

Report0325

Linyi Guo

2020/3/25

Hi Professor,

This report is mainly composed by the content I sent you last time with more figures and some hypothesis tests.

Part 1: Simulated data

Review

The simulated data we used in the following analysis contains 1000 datasets at length 180(15 years) simulated from the same state space model as following,

$$Y_t = T_t + S_t + I_t \quad \text{where } I_t \sim N(0, 10) \quad (1)$$

$$T_t = T_{t-1} + \epsilon_t \quad \text{where } \epsilon_t \sim N(0, 5) \quad (2)$$

$$S_t = - \sum_{j=1}^{11} S_{t-j} + \eta_t \quad \text{where } \eta_t \sim N(0, 1) \quad (3)$$

Then I used the first 14 years to build various models and compare their **decompositions** and the last year is saved for **forecasting**.

As a contrast, we shall still use *half-Cauchy*(with its own parameter equals to $\sqrt{10}$ and $\sqrt{5}$ separately) as another prior on standard error σ_I and σ_T instead of σ^2 this time.

Decomposition

Based on the decomposition from X-11 plus we mainly care about the trend and seasonal components, the error we defined is

$$Error = \sum (Trend_{X-11} - Trend_{Object})^2 + \sum (Seasonal_{X-11} - Seasonal_{Object})^2$$

The table below is the **summary** result of three models' error(they are all SSMS with different parameter estimations)

##	MLE	MAP	MAP_halfcauchy
##	Min. : 172.5	Min. : 176.7	Min. : 171.1
##	1st Qu.: 341.9	1st Qu.: 316.6	1st Qu.: 332.1
##	Median : 405.8	Median : 373.4	Median : 389.4

```
## Mean    : 422.6    Mean    : 390.5    Mean    :405.8
## 3rd Qu.: 487.0    3rd Qu.: 444.3    3rd Qu.:464.7
## Max.    :1012.3    Max.    :1029.6    Max.    :960.9
```

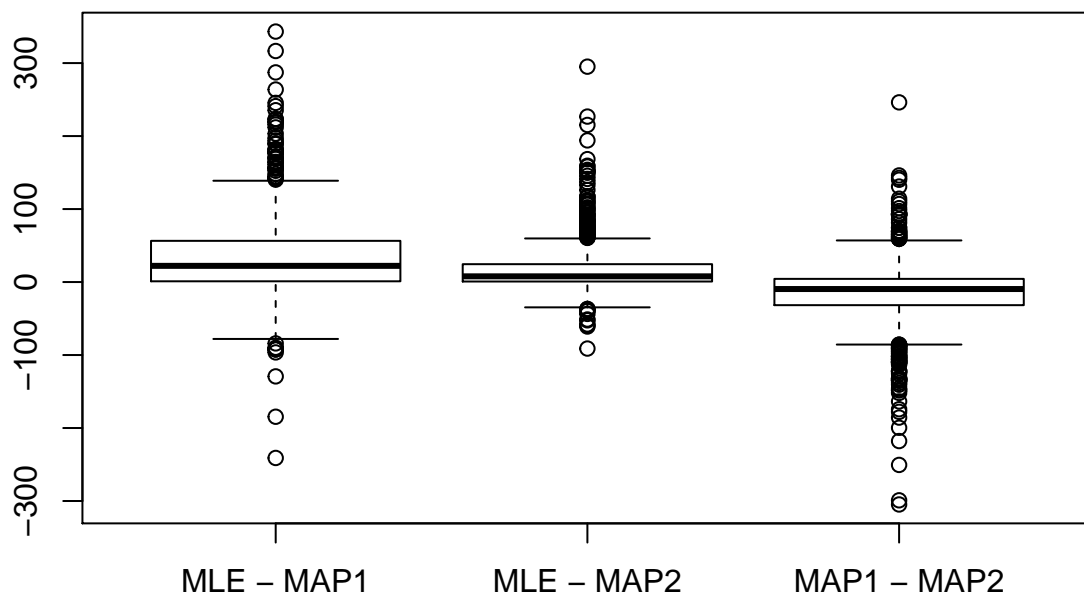
The standard deviation and standard error(which is the standard deviation adjusted by the group size 1000) are

```
##          MLE          MAP MAP_halfcauchy
##      112.3569      110.3803      105.0638
```

```
##          MLE          MAP MAP_halfcauchy
##      3.553037      3.490531      3.322409
```

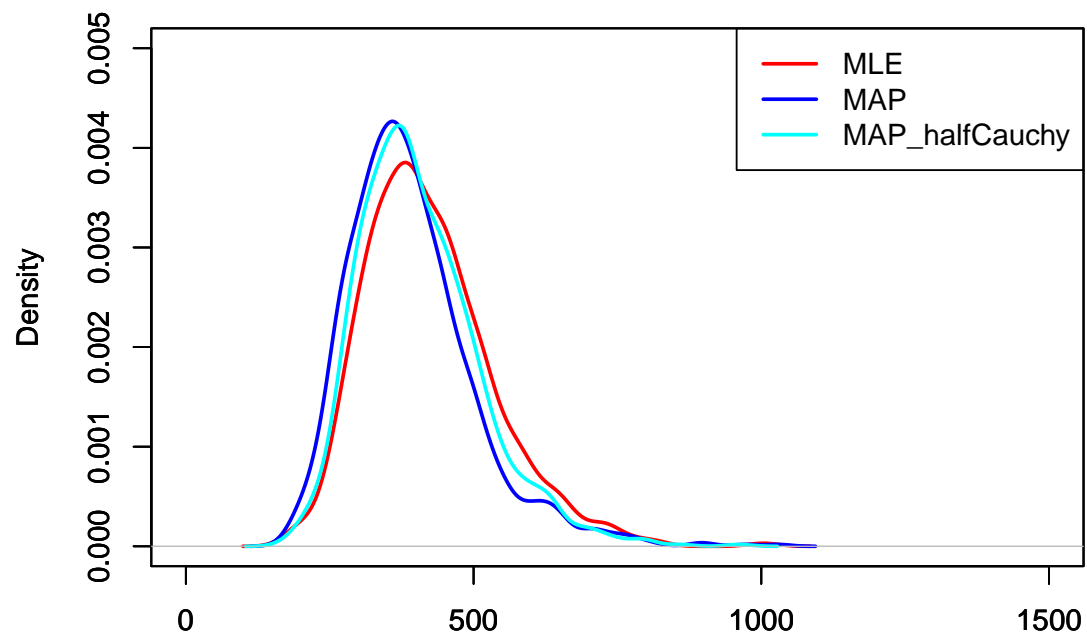
These numbers above give us an intuition that the MAP/MAP_halfCauchy behaves better than MLE, and compared with the empirical prior, the half-cauchy seems to converge too much(cause the parameter I set up is *too small*), but for the mean error, empirical prior did a better job.

This is the boxplot of their differences:(*MAP1* is the estimator from our informative prior; *MAP2* is from half-Cauchy)

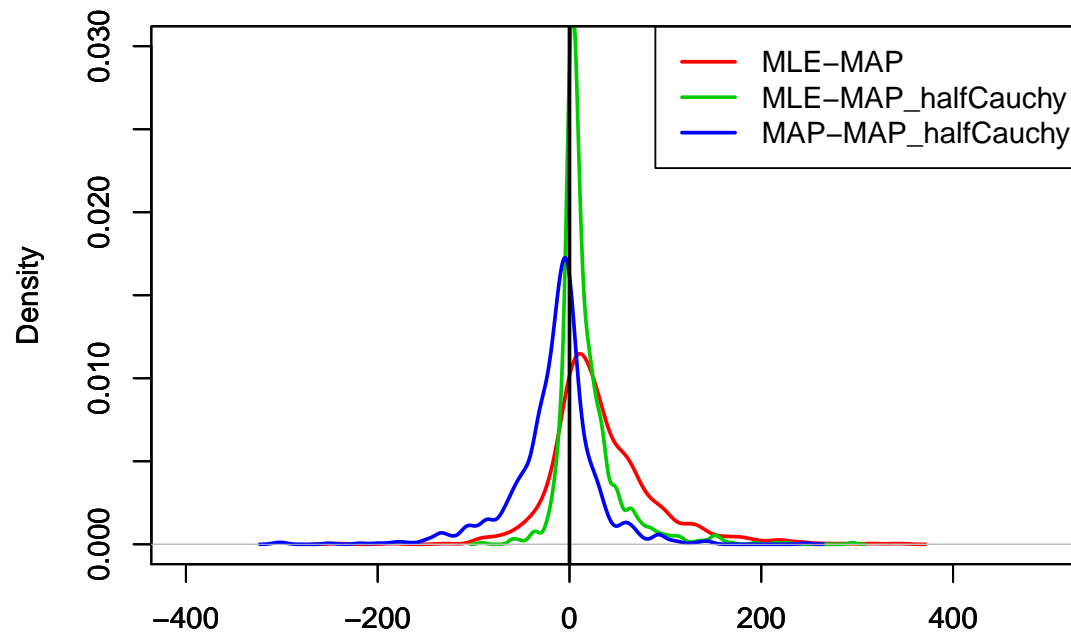


And the distributions of their error and errors' difference:

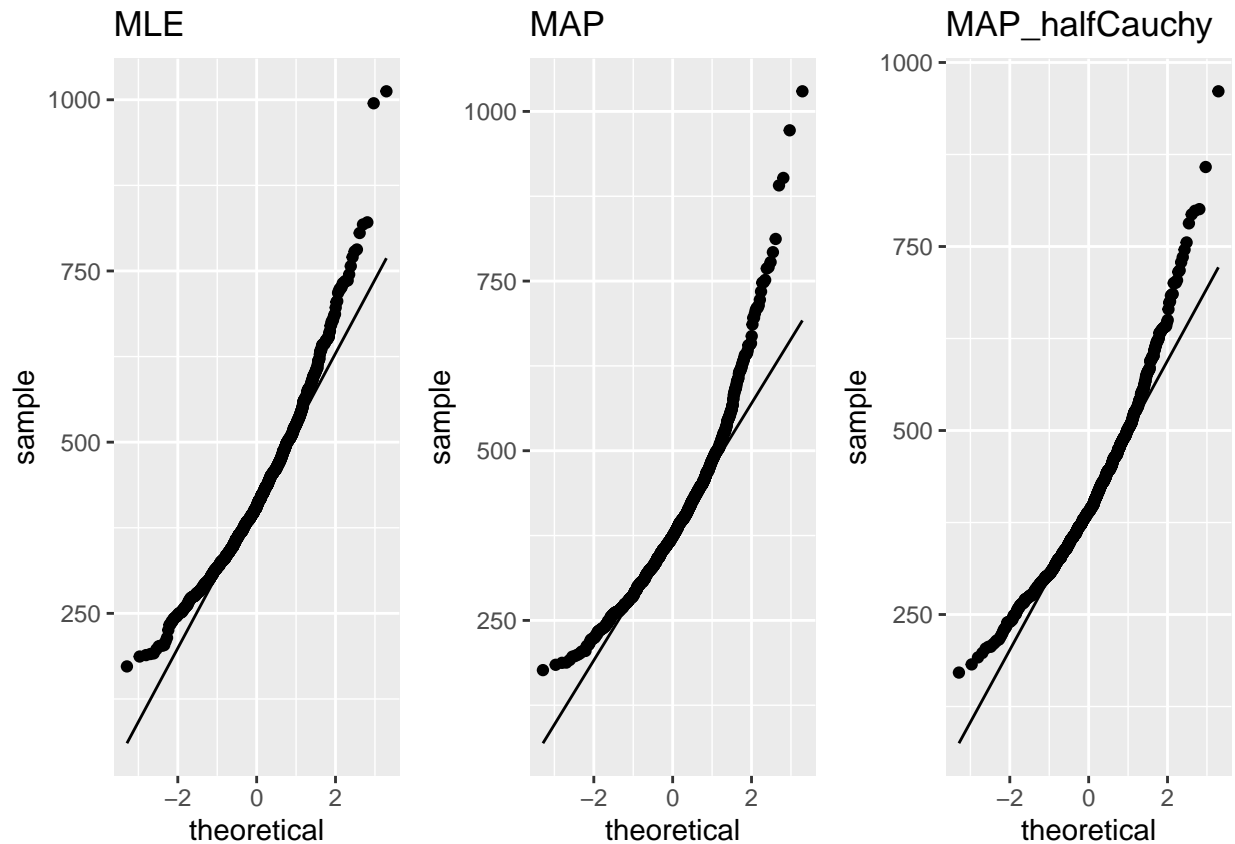
Decomposition error comparison(X11 is standard)



Distribution of differences among decomposition errors



I plotted the *normal qq plot* of their decomposition errors and it seems that they are not normally distributed, so after some explorations, I found the **Friedman test** and **Mann–Whitney U test** fit our problem well. And the results showed that the difference between each method is prominent. Therefore, we have reasons to believe these three methods did differ from each other in this part.



Results from Friedman test:(first one is to test three methods together, and the rest three are for every two of them)

```
##
## Friedman rank sum test
##
## data: Error and Method and No
## Friedman chi-squared = 520.86, df = 2, p-value < 2.2e-16

##
## Friedman rank sum test
##
## data: as.matrix(error_decomp[, c(1, 4)])
## Friedman chi-squared = 280.9, df = 1, p-value < 2.2e-16

##
## Friedman rank sum test
##
## data: as.matrix(error_decomp[, c(1, 6)])
## Friedman chi-squared = 338.72, df = 1, p-value < 2.2e-16
```

```
##
## Friedman rank sum test
##
## data:  as.matrix(error_decomp[, c(4, 6)])
## Friedman chi-squared = 133.96, df = 1, p-value < 2.2e-16
```

Results from Mann–Whitney U test:(for each two of them)

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  error_decomp$MLE and error_decomp$MAP
## W = 592348, p-value = 8.589e-13
## alternative hypothesis: true location shift is not equal to 0
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  error_decomp$MLE and error_decomp$MAP_halfcauchy
## W = 544516, p-value = 0.0005662
## alternative hypothesis: true location shift is not equal to 0
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  error_decomp$MAP and error_decomp$MAP_halfcauchy
## W = 449905, p-value = 0.0001047
## alternative hypothesis: true location shift is not equal to 0
```

Prediction

With the models obtained before, we could predict each series for the next year and then we compared our prediction with the true values.

Similar to the previous section, we shall have the following results:

MEAN

```
##      x11      MLE      MAP      map2
## 735.4285 619.9037 621.9331 618.5734
```

Standard deviation and standard error

```
##      x11      MLE      MAP      map2
## 775.4288 571.0056 581.0846 570.7436
```

```
##      x11      MLE      MAP      map2
## 24.52121 18.05678 18.37551 18.04850
```

As we see, differences here among SSMs are not obvious, but the behaviour of X11 is poor.

And when I applied the testing methods, *Friedman test* supported our argument above, which is the prediction error from x-11 is different from those from SSMs but SSMs have a similar prediction error distribution compared with each other. However, *Mann-Whitney U test* showed that all prediction errors' difference of these methods are not prominent, although the p-values of tests between x11 and other SSMs are not very high(0.13-0.16).

Friedman test

1,2,3,7 stand for x11, MLE, MAP, MAP_halfCauchy separately.

```
##
## Friedman rank sum test
##
## data:  as.matrix(error_pre[, c(1, 2, 3, 7)])
## Friedman chi-squared = 20.628, df = 3, p-value = 0.0001258

##
## Friedman rank sum test
##
## data:  as.matrix(error_pre[, c(1, 2)])
## Friedman chi-squared = 9.72, df = 1, p-value = 0.001823

##
## Friedman rank sum test
##
## data:  as.matrix(error_pre[, c(1, 3)])
## Friedman chi-squared = 12.813, df = 1, p-value = 0.0003442

##
## Friedman rank sum test
##
## data:  as.matrix(error_pre[, c(1, 7)])
## Friedman chi-squared = 11.213, df = 1, p-value = 0.0008121

##
## Friedman rank sum test
##
## data:  as.matrix(error_pre[, c(2, 3, 7)])
## Friedman chi-squared = 0.60667, df = 2, p-value = 0.7384
```

Mann-Whitney U test

```
##
## Wilcoxon rank sum test with continuity correction
##
## data:  error_pre$x11 and error_pre$MLE
## W = 48036, p-value = 0.1528
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
```

```
##
## data: error_pre$x11 and error_pre$MAP
## W = 48136, p-value = 0.1397
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
##
## data: error_pre$x11 and error_pre$map2
## W = 48159, p-value = 0.1368
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
##
## data: error_pre$MLE and error_pre$MAP
## W = 45108, p-value = 0.9596
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
##
## data: error_pre$MLE and error_pre$map2
## W = 45093, p-value = 0.9652
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test with continuity correction
##
## data: error_pre$MAP and error_pre$map2
## W = 44966, p-value = 0.9874
## alternative hypothesis: true location shift is not equal to 0
```

Real data

Here I extracted 27 datasets from Statcan for our analysis. For each dataset, we left the last year for prediction and used the previous to build models and do decompositions. But I haven't applied any empirical prior here, the only prior distribution is the half-Cauchy defined above with the same parameter values.

Note: We removed the outliers and calendar effects at first. And X-11 showed that the relations among components for these data are all multiplicative so to use SSM I had to take log in all models, which would makes these errors very tiny.

Decomposition

MEAN

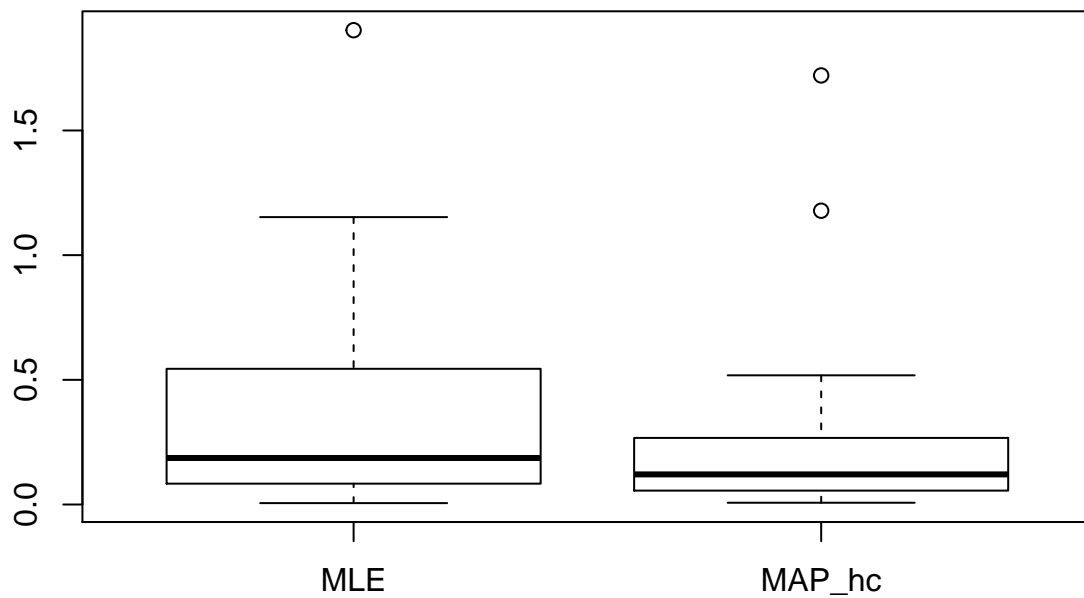
```
##          MLE      MAP_hc
## 0.3606837 0.2614593
```

Standard deviation and standard error

```
##      MLE      MAP_hc
## 0.4228153 0.3812444
```

```
##      MLE      MAP_hc
## 0.08137083 0.07337052
```

Boxplot



Test

Friedman test showed the difference of their decomposition error is significant but Mann-Whitney U test did not.

```
##
## Friedman rank sum test
##
## data: as.matrix(realerror_decomp)
## Friedman chi-squared = 4.4815, df = 1, p-value = 0.03426

##
## Wilcoxon rank sum test
##
## data: realerror_decomp$MLE and realerror_decomp$MAP_hc
## W = 430, p-value = 0.2632
## alternative hypothesis: true location shift is not equal to 0
```


Prediction

MEAN

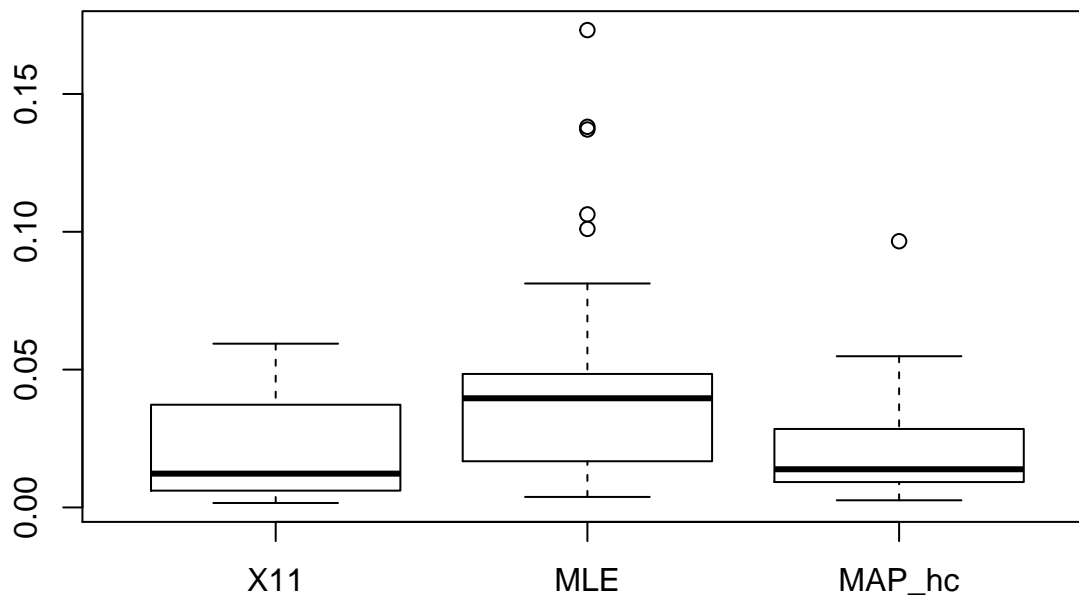
```
##          X11          MLE          MAP_hc
## 0.02118249 0.04732707 0.02065424
```

Standard deviation and standard error

```
##          X11          MLE          MAP_hc
## 0.01997234 0.04570643 0.01978781
```

```
##          X11          MLE          MAP_hc
## 0.003843679 0.008796206 0.003808167
```

Boxplot



Test

1,2,3 stand for X11, MLE and MAP_halfCauchy.

```
##
## Friedman rank sum test
##
## data:  as.matrix(realerror_pre)
## Friedman chi-squared = 18.074, df = 2, p-value = 0.0001189
```

```

##
## Friedman rank sum test
##
## data: as.matrix(realerror_pre[, c(1, 3)])
## Friedman chi-squared = 0.33333, df = 1, p-value = 0.5637

##
## Friedman rank sum test
##
## data: as.matrix(realerror_pre[, c(1, 2)])
## Friedman chi-squared = 10.704, df = 1, p-value = 0.001069

##
## Friedman rank sum test
##
## data: as.matrix(realerror_pre[, c(2, 3)])
## Friedman chi-squared = 13.37, df = 1, p-value = 0.0002556

##
## Wilcoxon rank sum test
##
## data: realerror_pre$X11 and realerror_pre$MLE
## W = 221, p-value = 0.01252
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test
##
## data: realerror_pre$X11 and realerror_pre$MAP_hc
## W = 356, p-value = 0.8908
## alternative hypothesis: true location shift is not equal to 0

##
## Wilcoxon rank sum test
##
## data: realerror_pre$MLE and realerror_pre$MAP_hc
## W = 515, p-value = 0.008684
## alternative hypothesis: true location shift is not equal to 0

```

Summary: although we didn't apply empirical prior, the weakly-informative prior(maybe not very weak) also did a better job compared with MLE. The interesting thing is compared with the prediction of simulated data, X-11 had a better performance in these real data. In conclusion, the result after considering prior information is better/equivalent to those from X-11 or MLEs.