7.954096
## 2 1.0468202        1.147307         2.292823   4.071251       8.507181
## 3 1.1227078        1.129190         2.086752   4.066371       8.915970
## 4 1.1500383        1.120701         2.037671   4.060840       9.005727
```

Log-transformed data

VAR(3) model for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate wiki_common_cold wiki_cough wiki_influenza
## 1 0.8986496        1.182042         1.803818   3.486698       6.019834
## 2 0.7808725        1.136675         1.682593   3.608830       6.069917
## 3 0.8434721        1.170725         1.565195   3.715331       6.007187
## 4 0.8689152        1.129430         1.514099   3.712698       6.052586
```

Log-transformed data

VAR(4) for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate wiki_common_cold wiki_cough wiki_influenza
## 1 0.9078336        1.277612         1.857206   3.490553       6.041129
## 2 0.8141498        1.294224         1.913812   3.746250       6.124945
## 3 0.8808668        1.487331         1.870109   3.886401       6.008082
## 4 0.9702296        1.154199         1.605047   3.869923       6.023537
```

Power-transformed data

VAR(3) for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate wiki_common_cold wiki_cough wiki_influenza
## 1 0.9128706        1.124127         1.907656   3.671340       6.615220
## 2 0.8378080        1.131402         1.778972   3.769637       6.608382
## 3 0.8991859        1.124014         1.660729   3.852198       6.685216
## 4 0.9344975        1.126901         1.619715   3.832551       6.838518
```

VAR(2) for seasonally-differenced data

```
##    wili_jfk wili_ny_upstate wiki_common_cold wiki_cough wiki_influenza
## 1 0.9037848        1.129835         1.995033   3.686904       6.606653
## 2 0.8840313        1.183470         1.824293   3.786966       6.621071
## 3 0.8828999        1.188968         1.669211   3.858346       6.692752
## 4 0.8918530        1.214365         1.619885   3.894634       6.781241
```