A Trading Strategy Based on Finite Mixture Models

Bingyan HAN

Abstract

In this report, I will briefly discuss the Bayesian Averaging trading strategy based on finite mixture models, appeared in Anderson et al. (2016). Numerical results are presented.

1. Algorithm

Suppose there are n risky assets (stocks) and one risk-free asset (bank account) with risk-free rate r.

Then we try to use finite mixture models to estimate the distributions of stock returns. We assume

$$R_t \sim \sum_{k=1}^{K} \pi_k f_k(R_t; \theta_k). \tag{1}$$

Instead of assuming the fixed K components in total, we use a dynamic procedure to add new components to the mixture and adjust the probability π_k with certain method. We will call each component as a 'model' in later context. And we use $P_t(m|\mathcal{F}_{t-1})$ instead of π_m to indicate that the probability will change with time.

Each component assumes the excess return R_t follows a multivariate normal distribution with constant mean μ and covariance matrix Σ:

$$R_t \sim \mathcal{N}(\mu, \Sigma)$$

It means $f_k(R_t; \theta_k)$ is the normal pdf. At time $t-1$, prior belief about mean returns, μ and covariance matrix Σ are:

$$\Sigma \sim \mathcal{IW}(\Lambda_{m,t-1}, \nu_{m,t-1})$$
$$\mu|\Sigma \sim \mathcal{N}(\mu_{m,t-1}, \Sigma/\kappa_{m,t-1})$$

where $\mathcal{IW}(\cdot)$ is the inverse-Wishart distribution. There are two subscripts and $m, t-1$ indicates it's model m 's parameter at time $t-1$. Therefore, μ, Σ follows the Normal-inverse-Wishart distribution with $\Lambda_{t,t-1} = \delta_{t,t-1}\Sigma_{t,t-1}, \nu_{t,t-1} = \delta_{t,t-1} + n + 1$. The subscript $t, t-1$ indicates these are the beliefs in the particular model t at time $t-1$.

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{t,t-1}, \kappa_{t,t-1}, \Lambda_{t,t-1}, \nu_{t,t-1})$$

The new component is introduced in the beginning of new trading period, before new data is observed. The time line of this strategy is shown in Figure 1.

New model is born. We set the initial value for new model t's parameter as:

$$\mu_{t,t-1} = \bar{\mu}_{t-1}\mathbf{1}_n, \quad \kappa_{t,t-1} = 1, \quad \Sigma_{t,t-1} = \bar{\lambda}_{t-1}\mathbf{I}_n, \quad \delta_{t,t-1} = 1.$$

Date		Events
$t-1$	▷	There are $t-1$ existing models that possibly describe past asset returns. Each model m has probability $P_{t-1}(m\vert\mathcal{F}_{t-1})$.
t	▷	A new model, model t, is born and probabilities for all models are adjusted for its inclusion. Each model m now has probability $P_t(m\vert\mathcal{F}_{t-1})$.
	▷	Excess returns at time t, R_t, are observed. Parameters for each model are updated. Probabilities are updated so that each model m now has probability $P_t(m\vert\mathcal{F}_t)$.
	▷	Expectations of excess returns at time $t+1$, R_{t+1}, are formed.
	▷	Optimal portfolio choices, ϕ_t, at time t are computed.
$t+1$	▷	Excess returns at time $t+1$, R_{t+1}, are observed and excess portfolio returns $\phi_t' R_{t+1}$ are realized.

Figure 1: Timeline for BA strategy.

where

$$\bar{\mu}_{t-1} = \frac{1}{n}\sum_{i=1}^{n}\bar{\mu}_{i,t-1},$$

$$\bar{\lambda}_{t-1} = \frac{1}{n(t-2)}\sum_{i=1}^{n}\sum_{s=1}^{t-1}(R_{i,s}-\bar{\mu}_{i,t-1})^2,$$

$$\bar{\mu}_{i,t-1} = \frac{1}{t-1}\sum_{s=1}^{t-1}R_{i,s}.$$

Then we adjust the probabilities of each model according to the following rule. This is called the sharing priors.

$$P_t(m|\mathcal{F}_{t-1}) = \begin{cases} \sum_{q=1}^{m}\frac{1}{t-q+1}P_{t-1}(q|\mathcal{F}_{t-1}) & \text{if } m < t, \\ \sum_{q=1}^{t-1}\frac{1}{t-q+1}P_{t-1}(q|\mathcal{F}_{t-1}) & \text{if } m = t. \end{cases} \tag{2}$$

The second subscript is still $t-1$ since return at time t is not observed at this stage.

New excess return is observed. We update parameters when new observation R_t arrives. According to Bayes rule, each model's parameter

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{m,t-1}, \kappa_{m,t-1}, \delta_{m,t-1}\Sigma_{m,t-1}, \delta_{m,t-1}+n+1)$$

will be updated to

$$\mu, \Sigma \sim \mathcal{NIW}(\mu_{m,t}, \kappa_{m,t}, \delta_{m,t}\Sigma_{m,t}, \delta_{m,t}+n+1)$$

where

$$\mu_{m,t} = \frac{\kappa_{m,t-1}\mu_{m,t-1} + R_t}{\kappa_{m,t}},$$
$$\Sigma_{m,t} = \frac{\delta_{m,t-1}\kappa_{m,t}\Sigma_{m,t-1} + \kappa_{m,t-1}(R_t - \mu_{m,t-1})(R_t - \mu_{m,t-1})'}{\delta_{m,t}\kappa_{m,t}}$$

and $\kappa_{m,t} = \kappa_{m,t-1} + 1, \delta_{m,t} = \delta_{m,t-1} + 1$. The proof can be found at Murphy (2012), Page 134, Equation (4.210) to (4.214).

Updating model probabilities when new return observed. By Bayes rule, we have following results, the proof can be found in Murphy (2007), Page 21, Equation (264) to (266). M_t is the set of all the models available in time t.

$$P_t(m|\mathcal{F}_t) = \frac{L(R_t|m, \mathcal{F}_{t-1})P_t(m|\mathcal{F}_{t-1})}{\sum_{m\in M_t} L(R_t|m, \mathcal{F}_{t-1})P_t(m|\mathcal{F}_{t-1})}.$$

$$L(R_t|m, \mathcal{F}_{t-1}) = \frac{\kappa_{m,t-1}^{n/2}|\Lambda_{m,t-1}|^{\nu_{m,t-1}/2}\Gamma_n(\nu_{m,t}/2)}{\pi^{n/2}\kappa_{m,t}^{n/2}|\Lambda_{m,t}|^{\nu_{m,t}/2}\Gamma_n(\nu_{m,t-1}/2)}.$$

where $\Lambda_{m,t} = \delta_{m,t}\Sigma_{m,t}, \nu_{m,t} = \delta_{m,t} + n + 1$.

The second subscript changes to t since return at time t is observed.

Strategy under mean-variance preference. If the investor uses mean-variance preference, and she wants to choose the investment strategy ϕ_t to maximize

$$E(\phi_t' R_{t+1} + r|\mathcal{F}_t) - \frac{\theta}{2} Var(\phi_t' R_{t+1} + r|\mathcal{F}_t)$$

where θ is the trade-off between risk and expected profit. The optimal strategy is given by

$$\phi_t = \frac{1}{\theta}\hat{\Sigma}_t^{-1}\hat{\mu}_t.$$

where

$$\hat{\mu}_t = \sum_{m\in M_t} \mu_{m,t}P_t(m|\mathcal{F}_t),$$
$$\hat{\Sigma}_t = \sum_{m\in M_t} \Big(\frac{1+\kappa_{m,t}}{\kappa_{m,t}}\Sigma_{m,t} + \mu_{m,t}\mu_{m,t}'\Big)P_t(m|\mathcal{F}_t) - \hat{\mu}_t\hat{\mu}_t'.$$

ϕ_t is used as the trading strategy at time $t+1$.

2. Numerical Results

I randomly select 12 stocks with weekly price data from Jan 01, 2000 to Dec 31, 2016. The ticker for these stocks are 'AAPL', 'AVB', 'AMD', 'CCC', 'BA', 'BLL', 'C', 'INTC', 'GPS', 'MMM', 'T', 'TJX'. The dataset is downloaded from Yahoo! Finance. A graph drawn from the data can be found in Figure 3. One important thing is the correlation between any two stock. If some of them are highly correlated, we may have problems in calculating the inverse of covariance matrix Σ. In other words, we assume $\Sigma \succeq \epsilon I$ for some constant $\epsilon > 0$. This is called the complete market hypothesis. The biggest estimated correlation is roughly 0.54. So this assumption is satisfied.

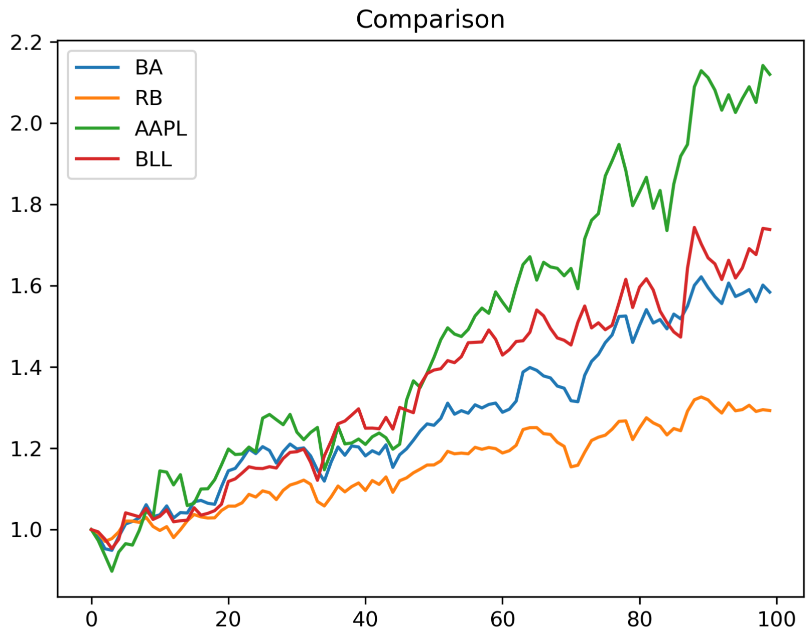

Figure 2: Comparison between Bayesian Averaging (BA) strategy, Rebalanced strategy (RB) and best two stocks, AAPL and BLL.

We start trading from week 701 and ended at week 800. The first 700 weeks' data is used to train our models. Suppose we have 1 dollar at the beginning of week 701. We choose $\theta = 0.001$. After calculation, we get the annualized Sharpe ratio of the Bayesian averaging strategy is 1.36 and annualized return 27%. Annualized Sharpe ratio is calculated as

$$\text{Annualized Sharpe ratio} = \frac{\text{portfolio return - risk-free rate}}{\text{standard deviation of portfolio return}} \times \sqrt{52} \tag{3}$$

To make our results more reliable, we compare this strategy with other two benchmarks. The first one is the so-called rebalanced strategy. The idea is as follows: we also suppose the initial wealth is 1 dollar. The investor equally puts his money in 12 stocks and one risk-free asset, i.e., each stock with 1/13 dollar and 1/13 dollar for savings account. After one week, the investor has x dollars now. He changes his position by invest x dollars equally in 12 stocks and one risk-free asset, i.e., $x/13$ for each stock and $x/13$ for the risk-free asset. He repeats the same operation until stop trading. That's why we can it rebalanced. The second benchmark is a holding strategy. We suppose the investor has superpower and he can know which stock performs best in next 100 weeks. And he invests his all 1 dollar into this stock and do nothing else. We get the Sharpe ratio of rebalanced strategy is 0.79. Annualized return is 14%. The final value of holding each stock is 2.12, 1.33, 0.59, 1.26, 1.47, 1.74, 1.04, 1.44, 0.99, 1.48, 1.08, 1.32, with the same order as the tickers given before. We can see Apple is the best stock in these 100 weeks and our BA strategy can't beat it. However, BA strategy can beat 10 stocks out of 12 stocks in this portfolio. Figure 2 shows the comparison between them in a graph. The whole investment strategy is plotted in Figure 4. Recall that we have 1 dollar at the beginning.

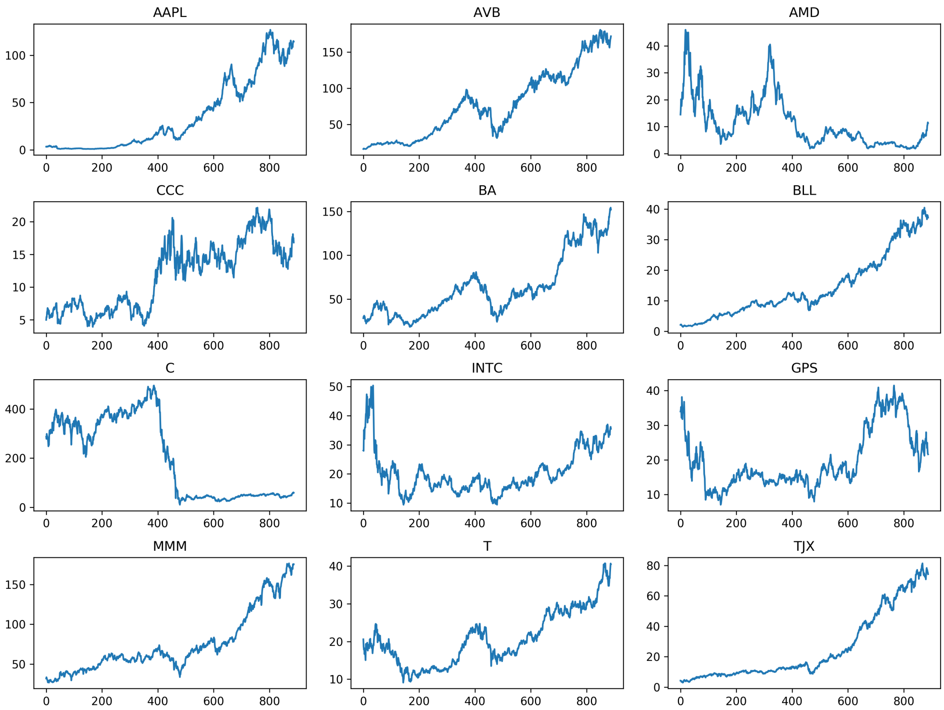

Figure 3: 888 weeks data of 12 stocks, adjusted from stock split and dividends. Downloaded from Yahoo! Finance.

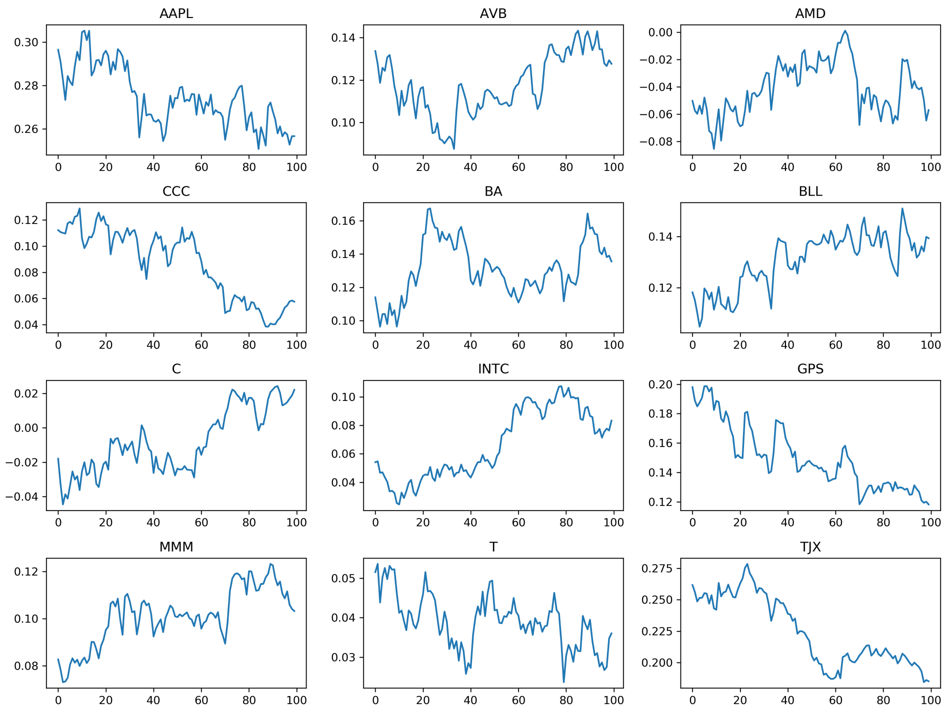

Figure 4: Money invested in each stock. Short selling is allowed in our strategy.

3. Further Discussion

- I didn't consider the transaction costs in our strategy. Usually, it can cause significant impact on profit if we change our position very dramatically. The graph showed that our strategy didn't change position very often.

- The correct way to calculate the risk-free rate is using the US Treasury bills. For example, since we use weekly data, the risk-free rate should be calculated from month T bills. However, I use constant $r = 0.001$ for all the time for simplicity.

- The number of components in our model is change with time. It's interesting to investigate some possible alternatives to choose the number of components.

- Final issue is the reliability of our numerical studies. We should test more data sets including real ones and simulated ones to get a more reliable result.

- Penalty term on model misspecification can be considered. That's why the reference paper is called 'robust'. No short selling constraint should also be considered if the market doesn't allow invest negative money in stocks.

References

Anderson, E. W., Cheng, A.-R., 2016. Robust Bayesian Portfolios Choices, *The Review of Financial Studies*, 29(5):1330–1375.

Murphy, K.P., 2007. Conjugate Bayesian analysis of the Gaussian distribution, *Lecture Notes*.

Murphy, K.P., 2012. Machine learning: A probabilistic perspective, *The MIT Press*.