STAT 7330 Project : Change-point Detection for Financial Markets Using Bayesian Inference

Yanghong Guo
Department of Statistics
University of Texas at Dallas
yxg190031@utdallas.edu

Abstract

The ability to detect change-points in the stock markets is of vital importance for both the investors and researchers. It is an established fact that the fluctuation of stock price in one country can have a serious impact on other markets across the globe. In this paper we introduce a Bayesian fusion method using t-shrinkage prior and Horseshoe shrinkage prior to detect the change-points in the main stock market and the cryptocurrency market. Fast and efficient computational procedures are presented via Markov Chain Monte Carlo methods exploring the full posterior distributions of all the parameters of interest. The model is then implemented on two sets of simulated data, the historical data of Dow Jones Industrial Average (DJI) cumulative return, and Bitcoin-USD cumulative return. The results show that our model works efficiently detecting the potential crash or thrive in certain stock market or cryptocurrency market in advance and helps make investment decision accordingly.

1 Introduction

1.1 Background

The credit crisis of 2008 has propelled a volume of research on stock market. Banerjee and Guhathakurta (2020) introduced a kernel-based approach to detect the existing and potential change-points in the mainstream stock markets across the globe. Inspired by Song and Cheng (2020), Banerjee and Shen (2022) introduced a heavy-tailed t-shrinkage prior on a chain graph, the excellent empirical performance of which method is demonstrated via simulation studies and applications to stock market data. Banerjee (2022) also came up with a similar Bayesian fusion method, but rather than t-shrinkage prior, they utilized Horseshoe shrinkage prior. Piironen and Vehtari (2017) also points out that Horseshoe prior is a noteworthy alternative for sparse Bayesian estimation. Both t-shrinkage prior and Horseshoe shrinkage prior induce a strong shrinkage effect on successive differences in the mean parameters thanks to their heavy-tailed structure, thus impose sufficient prior concentration for non-zero values of the same, and give block-shaped estimates for mean parameter.

Many time series are characterised by abrupt changes in structure. We consider change-points as time points which divide a data set into several distinct homogeneous segments. Numbers of methods have be developed to solve change-point detection problems in time series, and Killick et al. (2012) discussed the mainstream change-point detection methods in this case. Besides, Bayesian Inference is also widely used in change-point detection problems. There are two main Bayesian approach, the first one being classical Dirichlet process, and the second one being designed fusion prior. In this paper, the method we discuss with different prior settings is classified as a designed fusion prior.

1.2 Model specification

We first denote the time-series data we obtained from stock markets by $y_{1:n} = (y_1, ..., y_n)^T$, here y_i represents the cumulative return on day i. The main focus of this paper is the estimation of a piecewise constant mean value $\boldsymbol{\theta} = (\theta_1, ..., \theta_n)^T$. In particular we consider the following Gaussian mean model:

$$y_i = \theta_i + \epsilon_i$$

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., $i = 1, 2, 3, ..., n$,in this case 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^*, $i = 1, ..., n$, could be clustered by the criteria $I(|\theta_i^* - \theta_j^*| \leq c)$, where c is a chosen critical value. We assume that θ_i is a constant within each block, and differs between blocks.

2 Bayesian model

In our model, we conduct Bayesian analysis of $\boldsymbol{\theta}$. Motivated by the recent work in Song and Cheng (2020) and Banerjee (2022), we consider heavy-tailed priors for the difference between blocks to facilitate better convergence. In particular, we use t-shrinkage prior and Horseshoe shrinkage prior on the successive differences of mean values, i.e., $(\theta_i - \theta_{i-1})$ for every $i = 2, .., n - 1$. Thanks to the shrinkage property of these two priors, the estimated $\theta_i^* - \theta_{i-1}^*$ tends to be more sparse, and the blocks are more distinguishable.

2.1 Prior settings

t-shrinkage prior The t-shrinkage prior specification is given as:

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

where λ_0, a_σ, b_σ are positive hyperparameters, and $t_v(\xi)$ denotes a t-distribution with v degrees of freedom 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:

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

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 the condition $P(|t_v(m)|) \geq \sqrt{\log(n)/n} \approx n^{-1}$. The chioce of hyperparameters in a hierarchical Bayesian model setup plays a crucial role. In our case, the hyperparameters correspnding to the Inverse Gamma prior for the variance parameter σ^2 are chosen such that the prior is almost non-informative, e.g.,$a_\sigma = b_\sigma = 0.5$. For the choice of a_t, we follow the approach in Song and Cheng (2020), which sets $a_t = 2$. We choose $\lambda_0 = 5$ based also on the recommendation in Song and Cheng (2020).

Horseshoe shrinkage prior Inspired by Banerjee (2022), the Horseshoe shrinkage prior specification is:

$$\theta_1|\lambda_1^2, \sigma^2 \sim N(0, \lambda_1^2\sigma^2)$$
$$\eta_i|\lambda_1^2, \tau^2, \sigma^2 \overset{ind}{\sim} N(0, \lambda_i^2\tau^2\sigma^2), i = 2, ..., n$$
$$\lambda_i \overset{ind}{\sim} C^+(0, 1), i = 2, ..., n$$
$$\tau \sim C^+(0, 1)\sigma^2 \sim IG(a_\sigma, b_\sigma)$$

where $C^+(0, \psi)$ denotes a half-Cauchy distribution. The half-Cauchy distribution can further be written as a scale mixture of Inverse-Gamma distributions. For a random variable $D \sim C^+(0, \psi)$, we can write

$$D^2|\phi \sim IG(1/2, 1/\phi), \quad \phi \sim IG(1/2, 1/\psi^2)$$

Thus, the full hierarchical prior specification for our model is given by:

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

The hyperparameters a_σ and b_σ can also 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.

2.2 Posterior computation

t-shrinkage prior Given the Gaussian likelihood and prior specification, the conditional posterior distribution of the parameters of interest are in closed form and can be estimated by Gibbs sampler.

$$\begin{aligned}
\lambda_i| &\sim IG(a_t + \frac{1}{2}, b_t + \frac{(\theta_i - \theta_{i-1})^2}{2\sigma^2}), \;\; i = 2, ..., n \\
\sigma^2|\cdot &\sim IG(a_\sigma + n, b_\sigma + \frac{\theta_1^2}{2\lambda_0} + \|y - \theta\|_2^2 + \sum_{i=2}^{n} \frac{(\theta_i - \theta_{i-1})^2}{2\lambda_i}) \\
\theta_i|\cdot &\sim N(\mu_i, \nu_i), \;\; i = 1, ..., n
\end{aligned}$$

where

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

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

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

Horseshoe shrinkage prior Similarly to t-shrinkage prior, the conditional posterior distributions for the parameters of interest are given by:

$$\begin{aligned}
\lambda_i^2|\cdot &\sim IG(1, \frac{1}{\nu_i} + \frac{(\theta_i - \theta_{i-1})^2}{2\tau^2\sigma^2}), \;\; i = 2, ..., n \\
\sigma^2|\cdot &\sim IG(n + a_\sigma, b_\sigma + \frac{1}{2}\Big[\|y - \theta\|_2^2 + \frac{1}{\tau^2}\sum_{i=2}^{n} \frac{(\theta_i - \theta_{i-1})^2}{\lambda_i^2} + \frac{\theta_1^2}{\lambda_1^2}\Big] \\
\tau^2|\cdot &\sim IG(\frac{n}{2}, \frac{1}{\xi} + \frac{1}{2\sigma^2}\sum_{i=2}^{n} \frac{(\theta_i - \theta_{i-1})^2}{\lambda_i^2}) \\
\nu_i|\cdot &\sim IG(1, 1 + \frac{1}{\lambda_i^2}), \;\; i = 2, ..., n \\
\xi|\cdot &\sim IG(1, 1 + \frac{1}{\tau^2}) \\
\theta_i|\cdot &\sim N(\mu_i, \zeta_i), \;\; i = 1, ..., n
\end{aligned}$$

where

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

$$\zeta_1^{-1} = \frac{1}{\sigma^2}(1 + \frac{1}{\lambda_2^2\tau^2} + \frac{1}{\lambda_1^2}), \quad \mu_1 = \frac{\zeta_1}{\sigma^2}(y_1 + \frac{\theta_2}{\lambda_2\lambda_2^2})$$

$$\zeta_n^{-1} = \frac{1}{\sigma^2}(1 + \frac{1}{\lambda_{n-1}^2\tau^2} + \frac{1}{\lambda_n^2}), \quad \mu_n = \frac{\zeta_n}{\sigma^2}(y_1 + \frac{\theta_{n-1}}{\lambda_n\lambda_2^2})$$

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

3 Simulation studies

We consider two sets of simulated data from normal distributions, each of which have a overall same variance and several blocks with different mean values.

3.1 Simulated data I

The first set of data was generated from three normal distributions with different mean values and same variance $\sigma = 1$ and the total number of data points are $n = 800$ (Figure 1):

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

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

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

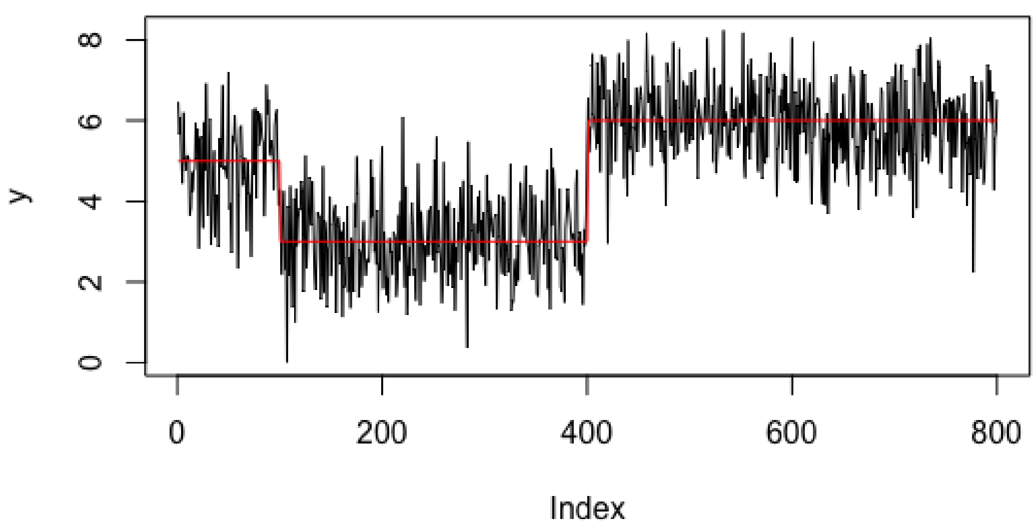

Figure 1: Simulated data set with $\sigma = 1$ and three different mean values. Black line represents the simulated data, and the red line is the true mean values of the data set.

By implementing the Bayesian fusion model with t-shrinkage prior and Horseshoe shrinkage prior, Figure 2 shows the results of the estimates for both applying these two priors. From the figure, we can see that Horseshoe shrinkage prior outperforms t-shrinkage prior in this case since the result generated by t-shrinkage prior contains some undesired jumps in the estimated mean value. In contrast, Horseshoe prior produces a more stable and smooth estimation.

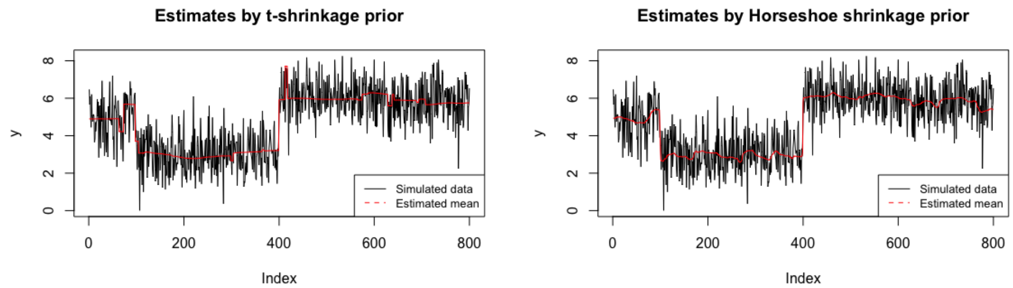

Figure 2: The performance of two priors on the first simulated data set. Black line represents the simulated data, and the red line is the estimated mean value.

3.2 Simulated data II

The second set of simulated data contains also $n = 800$ simulated data points, while different from the first set, this one has a larger overall variance $\sigma = 2$ (Figure 3). The length of each block narrows, and the difference of mean value between blocks decreases:

$$y1 \sim N(2,4),\ n1 = 200 \qquad y2 \sim N(4,4),\ n2 = 200$$
$$y3 \sim N(6,4),\ n3 = 200 \qquad y4 \sim N(3,4),\ n4 = 200$$

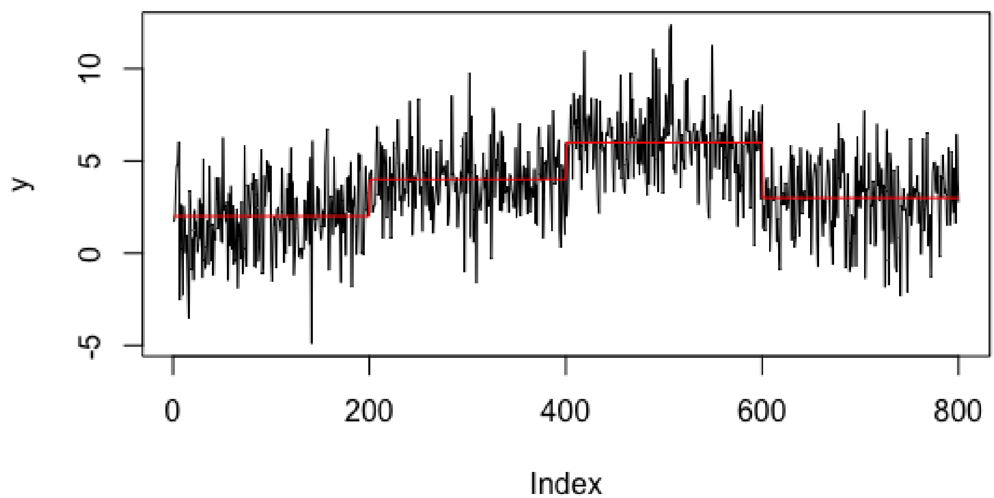

Figure 3: Simulated data set with $\sigma = 2$ and four different mean values. Black line represents the simulated data, and the red line is the true mean values of the data set.

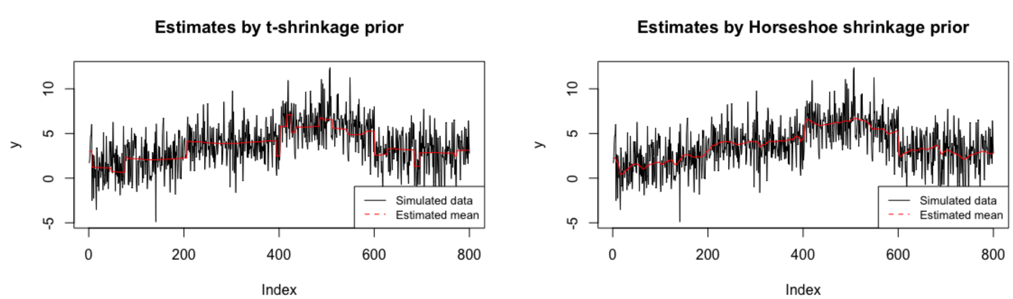

Figure 4: The performance of two priors on the second simulated data set. Black line represents the simulated data, and the red line is the estimated mean value.

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 by successfully distinguishing between different mean values among blocks (Figure 4).

4 Real data estimation

In the real world, the stock markets has a high volatility from their unstable nature, and the difference of mean values between blocks is not clear, so from the characteristics of t-shrinkage prior and Horseshoe shrinkage prior, we can deduce that t-shrinkage prior may have a better performance when applying to real financial data set.

We consider the problem of change-point detection in the cumulative returns of two different markets in two different time periods on a daily basis, 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.

4.1 DJI return from 2007 to 2010

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.

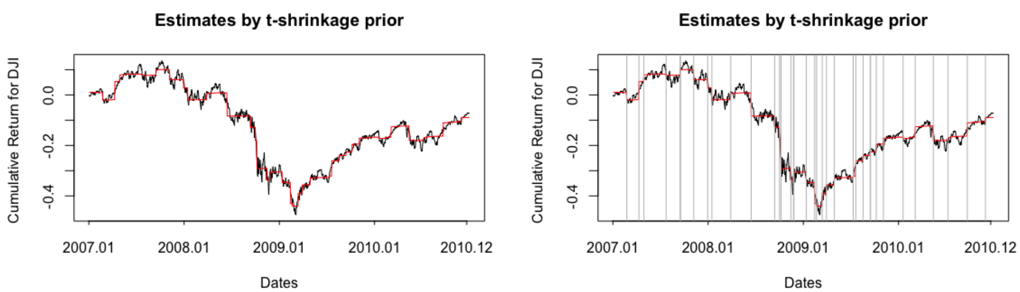

Figure 5: The performance of t-shrinkage prior on DJI returns from 2007.1 to 2010.12. The black line represents the real data, the red line is the estimated mean value, and the grey vertical lines are the detected change-points.

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. The grey vertical lines in Figure 5 (right) show the position of the detected change-points.

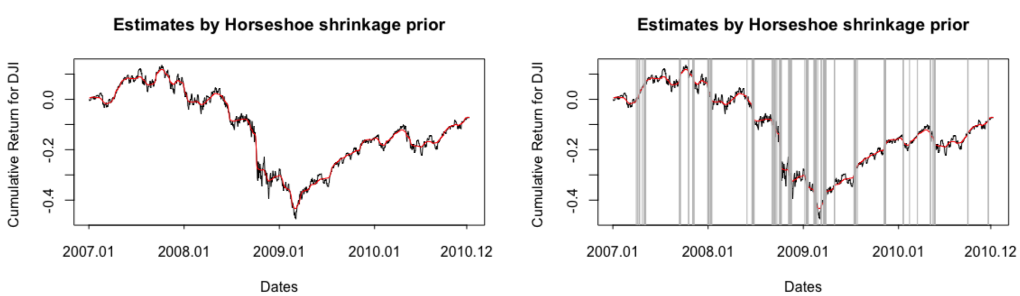

Figure 6: The performance of Horseshoe shrinkage prior on DJI returns from 2007.1 to 2010.12. The black line represents the real data, the red line is the estimated mean value, and the grey vertical lines are the detected change-points.

By comparing the two results, we find that the solution of Horseshoe shrinkage prior model (Figure 6) suffers from severe over-fitting problems, while t-shrinkage prior model produces almost piecewise constant estimates. It's difficult to locate the change-points when the estimation is too smooth, as what is generated by Horseshoe shrinkage prior in this case.

4.2 Bitcoin return from 2017 to 2019

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.

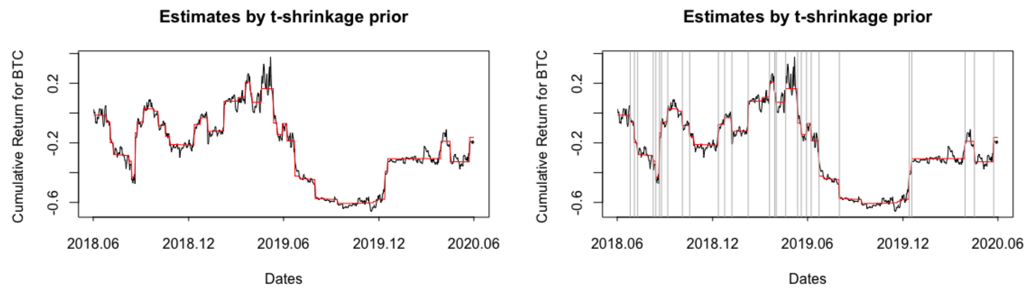

Figure 7: The performance of t-shrinkage prior on DJI returns from 2018.6 to 2020.6. The black line represents the real data, the red line is the estimated mean value, and the grey vertical lines are the detected change-points.

Unlike the traditional stock market, the cryptocurrency market fluctuates severe among time (Figure 7). The model with t-shrinkage prior successfully captures the most volatility period (2018.6 - 2019.6) and then a relatively stable period (2019.6 - 2020.6).

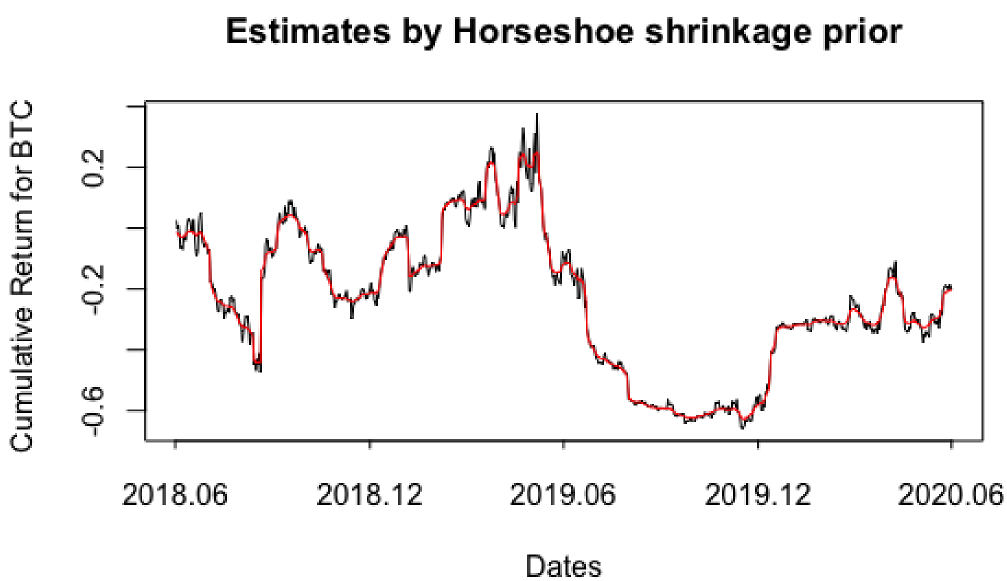

Figure 8: The performance of Horseshoe shrinkage prior on DJI returns from 2007.1 to 2010.12. The black line represents the real data, and the red line is the estimated mean value.

Horseshoe shrinkage prior method solution again suffers from severe over-fitting problems (Figure 8). By the same criteria, it is impossible to locate the change-point positions.

5 Conclusion

In this paper we discuss the choice of prior in a designed fusion model for the change-point detection problem, and thanks to the properties of Horseshoe shrinkage prior and t-shrinkage prior, these two priors can both be applied in this case. Horseshoe shrinkage prior outperforms t-shrinkage prior when the variance of the data is relatively small and the differences between blocks are significant, while 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 prior has an overall better performance on these data sets.

References

Banerjee, S.; Guhathakurta, K. *Stat* **2020**, *9*, e269.

Song, Q.; Cheng, G. *Sankhya A* **2020**, *82*, 353–385.

Banerjee, S.; Shen, W. *Journal of Statistical Planning and Inference* **2022**,

Banerjee, S. *Computational Statistics & Data Analysis* **2022**, 107450.

Piironen, J.; Vehtari, A. *Electronic Journal of Statistics* **2017**, *11*, 5018–5051.

Killick, R.; Fearnhead, P.; Eckley, I. A. *Journal of the American Statistical Association* **2012**, *107*, 1590–1598.