

Meeting20200302_StatCan

Lin

2020/2/29

Contents

| | |
|------------------------------------------|----|
| Review | 2 |
| State space modelling with MLEs | 2 |
| State space modelling with LOSS function | 4 |
| Bayesian analysis structure building | 6 |
| Results under Bayes analysis | 6 |
| Improvements and Questions | 13 |

Review

When we met last time, I said *the seasonal adjusted series is smoother than those from X-11 and SEATS*, but the problem is: *I fixed the $\sigma_I = \sigma_T = \sigma_S = 1$ at that time*. Apparently it is unreasonable to pick up some numbers and assign them to these parameters manually.

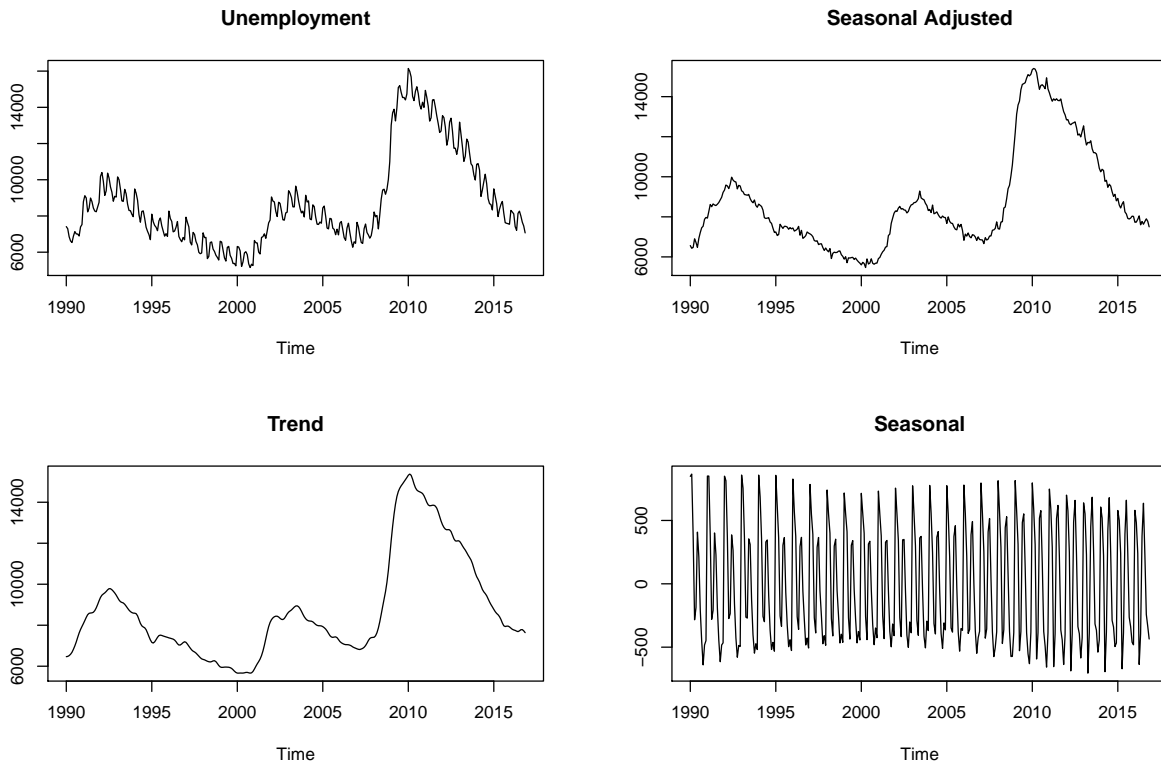
Thus, we have a parameter estimation problem here, which is, to some extent, aligned with the first point I mentioned in the **Improvement** part of our last meeting. (My original sentence was *We have mentioned before: the seasonal component from X-13ARIMA-SEATS should be invariant within each year, why are our results changeable? And given they are changeable, how to control them?*, but now we will focus on how to control *SSM*.) Mathematically, we need to estimate three variances in the following state space model:

$$Y_t = T_t + S_t + I_t \quad \text{where } I_t \sim N(0, \sigma_I^2) \quad (1)$$

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

$$S_t = - \sum_{j=1}^{11} S_{t-j} + \eta_t \quad \text{where } \eta_t \sim N(0, \sigma_S^2) \quad (3)$$

Let's take a look at the decomposition from **X-11** of the unemployment of US between 1990 and 2015 at first:

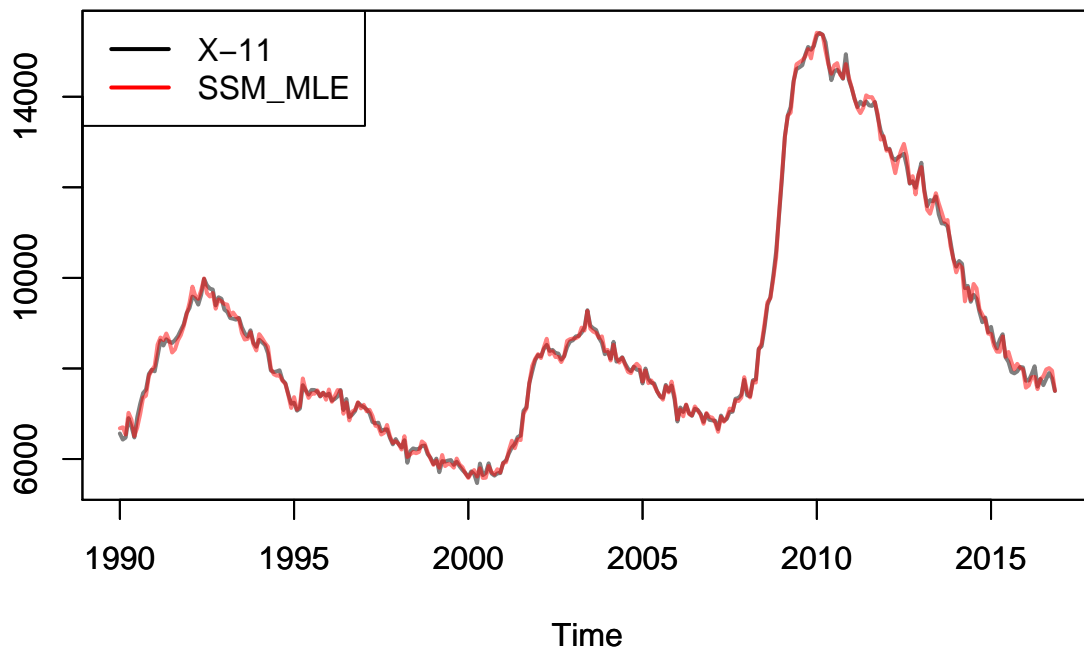


State space modelling with MLEs

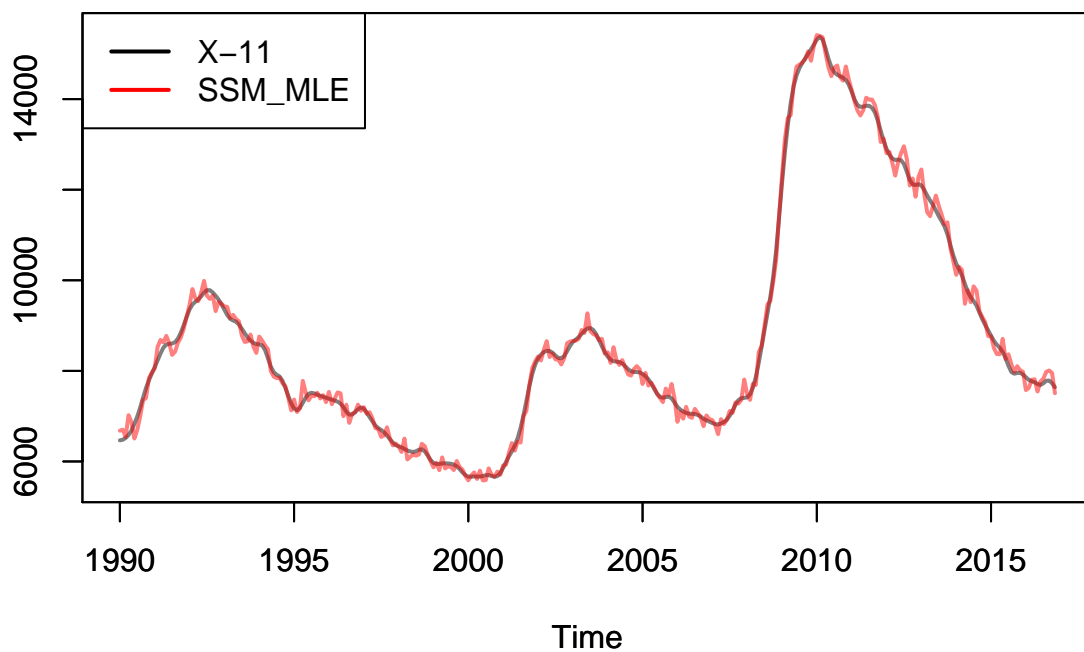
Naturally, we should try the **MLE** as the estimator of our parameter at first. And luckily, R package **KFAS** could help us to find the MLE of these variances. (I will **SKIP** the technical details here at first.) Then we can apply the MLE to the **State space model** and obtain the decompositions.

The comparison of seasonal adjusted series and trend series from **X-11** and **SSM_MLE** are as following:

Seasonal adjusted



Trend



As we can see here, the results generating from the model **SSM_MLE** is far away from those from **X-11**. Therefore, instead of using MLEs, we can define a *loss function* based on our **NEED** and search the *good* values of parameters.

State space modelling with LOSS function

In this section, our loss function is defined as

$$Loss = \sum (Seasonal_{X-11} - Seasonal_{SSM})^2 + \sum (Trend_{X-11} - Trend_{SSM})^2 + \sum (D(Trend_{X-11}) - D(Trend_{SSM}))^2$$

where $D(Trend) = Trend_{2:T} - Trend_{1:(T-1)}$.

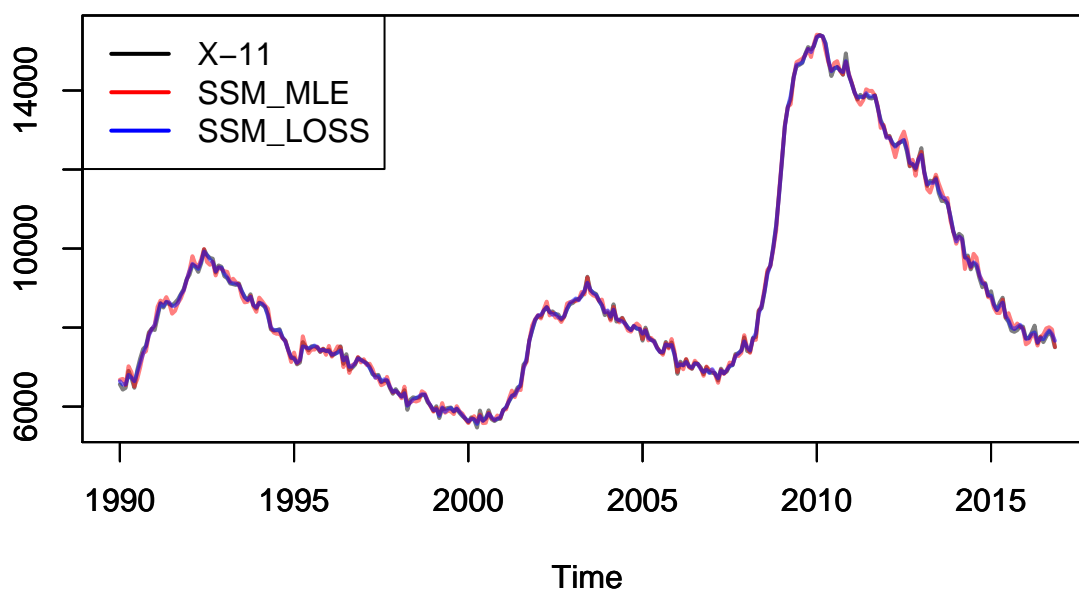
EXPLAIN

- the *first sum* is to punish the difference of *seasonal series*, which could be viewed to punish the *seasonal adjusted series* as well;
- the *second* one is to punish the difference of *trend series*;
- the *last sum* can be regared to punish the '*derivative*' of our trend series.

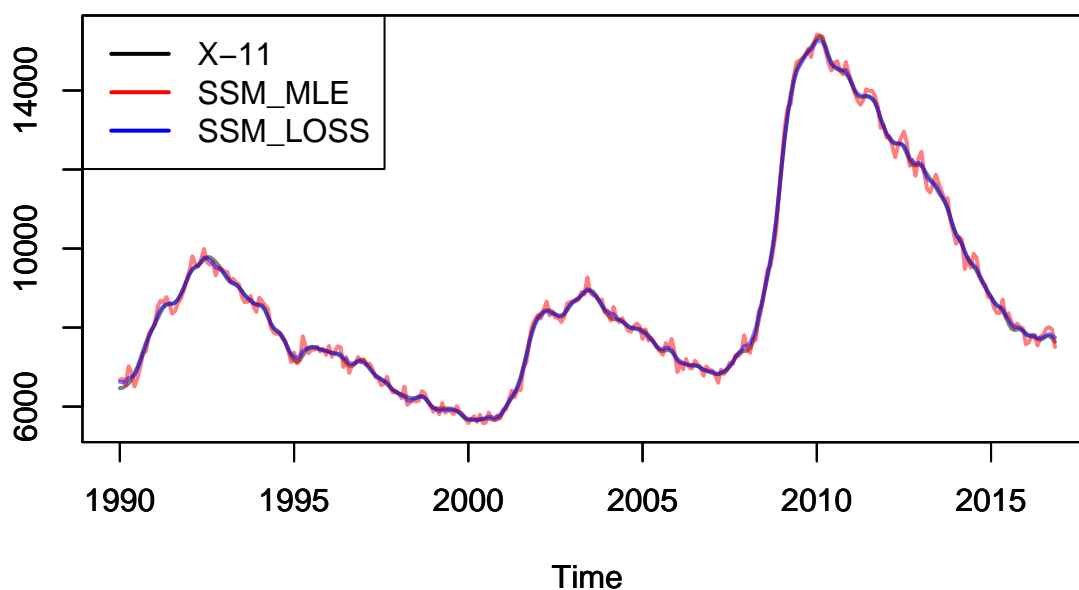
For the moment I used the *grid search* to find the *good* values(*TO BE IMPROVED*).

The following figures are the comparisons of results from *X-11*, *SSM_MLE*, *SSM_LOSS*:

Seasonal adjusted



Trend



We can see that the result from the new model **SSM_LOSS** has a better fit with that from **X-11**, which means our loss function does a better job than MLE.

BUT We need to notice one thing: although the **SSM_LOSS** model behaves good, if we want to use it, we still need to *rely* on X-11/SEATS to get the *loss* and the computation is quite heavy based on grid search.

Bayesian analysis structure building

To avoid both problems and make use of our loss function at the same time, we could try the following steps, which is based on the empirical Bayesian:

1. Simulate numerous datasets from the **SAME** model;
2. Apply *LOSS function* and *grid search* to 70% simulated datasets, then record the *good* estimators for each dataset;
3. Compute the distributions of *good* estimators from *Step 2* w.r.t trend, seasonal and noise variances separately, and treat them as our *prior* distributions;
4. For the rest 30% datasets, instead of using *LOSS* function to search the ideal value, based on **Bayesian analysis**, under the prior information we gained above, we use the **MAP** values as parameter estimators;
5. **Repeat** Step2 on the rest 30% datasets, and also compute their **MLEs**;
6. Compare three different estimators of our variances.

Results under Bayes analysis

REMARK: based on Kalman filter, we can prove: for the same structured time series dataset, as long as the ratio of three variances does NOT change, the decompositions will NOT change. Because of this fact, I **fixed** the variance of seasonal series at 1 so only need to care about the values of the other two variances.

According to steps above:

I. I first simulated 1000 datasets from the **SAME** state space model as following:

$$Y_t = T_t + S_t + I_t \quad \text{where } I_t \sim N(0, 10) \quad (4)$$

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

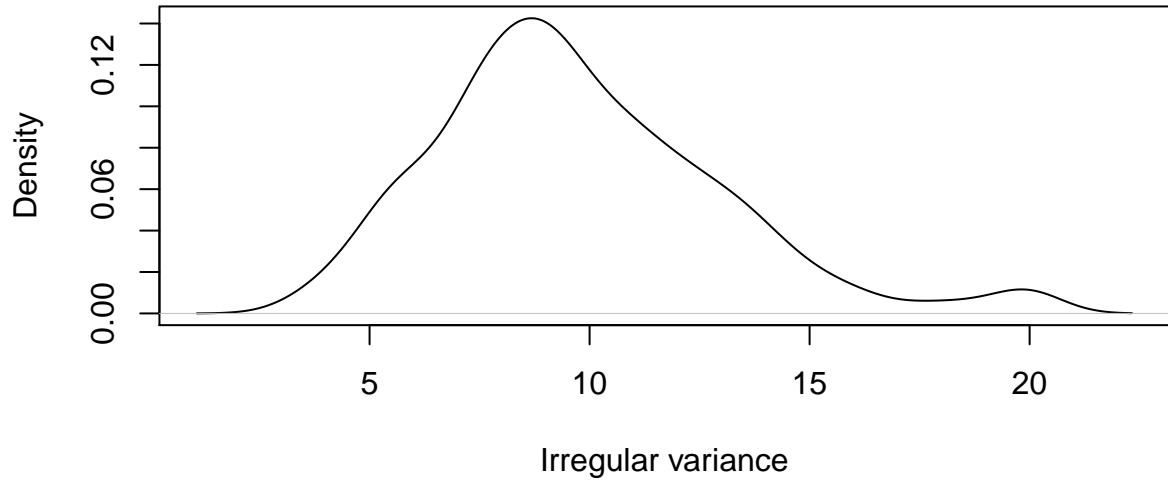
$$S_t = - \sum_{j=1}^{11} S_{t-j} + \eta_t \quad \text{where } \eta_t \sim N(0, 1) \quad (6)$$

II. I used the following loss function(*the last term is different but also to control the derivative of trend series*) to build our prior distributions:

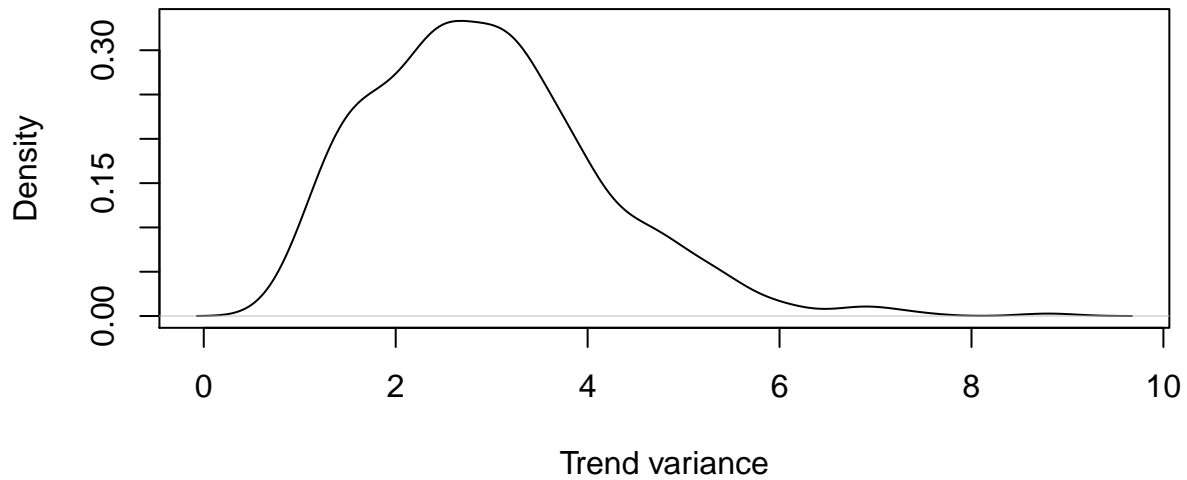
$$Loss = \sum (Seasonal_{X-11} - Seasonal_{SSM})^2 + \sum (Trend_{X-11} - Trend_{SSM})^2 + [\sum D(Trend_{X-11}) - \sum D(Trend_{SSM})]^2$$

III. The distribution of the *good* estimators of the other two variances is as following:(based on the 101:600 datasets)

Prior distribution of irregular variance estimators



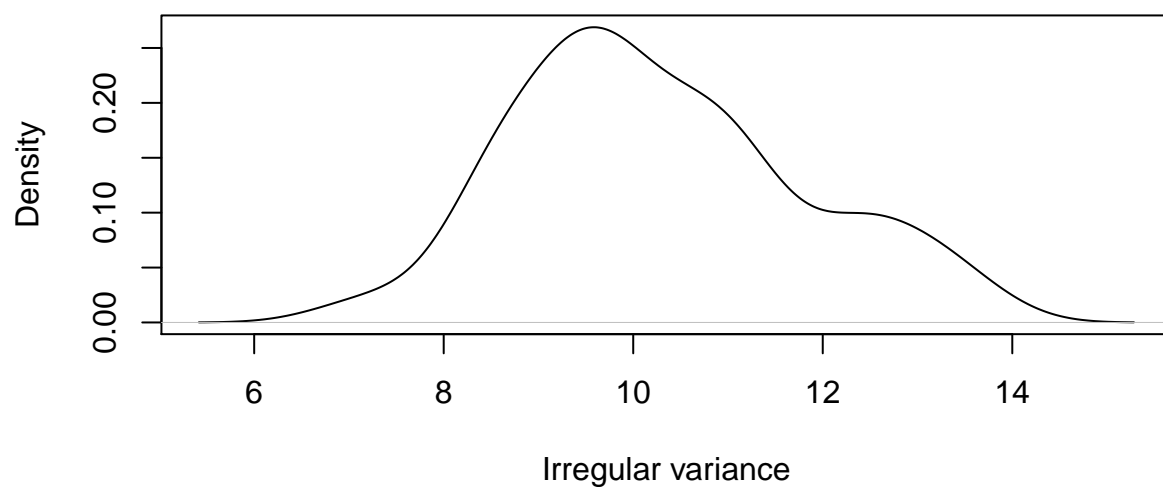
Prior distribution of trend variance estimators



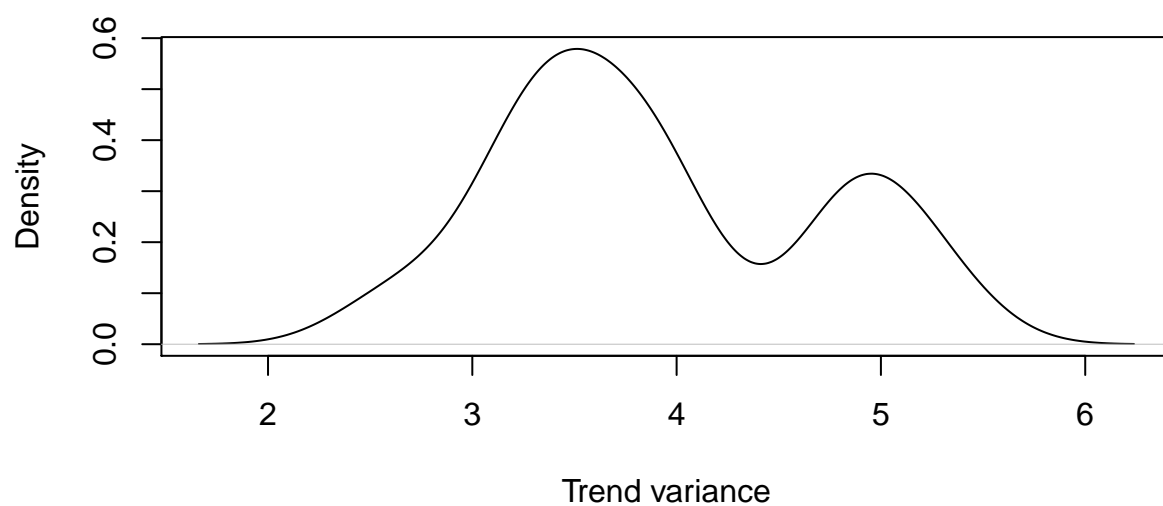
NOTE: The reason we have a little bump at 20 in the first figure is that our *upper limit* of irregular variance in grid search is 20, in another word, some *good* estimators of irregular variance could be greater than 20. Since the computation cost a lot of time, we will use this prior at first. It is not hard to fix it :)

IV. At the moment, I only used the first 100 datasets to test my method, and the MAPs' distribution are as following:

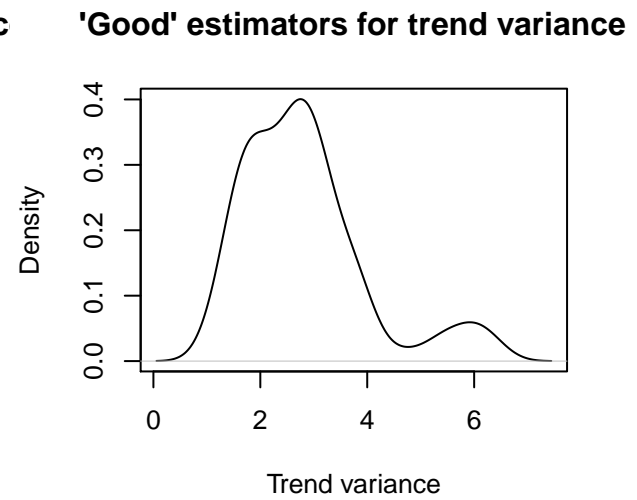
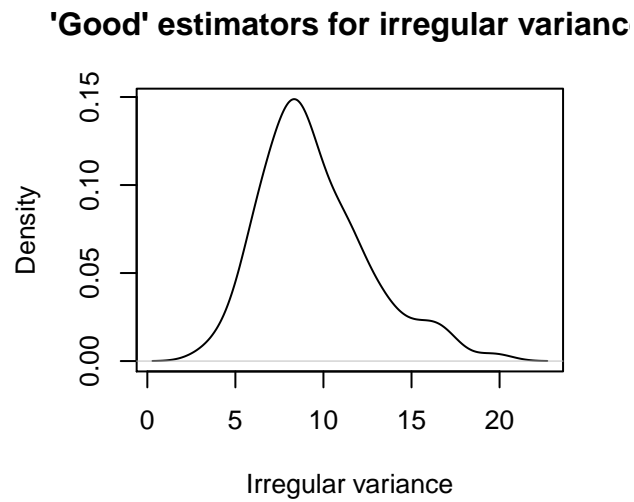
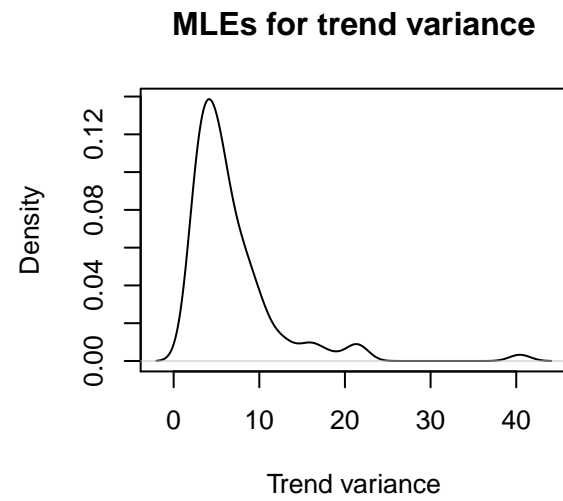
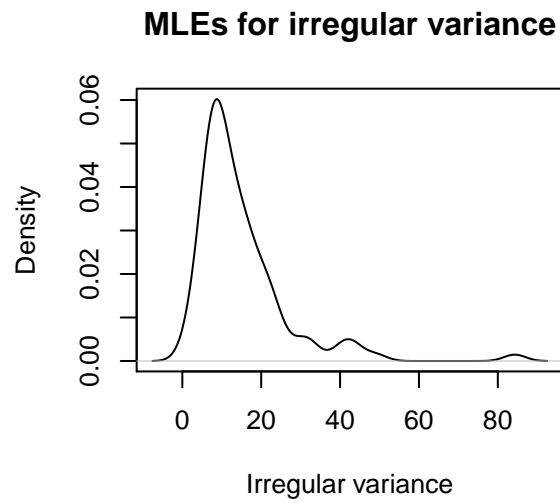
Distribution of MAP estimators for irregular variance



Distribution of MAP estimators for trend variance



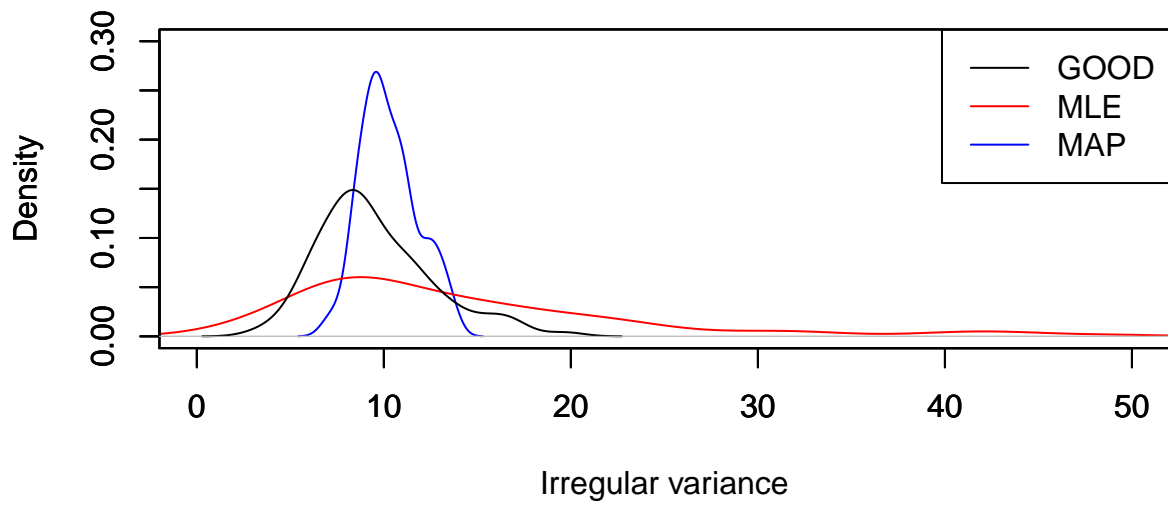
V. The MLEs and the **ideal** estimators of these 100 datasets are distributed as following:



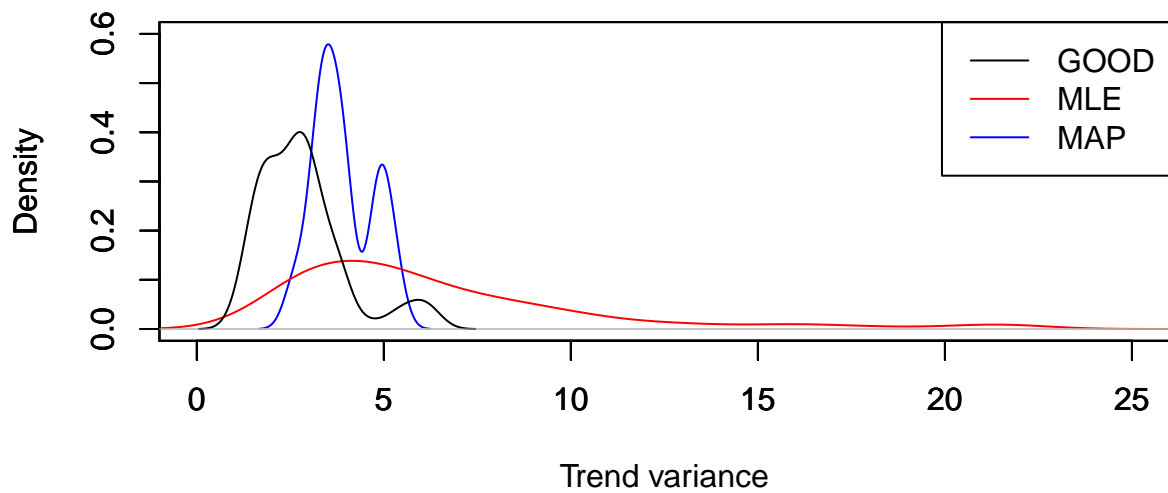
VI. To compare these three models, I used two graphs here:

- the **density** distributions;
- their **scatterplots**.

Comparison of different estimators for irregular variance



Comparison of different estimators for trend variance



The parallel coordinate scatterplots are:

Figure 1: Comparison of irregular variance estimators
Parallel Coordinate Plot for the variance.1

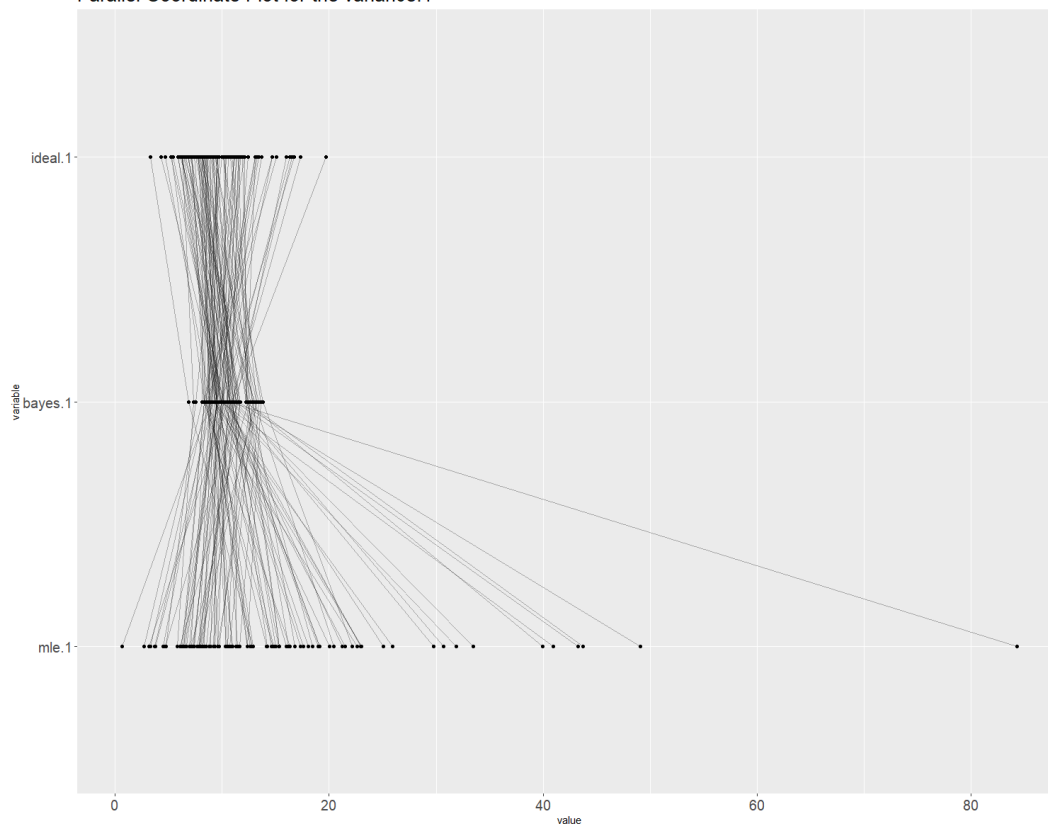
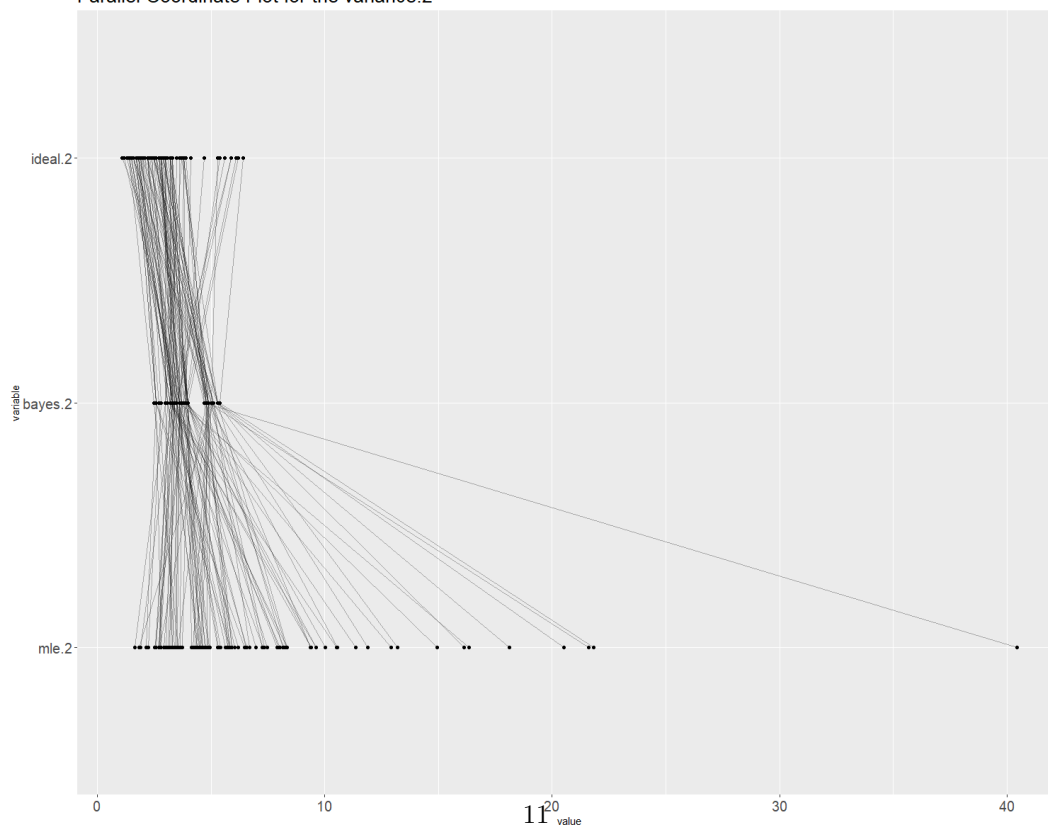


Figure 2: Comparison of trend variance estimators
Parallel Coordinate Plot for the variance.2



Overall, we can tell from the above two figures that our informative prior(bayes) **DOSE WORK** compared with MLE, since the **bayes** are closer to ideal values.

To **verify** the effect of our prior, I used the **uniform** distribution as another prior at the same time *as a contrast*.

Note: the uniform prior for irregular variance is defined on $[0,20]$ and that of trend variance is $[0,10]$. So both uniform priors are also informative to some extent.

The results are as following:

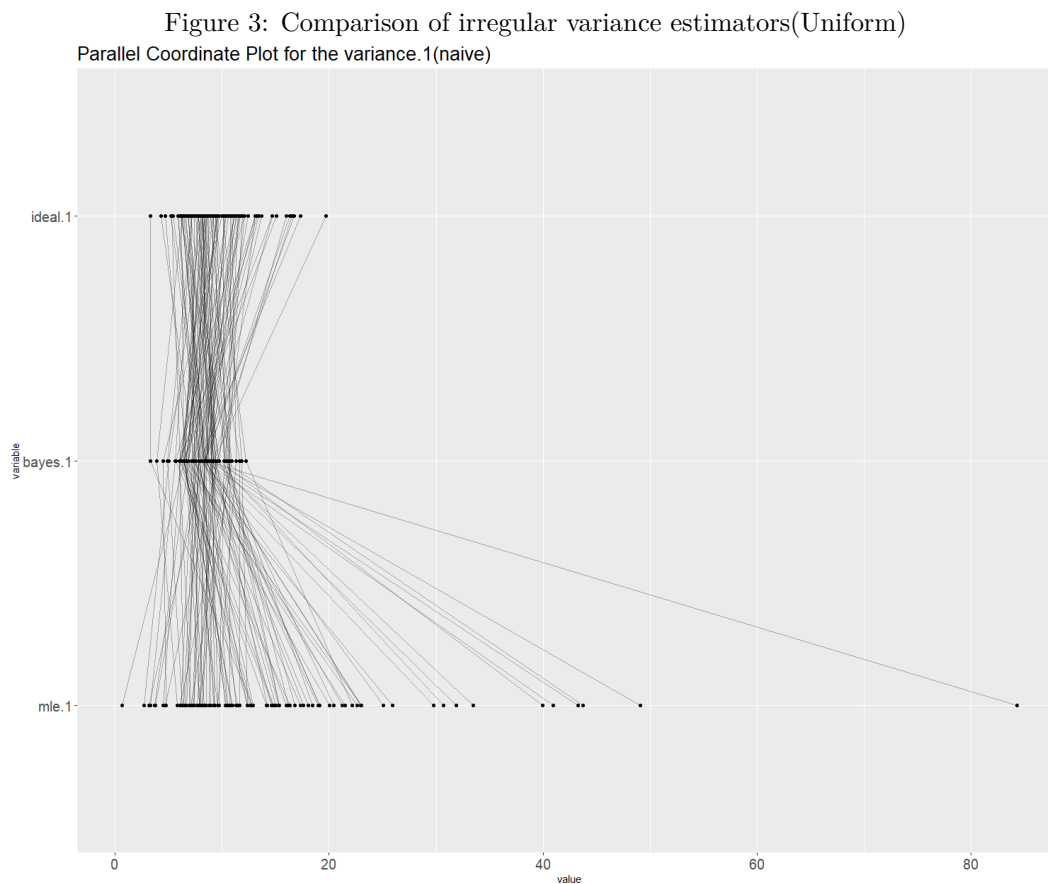
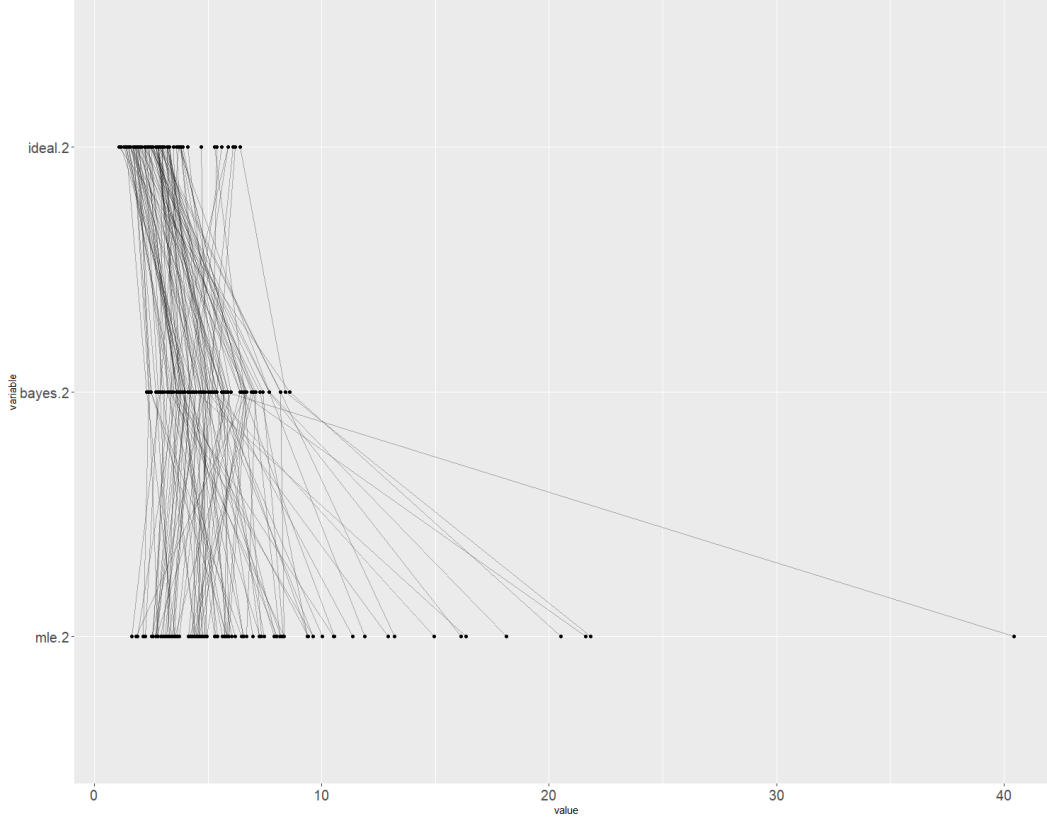


Figure 4: Comparison of trend variance estimators(Uniform)
Parallel Coordinate Plot for the variance.2(naive)



We can see that the convergence under *uniform* prior is not very quick compared with the previous one. To have a better understand of the difference among *MLE*, *SSM_Bayes*, *SSM_Bayes(uniform)*, I define

$$d = ||Var1_{ideal} - Var1_{model}||_1 + ||Var2_{ideal} - Var2_{model}||_1$$

that is

$$d = |Var1_{ideal} - Var1_{model}| + |Var2_{ideal} - Var2_{model}|$$

where Var_{ideal} means the result from our loss function and $Var1, Var2$ stand for irregular variance and trend variance respectively. The following table gives the number of datasets satisfying the corresponding condition:(*Reminder: we have 100 datasets for testing in total*)

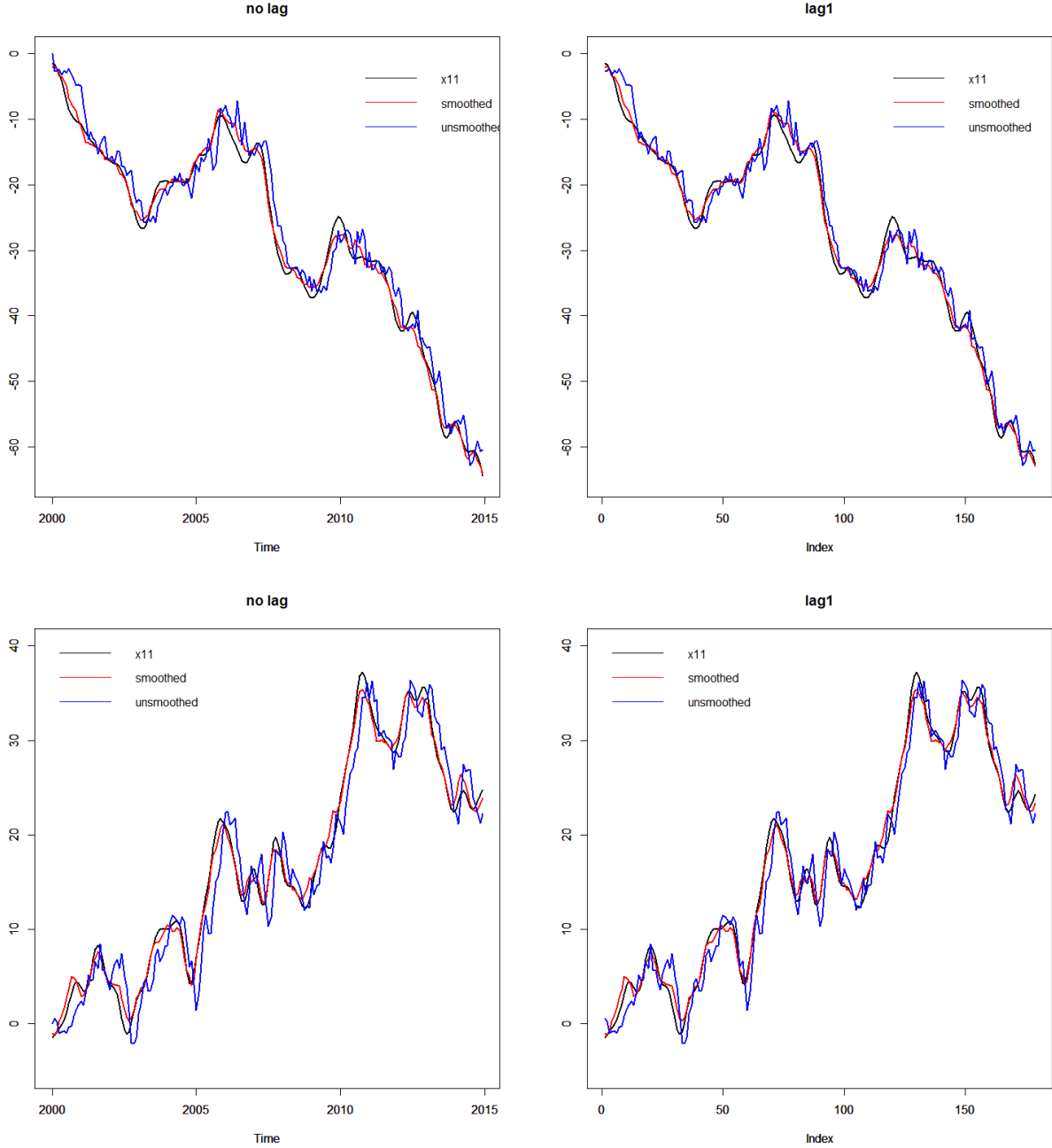
Table 1: Comparison of estimate error among different methods

| Method | $d \leq 1$ | $d \leq 2$ | $d \leq 3$ |
|---------------|------------|------------|------------|
| MLE | 2 | 9 | 17 |
| Uniform Prior | 2 | 9 | 25 |
| Our Prior | 5 | 21 | 35 |

Improvements and Questions

- In Kalman filter, the filtering process is fixed, but for smoothing process, whether there is any **better smoother**?

- As we can see in the following **TREND** figures, it seems that at each bump, our SSM will **underestimate** a little bit. If we can't figure out the smoother problem, we could add one more step to fix this difference after smoothing process. It is worthy to explore this potential step here.



- Although the results after considering the Bayesian analysis are better than MLEs **as a whole**, I still need to admit the effect is not very good, especially when we look at the **single** dataset, just like the figures above. I guess this problem would be solved if we figure out the second one.
- Computing a *good* estimator of variance is time-consuming, and to build a strong informative prior we need to compute numerous datasets, so finding a better optimization method will **accelerate** our

process a lot. (The difficulty here is that we do **NOT** have an *explicit* expression of loss function w.r.t variances. I skipped this before since it won't influence our final conclusions.)

- Suppose we are happy with the results we obtained (that is we don't care about these *tiny* difference anymore), the next thing we could try is to check the forecasting result based on our model. (I haven't do this before, but it should be pretty quick I think.)
- Finally, a very essential and important question is what if the results from *X-11* or *SEATS* are not correct/accurate?