

A Markov Chains Application To Stock Markets

Federico Camerota

Main idea

- We aim to use Markov Chains (MC) to model the state of a stock market and the price of the single stocks in it. The behaviors are studied on a daily basis so we use discrete time MCs.
- The purpose is to study the ability of predicting future trends of stock prices of such a model and to investigate if the information obtained by simulating the stock market with the model can be used in a portfolio selection strategy with good results.

Market chain

- For the market we consider two possible states: **bad** and **good**.
- A transition to the state *good* is defined as an increase in the overall market value. While a transition to the *bad* state is defined as a decrease of such quantity.
- So the MC describing the market is characterized by the following transition matrix

	<i>bad</i>	<i>good</i>
<i>bad</i>	$1 - p_{b2g}$	p_{b2g}
<i>good</i>	p_{g2b}	$1 - p_{g2b}$

Stock chain

- We consider as initial state the last observed stock price and assume that the daily price variation is a decrease or an increase by a fixed δ that depends on the state of the market.
- The market state influences both the daily variation *delta* and the probability of an increase/decrease. However these probabilities are assumed to be independent of the current stock price.

Stock chain

- We end up with a complex transition matrix, however, due to the previous assumptions, the stock MC can be described by the following parameters:
 - 1 $\delta_{g,m}$ and $\delta_{b,m}$ that describe the daily price variation of stock m when the market is in good and bad state.
 - 2 $p_{g,m}$ and $p_{b,m}$ that are the probabilities of an increase in the price of stock m when the market is in good and bad state.

Comments and possible extensions

- The definition of market states is quite naive, this could be improved by considering trends and the amount of variation in the market value.
- Another underlying assumption is that the stock chains are independent. The model could be improved by considering correlations, for example clustering stocks with similar time evolution and then simulating a common trajectory for the cluster and the discrepancy from it for each stock in the cluster.
- In the stock chain the daily variation is kept fixed, an extension of the model that has been explored is the usage of a proposal distribution whose parameters are estimated from the data. We tested normal distributed daily variations.

Data Set

- The data used has been obtained from Kaggle and are daily prices of stocks in the New York Stock Exchange (NYSE).
- From the data we define the overall market price as the sum of all stock prices.

Data model

- To estimate the parameters of the model we consider the time series of ups and downs is the prices of both the market and the stocks. Encoding an up with 1 and a down with 0.
- Under the assumptions of the model the market trend series is a sequence of Bernoulli random variables with parameter that depends on the previous observation.

$$M_t \sim \text{Bernoulli}(p_t)$$

where p_t is equal to p_{b2g} if $M_{t-1} = 0$ and $1 - pg2b$ otherwise.

Data model

- The stock trend series instead turn out to be independent Bernoulli variables with parameter that depends on the current market state

$$S_{t,m} \sim \text{Bernoulli}(p_{t,m})$$

where $p_{t,m}$ is equal to $p_{g,m}$ if $M_t = 1$ and $p_{b,m}$ otherwise.

Parameter Estimation

- Considering the market trend series and a single stock trend series, the parameters in the two MCs can be estimated through maximum likelihood quite easily.
- The estimators found this way can be extended to all the stocks considered due to the hypothesis that the stock chains are independent.
- In fact, the estimators for the market chain depend only on the market behavior while the estimators for the stock chain depend only on the stock observed trends.
- For the *delta* parameters in the stock chains we propose three different approaches:
 - 1 Mean daily variations
 - 2 Median daily variations
 - 3 10th percentile of daily variations

Comments and possible extensions

- Another possible extension of the model could be the usage of **continuous time** MCs. However, since the available data is on a daily basis we consider it won't make much sense in this framework. Given more frequent data this approach could yield a better model.
- In the parameter estimation we give the same importance to all the observations considered, we could instead use different weight profiles for each stock in order to exploit the information present in the time series and have better parameter estimates. On the proposed data model the usage of **weighted likelihood** seems a bit difficult so this extension has not been studied.

Experiments

- We propose two study cases to test our model.
- The first experiment (*portfolio selection*) consists in using the results of a fixed number of simulations ran using our model as data in a portfolio selection problem and compare its performance with respect to a portfolio obtained only using historical data.
- In the second experiment we study the ability of the model to predict future stock trajectories.

Portfolio selection

- As selection strategy we use the maximization of the *Sharpe ratio* defined as the ratio of the expected return and variance.
- To solve the problem we use standard numerical methods provided in the *scipy* module. We also tested the usage of an heuristic method to solve the problem but the runtime is considerably higher and results don't seem to be superior.
- The strategy based on historical data uses expected returns and variances estimated over the last 200 days while the strategy based on simulations uses expected returns and variance of the simulated return for a given time horizon.
- The experiment has been conducted using three different time horizons: 7, 15 and 30 days.

Portfolio selection results

As a benchmark we consider a strategy that invests the same amount of capital on each stock in the market.

content...