MARKOV CHAINS

Transition matrix for S&P 500 historical prices

Author: Hakob Hakobyan

Assuming that the stock price of S&P 500 obeys Markov property, the **goal** of this research is to construct the transition matrix for this Markov Chain.
For this research I have taken **data** of [S&P 500 historical prices] for 10 years since 2012 (even though data of historical prices is available since 1928) on a daily basis assuming that trends for the past decade may vary. The time series we get consists of over 2500 observations.
The main perimetersI have used are Date (daily), Close price, Return (rate of stock price increase compared to the previous day), Statement (where B= big, S=small, D=decrease, I=increase), which reflects in what direction prices have changed compared to the previous-day prices, Status, which reflects the transition to current statement (BDSI = from statement SI to BD), Dmean (which is the mean of all negative returns), and Imean (mean of all positive returns).

C2 fx =(B2-B3)/B3

	A	B	C	D	E	F	G
1	Date	Close*	Return	Statement	Status	Dmean	Imean
2	Nov 09, 2022	3,748.57	-0.02077788	BD	BDSI	-0.00725716	0.00695301
3	Nov 08, 2022	3,828.11	0.00559788	SI	SIBI		
4	Nov 07, 2022	3,806.80	0.00961398	BI	BIBI		
5	Nov 04, 2022	3,770.55	0.01361868	BI	BIBD		
6	Nov 03, 2022	3,719.89	-0.01058598	BD	BDBD		
7	Nov 02, 2022	3,759.69	-0.02500194	BD	BDSD		
8	Nov 01, 2022	3,856.10	-0.00410126	SD	SDBD		
9	Oct 31, 2022	3,871.98	-0.00745438	BD	BDBI		
10	Oct 28, 2022	3,901.06	0.02462638	BI	BISD		
11	Oct 27, 2022	3,807.30	-0.00608260	SD	SDBD		
12	Oct 26, 2022	3,830.60	-0.00738771	BD	BDBI		
13	Oct 25, 2022	3,859.11	0.01626665	BI	BIBI		
14	Oct 24, 2022	3,797.34	0.01188195	BI	BIBI		
15	Oct 21, 2022	3,752.75	0.02372483	BI	BIBD		
16	Oct 20, 2022	3,665.78	-0.00795094	BD	BDSD		
17	Oct 19, 2022	3,695.16	-0.00667208	SD	SDBI		

Returns that are not positive are considered as a Decrease, elsewhere an Increase. If the value of return is lower than the Dmean, it is denoted as BD, if it is higher (but still negative), it is denoted as SD. On the other hand if the value of return is positive & smaller than Imean it is denoted as SI, and BI if higher than the Imean.
In this research I have taken the means instead of the median quartile values in order to define the statements by high/low-increase/decrease values.
Operations. In order to get a transition matrix we should take every t-1 statement, and count the probability of each transition from t-1 to t states. For example when we take “BI” as t-1 statement, we have 508 observations on that position, and should count the probability for every status containing “BI” as t-1 statement (BDBI, SDBI,

SIBI, BIBI). So we take the number of observations for any status and divide that into the respective statement (BDBI/BI). I have used RStudio to count the number of observations.

```
> table(data$Statement)

 BD  BI  SD  SI 
375 508 774 858 
> table(data$States)

BDBD BDBI BDSD BDSI BIBD BIBI BISD BISI SDBD SDBI SDSD SDSI SIBD SIBI SISD SISI 
  89   96  101   89  118  105  154  131   78  152  224  320   89  155  296  318 
```

As a result we get the probabilities for each status, when we already have a certain t-1 statement.

	BI	SI	SD	BD	SUM
BI(t-1)	0.2067	0.3051	0.2992	0.1890	1.00
SI(t-1)	0.1527	0.3706	0.3730	0.1037	1.00
SD(t-1)	0.1990	0.3824	0.2894	0.1305	1.00
BD(t-1)	0.3147	0.2373	0.2080	0.2373	1.00

Conclusion. As we can see the transition matrix we get by this research is not uniform as the values of probabilities in various states visually differ from each other. We have got the highest probability of transition from the state “Small Decrease” to “Small Increase” with the value of 0.38, and the lowest probability of transition from “Small Increase” to “Big Decrease” equal to 0.10.