STAT 7330
Final Project

Change-point Detection for Financial Markets Using Bayesian Inference

Yanghong Guo
yxg190031@utdallas.edu

May 2, 2022

Content

Background

- Financial markets have long been in the spotlight. A lot of researchers come up with methods to forecast upcoming market volatility and make investment decisions accordingly.

Background

- Financial markets have long been in the spotlight. A lot of researchers come up with methods to forecast upcoming market volatility and make investment decisions accordingly.
- There were several significant market movements in the history. The famous one being 2008 credit crisis.

Background

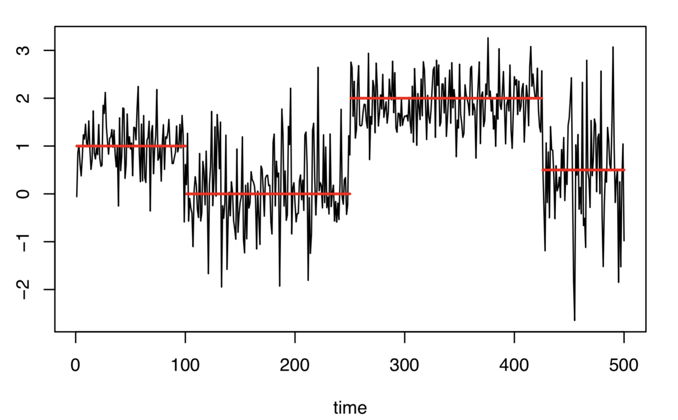
- Financial markets have long been in the spotlight. A lot of researchers come up with methods to forecast upcoming market volatility and make investment decisions accordingly.
- There were several significant market movements in the history. The famous one being 2008 credit crisis.
- Cryptocurrencies including Bitcoin have become an important investment instrument in the recent years from 2016. But this market fluctuates frequently.

Background

- Financial markets have long been in the spotlight. A lot of researchers come up with methods to forecast upcoming market volatility and make investment decisions accordingly.
- There were several significant market movements in the history. The famous one being 2008 credit crisis.
- Cryptocurrencies including Bitcoin have become an important investment instrument in the recent years from 2016. But this market fluctuates frequently.
- This project focuses on the change-point detection problems in the financial markets, i.e., stock market and cryptocurrency market.

Methods for Change-points Detection
Two Commonly Used Methods

Many time series are characterised by abrupt changes in structure. We consider change-points as time points that divides a data set into several distinct homogeneous segments.



Methods for Change-points Detection
Two Commonly Used Methods

Numbers of methods have be developed to solve change-point detection problems, and Bayesian Inference is widely used in those methods.

There are two main Bayesian approach as to solve change-point detection problems when the total number of segments is unknown:

Methods for Change-points Detection
Two Commonly Used Methods

Numbers of methods have be developed to solve change-point detection problems, and Bayesian Inference is widely used in those methods.

There are two main Bayesian approach as to solve change-point detection problems when the total number of segments is unknown:

- Classical Dirichlet process modeling.
- Designed Bayesian fusion modeling.

Methods With Designed Prior

Model Settings

In this work, we consider the following Gaussian mean problem:

$$y_i = \theta_i + \epsilon_i$$

where $\epsilon_i' s$ are iid normal error with unknown variance σ^2. Suppose we have in total n data points in the time series, i.e., $i = 1, 2, 3, ..., n$, and we aim at finding a partition $\{B_1, ..., B_s\}$ of $\{1, .., n\}$ such that $\theta_i' s$ are constant for all $i \in B_k$.

Since the number of blocks s with the partition is unknown, it can be treated as a clustering problem. The estimated θ_i^*, $i = 2, ..., n$, could be clustered by the criteria $I(|\theta_{i+1}^* - \theta_i^*| \leq c)$, where c is a chosen critical value. We assume that θ_i is a constant within each block, and differs between blocks.

There are multiple choices for the prior in this problem, such as **Laplace shrinkage prior, spike-and-slab shrinkage prior, Horseshoe shrinkage prior, and t-shrinkage prior.**

In this paper, we only consider Horseshoe shrinkage prior and t-shrinkage prior. These two have an advantage over the others, that thanks to the heavier tail of the distribution, the estimated θ_i^* tends to be less smooth, and the blocks are more distinguishable.

The t-shrinkage prior specification is

$$\theta_1|\sigma^2 \sim N(0, \lambda_0\sigma^2)$$
$$(\theta_i - \theta_{i-1})|\sigma^2 \overset{ind}{\sim} t_v(m\sigma^2),\ \textit{for}\ i = 2, .., n-1$$
$$\sigma^2 \sim IG(a_\sigma, b_\sigma)$$

where λ_0, a_σ, b_σ are positive hyperparameters, and $t_v(\xi)$ denotes a t-distribution with df= v and scale parameter ξ. Since the t-distribution can be expressed as a Normal-scale mixture with inverse Gamma, we can rewrite the prior as

$$\theta_1|\sigma^2 \sim N(0, \lambda_0\sigma^2)$$
$$(\theta_i - \theta_{i-1})|\sigma^2\lambda_i \overset{ind}{\sim} N(0, \lambda_i\sigma^2),\ \textit{for}\ i = 2, .., n-1$$
$$\lambda_i \overset{ind}{\sim} IG(a_t, b_t),\ \textit{for}\ i = 2, .., n-1 \qquad \sigma^2 \sim IG(a_\sigma, b_\sigma)$$

where hyperparameters a_t and b_t satisfy $v = 2a_t$ and $m = \sqrt{b_t/a_t}$. Here m is chosen so as to satisfy $P(|t_v(m)|) \geq \sqrt{\log(n)/n} \approx n^{-1}$

Methods With Designed Prior

t-shrinkage prior - Posterior Distribution

Given the Gaussian likelihood and prior specification, the conditional posterior distribution of the parameters of interest are in closed form and be estimated by Gibbs sampling.

$$\lambda_i | \sim IG(a_t + \frac{1}{2}, b_t + \frac{(\theta_i - \theta_{i-1})^2}{2\sigma^2}), i = 2, ..., n$$

$$\sigma^2 | \cdot \sim IG(a_\sigma + n, b_\sigma + \frac{\theta_1^2}{2\lambda_0} + \|y - \theta\|_2^2 + \sum_{i=2}^{n} \frac{(\theta_i - \theta_{i-1})^2}{2\lambda_i})$$

$$\theta_i | \cdot \sim N(\mu_i, \nu_i), i = 1, ..., n$$

where

$$\nu_i^{-1} = \frac{1}{\sigma^2}(1 + \frac{1}{\lambda_i} + \frac{1}{\lambda_{i+1}}), \quad \mu_i = \frac{\nu_i}{\sigma^2}(y_i + \frac{\theta_{i-1}}{\lambda_i} + \frac{\theta_{i+1}}{\lambda_{i+1}}), \quad i = 2, ..., n-1$$

$$\nu_1^{-1} = \frac{1}{\sigma^2}(1 + \frac{1}{\lambda_2}), \quad \mu_1 = \frac{\nu_1}{\sigma^2}(y_1 + \frac{\theta_2}{\lambda_2})$$

$$\nu_n^{-1} = \frac{1}{\sigma^2}(1 + \frac{1}{\lambda_n}), \quad \mu_n = \frac{\nu_1}{\sigma^2}(y_1 + \frac{\theta_{n-1}}{\lambda_n})$$

The Horseshoe shrinkage prior specification is

$$
\begin{aligned}
\theta_1|\lambda_1^2, \sigma^2 &\sim N(0, \lambda_1^2\sigma^2) \\
\eta_i|\lambda_1^2, \tau^2, \sigma^2 &\overset{ind}{\sim} N(0, \lambda_i^2\tau^2\sigma^2), i = 2, ..., n \\
\lambda_i^2|\nu_i &\overset{ind}{\sim} IG(1/2, 1/\nu_i), i = 2, ..., n \\
\tau^2|\xi &\sim IG(1/2, 1/\xi) \\
\nu_2, ..., \nu_n, \xi &\overset{ind}{\sim} IG(1/2, 1) \\
\sigma^2 &\sim IG(a_\sigma, b_\sigma)
\end{aligned}
$$

The hyperparameters a_σ and b_σ can be chosen in such a way that the corresponding prior becomes non-informative. The local scale parameter λ_1 is considered to be fixed as well.

Methods With Designed Prior

Horseshoe shrinkage prior - Posterior distributions

The conditional posteriors for the rest of the parameters are given by

$$\lambda_i^2|\cdot \sim IG(1, \frac{1}{\nu_i} + \frac{(\theta_i - \theta_{i-1})^2}{2\tau^2\sigma^2}),\ i = 2, ..., n$$

$$\sigma^2|\cdot \sim IG(n + a_\sigma, b_\sigma + \frac{1}{2}\left[\|y - \theta\|_2^2 + \frac{1}{\tau^2}\sum_{i=2}^{n}\frac{(\theta_i - \theta_{i-1})^2}{\lambda_i^2} + \frac{\theta_1^2}{\lambda_1^2}\right]$$

$$\tau^2|\cdot \sim IG(\frac{n}{2}, \frac{1}{\xi} + \frac{1}{2\sigma^2}\sum_{i=2}^{n}\frac{(\theta_i - \theta_{i-1})^2}{\lambda_i^2})$$

$$\nu_i|\cdot \sim IG(1, 1 + \frac{1}{\lambda_i^2}),\ i = 2, ..., n$$

$$\xi|\cdot \sim IG(1, 1 + \frac{1}{\tau^2})$$

By Gibbs sampler, we would generate MCMC samples for each parameter and associated hyperparameters.

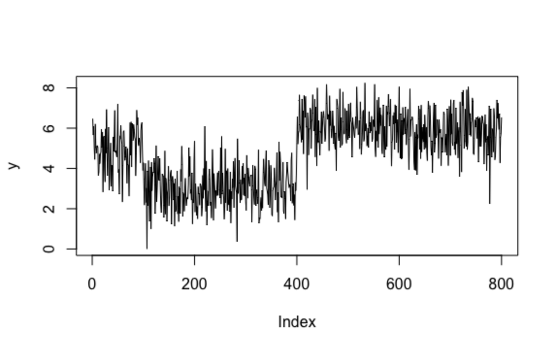
Simulation Data Analysis

The first set of data was generated from three normal distributions with different mean values and same variance $\sigma = 1$.

$$y1 \sim N(5, 1), \ \ n1 = 100$$

$$y2 \sim N(3, 1), \ \ n2 = 300$$

$$y3 \sim N(6, 1), \ \ n3 = 400$$



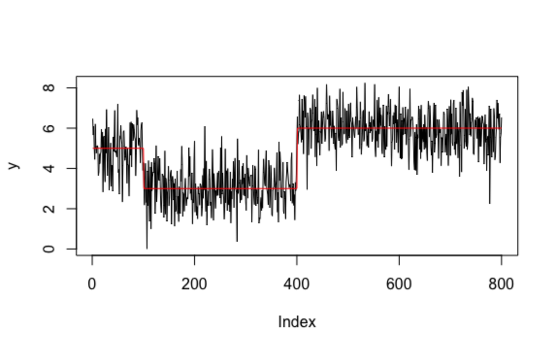
Simulation Data Analysis

The first set of data was generated from three normal distributions with different mean values and same variance $\sigma = 1$.

$$y1 \sim N(5, 1), \;\; n1 = 100$$

$$y2 \sim N(3, 1), \;\; n2 = 300$$

$$y3 \sim N(6, 1), \;\; n3 = 400$$



Simulation Data Analysis

Let's see how these two priors perform on this simulated data set.

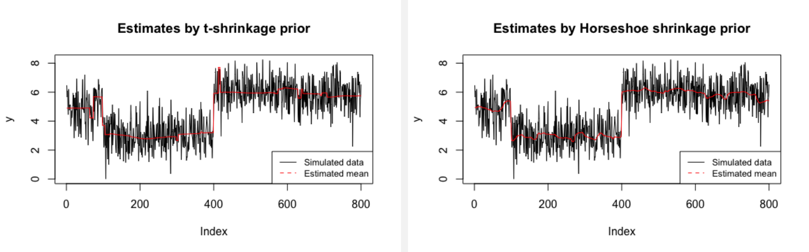


Figure: The performance of two priors on the first simulated data set.

We can see compared with Horseshoe prior, there are some undesired jumps in the figure of performance of t-shrinkage prior. Horseshoe prior performs a more stable and smooth estimation in this case.

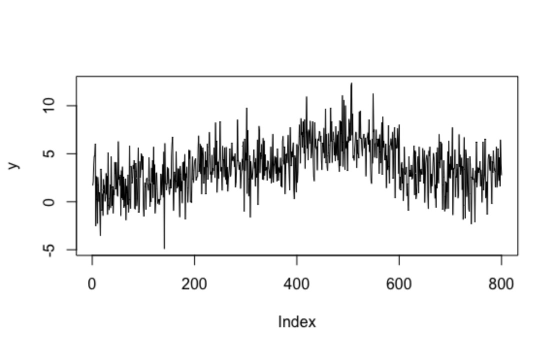
Simulation Data Analysis

For the second set of simulated data, we increase the variance to *sigma* $= 2$, and narrow the 'time interval' for each block.

$$y1 \sim N(2,4),\ \ n1 = 200 \qquad y2 \sim N(4,4),\ \ n2 = 200$$

$$y3 \sim N(6,4),\ \ n3 = 200 \qquad y3 \sim N(3,4),\ \ n3 = 200$$

The simulated data is



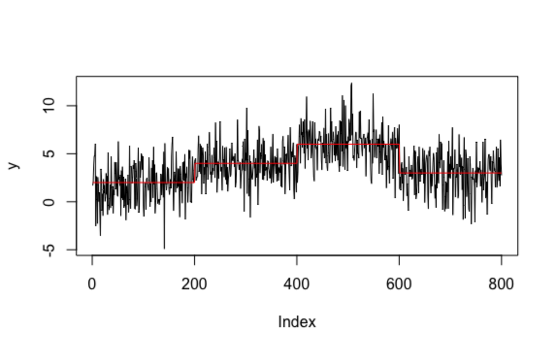
Simulation Data Analysis

For the second set of simulated data, we increase the variance, and narrows the 'time interval' for each block.

$$y1 \sim N(2,2),\ \ n1 = 200 \qquad y2 \sim N(4,2),\ \ n2 = 200$$

$$y3 \sim N(6,2),\ \ n3 = 200 \qquad y3 \sim N(3,2),\ \ n3 = 200$$

The simulated data is



Simulation Data Analysis

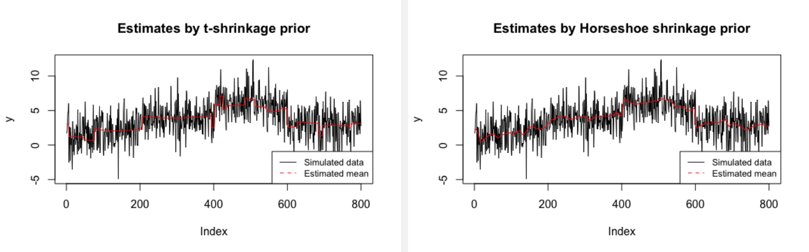
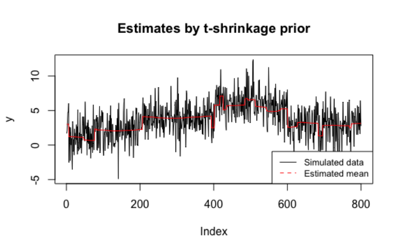


Figure: The performance of two priors on the second simulated data set.

Simulation Data Analysis



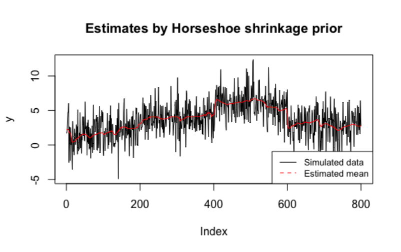


Figure: The performance of two priors on the second simulated data set.

The Horseshoe shrinkage prior generates a smooth estimation, while the result of t-shrinkage prior is more discrete. We can see in this case the t-shrinkage prior outperforms the horseshoe shrinkage prior.

Simulation Data Analysis

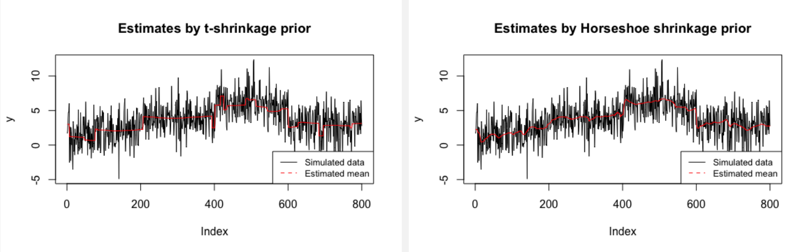


Figure: The performance of two priors on the second simulated data set.

The Horseshoe shrinkage prior generates a smooth estimation, while the result of t-shrinkage prior is more discrete. We can see in this case the t-shrinkage prior outperforms the horseshoe shrinkage prior.

In the real world, the stock market has a high volatility from it's unstable nature, and the difference of mean values between blocks is not that distinguishable, so we can deduce that t-shrinkage prior will have a better performance when applying to real financial data set.

Real Data Analysis

- We consider the change-point detection problem in the cumulative returns of two different markets in two different time periods, one for the Dow Jones Industrial Average (DJI) and the other for Bitcoin returns against USD (BTC). We carefully choose the two different time periods so as to include the known market crash or thrive phenomena.

Real Data Analysis

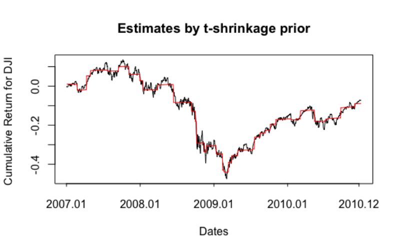
- We consider the change-point detection problem in the cumulative returns of two different markets in two different time periods, one for the Dow Jones Industrial Average (DJI) and the other for Bitcoin returns against USD (BTC). We carefully choose the two different time periods so as to include the known market crash or thrive phenomena.
- The time period that we choose to study DJI is from January 2007 to Dec 2010, that encompasses the 2008 Recession and the following recovery.

Real Data Analysis

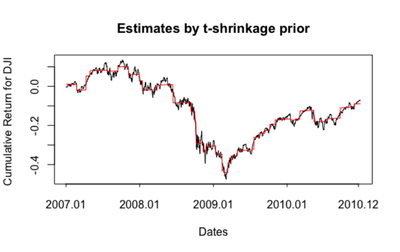
- We consider the change-point detection problem in the cumulative returns of two different markets in two different time periods, one for the Dow Jones Industrial Average (DJI) and the other for Bitcoin returns against USD (BTC). We carefully choose the two different time periods so as to include the known market crash or thrive phenomena.
- The time period that we choose to study DJI is from January 2007 to Dec 2010, that encompasses the 2008 Recession and the following recovery.
- The time period we choose to study BTC is from Jun 2018 to Jun 2020, that encompasses the main crash of Bitcoin in 2019 and the following recovery.

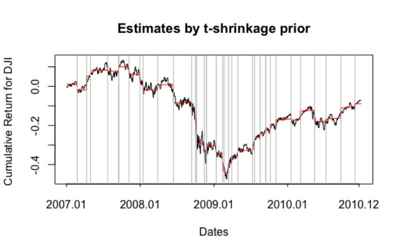
Real Data Analysis

DJI - t-shrinkage prior



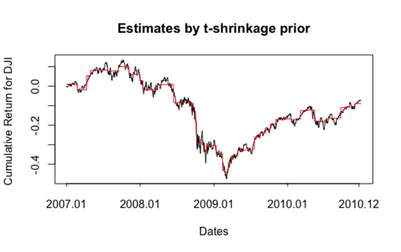
Real Data Analysis
DJI - t-shrinkage prior

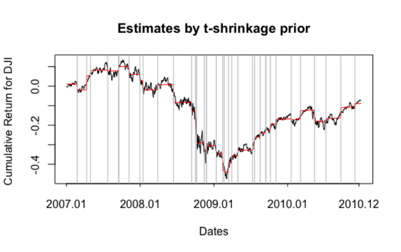




Real Data Analysis

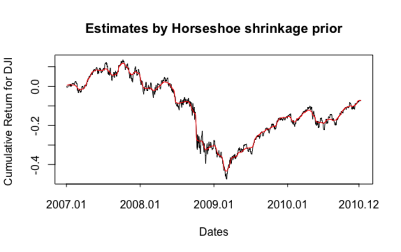
DJI - t-shrinkage prior





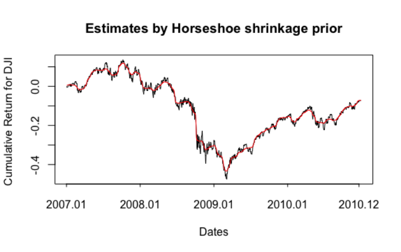
The criteria we applied to define a change-point is by $I(|\theta_i^* - \theta_j^*| \leq c)$, and the value c is chosen to be $c = \hat{\sigma} m t_{1/2n}$, where $\hat{\sigma}$ is the Bayesian estimates of σ given by the posterior mean, and $t_{1/2n}$ is the $(1 - 1/2n)$-quantile of a t distribution with $2a_t$ degrees of freedom.

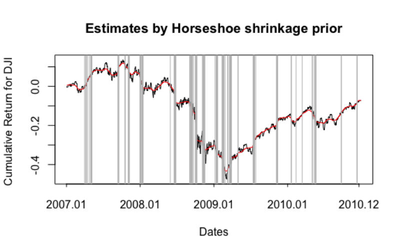
Real Data Analysis
DJI - Horseshoe shrinkage prior



Real Data Analysis

DJI - Horseshoe shrinkage prior

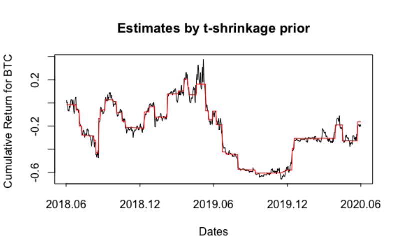




We find that the Horseshoe shrinkage prior method solution suffers from severe over-fitting problems, which is quite different from the solution given by t-shrinkage method that produces almost piecewise constant estimates. It's difficult to locate the change-points in this case.

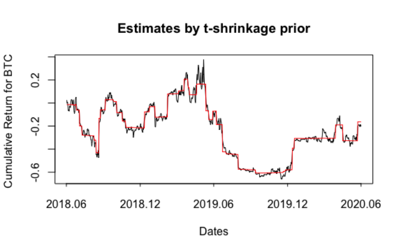
Real Data Analysis

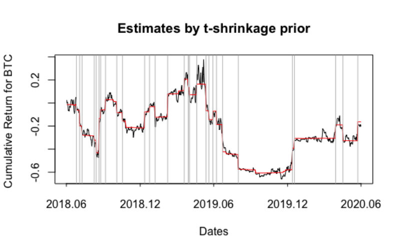
BTC - t-shrinkage prior



Real Data Analysis

BTC - t-shrinkage prior

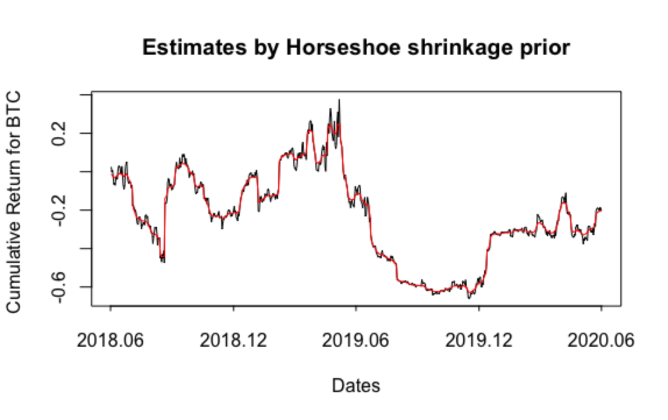




Unlike the traditional stock market, the cryptocurrency market fluctuates severe among time. t-shrinkage method successfully captures the most volatility period (2018.6 - 2019.6) and the relatively stable period (2019.6 - 2020.6).

Real Data Analysis

BTC - Horseshoe shrinkage prior



Horseshoe shrinkage prior method solution again suffers from severe over-fitting problems.

Discussion & Conclusion

- Horseshoe shrinkage prior and t-shrinkage prior can both be applied in change-point detection problems.

Discussion & Conclusion

- Horseshoe shrinkage prior and t-shrinkage prior can both be applied in change-point detection problems.
- Horseshoe shrinkage prior outperforms t-shrinkage prior when the variance of the data is relatively small and the differences between blocks are significant.

Discussion & Conclusion

- Horseshoe shrinkage prior and t-shrinkage prior can both be applied in change-point detection problems.
- Horseshoe shrinkage prior outperforms t-shrinkage prior when the variance of the data is relatively small and the differences between blocks are significant.
- t-shrinkage prior performs better when the variance of the data is relatively large, and the difference between blocks are insignificant.

Discussion & Conclusion

- Horseshoe shrinkage prior and t-shrinkage prior can both be applied in change-point detection problems.
- Horseshoe shrinkage prior outperforms t-shrinkage prior when the variance of the data is relatively small and the differences between blocks are significant.
- t-shrinkage prior performs better when the variance of the data is relatively large, and the difference between blocks are insignificant.
- The data set from financial market such as stocks and cryptocurrencies have a high volatility in nature, thus t-shrinkage fusion model has an overall better perform on these data sets.

Reference

Banerjee, S., Guhathakurta, K., 2020. Change-point analysis in financial networks. Stat 9 (1), e269.

Banerjee, S., Horseshoe shrinkage methods for Bayesian fusion estimation, Computational Statistics and Data Analysis, https://doi.org/10.1016/j.csda.2022.107450

Banerjee, S., Shen, W., 2022. Graph signal denoising using t-shrinkage priors. J. Stat. Plan. Inference219, 279–305.

Eckley, I.A., Fearnhead, P. and Killick, R., 2011. Analysis of changepoint models. Bayesian time series models, 205, p.224.

Piironen, J. and Vehtari, A., 2017. Sparsity information and regularization in the horseshoe and other shrinkage priors. Electronic Journal of Statistics, 11(2), pp.5018-5051.

Song, Q., Cheng, G., 2019. BayesIan fusion estimation via t shrinkage. Sankhya A 1–33.

Thank you for watching!