**VERJETNOST PREKOMERNEGA PRILEGANJA**

Naslov mojega projekta je verjetnost prekomernega prileganja. Moja naloga je bila za dani nabor trgovalnih strategij izračunati verjetnost prekomernega prileganja z novo metodo kombinatoričnega simetričnega prečnega preverjanja.

//

**Opis dela:**

**1. podatki:**

Pa bi kar predstavila kako sem se naloge lotila.

Najprej sem pridobila podatke za cene delnic, ki so trenutno v indeksu S&P 500.

*S&P 500 (oz. Standard & Poor's 500) je borzni indeks, ki temelji na tržni kapitalizaciji 500 velikih podjetji, katerih navadne delnice kotirajo na Newyorški borzi (NYSE) ali NASDAQ-u (National Association of Securities Dealers Automated Quotations).*

Podatke sem dobila iz spletne strani Yahoo finance s pomočjo funkcije vgrajene v R. Vzela sem dnevne cene delnic od 1. januarja 2000 do 1. novembra 2013 (slabih 13 let). Kasneje se je pri uporabi metode izkazalo, da je to preveč podatkov, ki jih računalnik ne bo mogel obdelati v doglednem času, zato sem vzela le podatke za 5 let, od začetka 2008 do konca 2012. Vzela sem zadnjo ceno delnic.

Tiste delnice, ki so imele več kot tri manjkajoče vrednosti zaporedoma sem izločila. To je bodisi zato ker delnica ne obstaja dovolj dolgo, se je z njo nehalo trgovati ali pa pač ni bilo podatkov. Ostale manjkajoče vrednosti, ki so še ostale sem nadomestila z zadnjo ceno prejšnjega trgovalnega dne. Tako mi je na koncu ostalo 411 delnic za katere sem imela podatke za 3330 trgovalnih dni, ki sem jih kasneje skrčila na 1260 trgovalnih dni.

//

Nato sem sestavila nekaj enostavnih trgovalnih strategij ter implementirala metodo kombinatoričnega simetričnega prečnega preverjanja.

//

**2. strategije:**

Zdaj bi vam na hitro predstavila trgovalne strategije, ki sem jih uporabila.

- Ena izmed strategij je strategija Random, ki naključno izbere, ali imam dano delnico ali ne.

//

- Naslednja strategija je Buy & Hold, kjer sem delnico na začetku kupila in jo imela do konca obdobja.

//

Ostale trgovalne strategije sem sestavila na podlagi indikatorjev. Uporabila sem indikatorje SMA, RSI in Bollinger.

//

- SMA (simple moving average) oz. preprosto drseče povprečje je aritmetična sredina gibanja cene delnice za izbrano dolžino obdobja. Uporabila sem dolžine 5, 25, 50 in 150 dni. Nakupni znak za to strategijo je, ko zadnja cena zraste čez krivuljo drsečega povprečja, prodajni pa, ko pade pod njo.

Tukaj je primer zadnje cene delnice indeksa SP500 v lanskem letu in indikatorja SMA25 V obdobju, ko je zadnja cena nad SMA25, bi imeli delnico, ko pa je pod SMA25, je ne bi imeli.

//

- Naslednja strategija temelji na indikatorju RSI (Relative Strength Indeks) oz. indeksu relativne moči. Ta indikator primerja velikosti rasti z velikostmi padcev cene delnice. Njegove vrednosti so med 0 in 100. Če vrednost indeksa pade pod 30, je to znak za nakup vrednostnega papirja *(podcenjenost),* če pa zraste nad 70, je to znak za prodajo vrednostnega papirja *(precenjenost).* Uporabila sem strategije z indikatorjem RSI na obdobjih dolžine 2 in 14 dni.

*Formula za izračun RSI: RSI = 100 – 100/(1-RS), kjer je RS povprečna vrednost pozitivnih period v opazovanem obdobju deljeno s povprečno vrednostjo negativnih period. Pozitivne in negativne periode so izračunane kot absolutna sprememba tečaja od zaključnega do zaključnega tečaja.*

Primer indikatorja RSI2 za indeks SP500 v zadnjem letu in meji 30 in 70. To so nakupni signali (spodaj) in to prodajni (zgoraj)

//

- Zadnja strategija, ki sem jo uporabila temelji na Bollingerjevih pasovih. Sestavljeni so iz treh krivulj: srednja krivulja predstavlja srednjeročni trend. Običajno je to preprosto drseče povprečje (SMA). Zgornja krivulja je za faktor krat standardni odklon premaknjena navzgor, spodnja krivulja pa je za faktor krat standardni odklon premaknjena navzdol. Jaz sem uporabila faktor 1. Za izračun SMA in standardnega odklona se uporablja enako časovno obdobje, jaz sem uporabila obdobje 20 dni.

*V času visoke razpršenosti cen se bollingerjevi pasovi razširijo, v času nizke razpršenosti cen pa zožijo.*

Če je cena pod spodnjim pasom, je to znak za nakup delnice, če pa je nad zgornjim pasom je to znak za prodajo.

Primer bollingejevega pasu za indeks SP500 v zadnjem letu. SMA20 in 1 standardni odklon.

//

**3. uporaba strategij:**

Predstavljene strategije sem uporabila na podatkih za 500 oz. 411 delnic. Vsaki delnici v portfelju sem na začetku namenila 1000 eurov kapitala. Ta kapital sem potem uporabila za trgovanje s to delnico. Če je bil znak za nakup sem z vsem denarjem, ki je bil namenjen tej delnici kupila delnice, če je bil znak za prodajo, sem prodala vse delnice te vrste in nato denar držala, do signala za nakup. Vrednost celotnega portfelja sem izračunala kot vsoto dnevnih vrednosti za posamezno delnico.

Slika prikazuje dnevne vrednosti portfelja za posamezno strategijo tekom 5 let (od 1.1.2008 do 1.1.2013).

Kot je razvidno iz grafa je imel na koncu 5 letnega obdobja največjo vrednost portfelj, pri katerem sem uporabila strategijo RSI14. Sledi ji Boillinger, ki temelji na SMA20 in 1 standardnem odklonu. Najmanjšo vrednost pa je imel portfelj, pri katerem sem uporabila strategijo SMA5, ostale strategije pa so si večino časa približno podobne.

//

**4. metoda kombinatoričnega simetričnega preverjanja:**

Ko sem imela dnevne vrednosti portfelja za vse strategije, sem se lotila metode kombinatoričnega simetričnega prečnega preverjanja, s katero sem kasneje izračunala verjetnost prekomernega prileganja.

Najprej, kaj sploh je prekomerno prileganje in kaj verjetnost prekomernega prileganja.

Ko imamo na voljo nekaj strategij se želimo odločiti, katera bi bila najboljša da bi jo uporabili za trgovanje. Običajno vzamemo tisto, ki ima največjo uspešnost na zgodovinskih podatkih. Lahko pa se zgodi, da bo ta strategija neuspešna na nekih drugih podatkih.

Za mero uspešnosti strategije sem uporabila Sharpovo razmerje, ki ga dobimo tako, da pričakovano stopnjo donosa portfelja, (*od katere običajno odštejemo netvegano stopnjo donosa*,) delimo s standardnim odklonom donosov portfelja. Večje je Sharpovo razmerje, bolj uspešna je trgovalna strategija.

Temu, če je strategija, ki je najboljša na učni množici, zelo slaba na testni množici rečemo prekomerno prileganje. Razlog za to je, da je strategija preveč povezana z učno množico.

Oz če povem definicijo

DEFINICIJA (prekomerno prileganje zgodovinskega preverjanja): Naj bo n najboljša strategija na učni množici, **to je ima največje Sharpovo razmerje na učni množici izmed vseh strategij**. Proces izbire strategije se prekomerno prilega, če za najboljšo strategijo na učni množici velja, da je njena pričakovana uspešnost na testni množici, manjša od mediane uspešnosti vseh strategij na testni množici**.**

DEFINICIJA (verjetnost prekomernega prileganja): Naj bo n najboljša strategija na učni množici. Ker n ni nujno tudi najboljša strategija na testni množici, obstaja pozitivna verjetnost, da je njena uspešnost na testni množici manjša od mediane uspešnosti vseh strategij na testni množici. Tej verjetnosti rečemo verjetnost prekomernega prileganja.

//

Kako metoda kombinatoričnega simetričnega prečnega preverjanja deluje:

Potrebujemo dnevne donose portfelja za posamezne strategije. Recimo da imamo N strategij, ki jih opazujemo na T trgovalnih dneh, lahko jih zložimo v matriko velikosti TxN

Nato obdobje v katerem opazujemo strategije razdelimo na več podobdobij enake dolžine, recimo na S.

Izberemo pol podobdobij, ki jih vzamemo za učno množico, drugo polovico podobdobij pa vzamemo za testno množico. To naredimo za vse možne kombinacije pobdobij – S nad S/2 kombinacij.

Za vsako kombinacijo naredimo naslednje:

Za vsako strategijo izračunamo uspešnost strategije na učni množici (Sharpovo razmerje) in določimo strategijo, ki je najbolj uspešna na učni množici (ima največje Sharpovo razmerje). Recimo ji strategija n. Nato izračunamo še uspešnost posamezne strategije na testni množici.

Za strategijo, ki je najboljša na učni množici, določimo relativni rank oz. relativno uspešnost na testni množici. Označimo ga z omega, ki je med 0 in 1.

Nato definiramo lambdo kot …. Lambda je 0, ko je ko je uspešnost strategije n na testni množici enaka mediani uspešnosti vseh strategij na testni množici. Visoke vrednosti lambde pomenijo konsistenco med uspešnostjo na učni in testni množici, kar kaže na nizko raven prekomernega prileganja

Za vsako izmed kombinacij razdelitve podatkov na učno in testno množico dobimo eno lambdo. Določimo frekvenčno porazdelitev teh lambd.

Verjetnost prekomernega prileganja lahko ocenimo z vsoto frekvenc lambd, ki so negativne, kar predstavlja odstotek pri katerem je optimalna strategija na učni množici slabša od mediane strategij na testni množici. Recimo ji Fi.

//

1. Če je Fi približno 0, je bila v veliki večini kombinacij strategija, ki je bila najboljša na učni množici tudi zelo dobra na testni množici. To pomeni, da ni velikega prekomernega prileganja, in izbor optimalne strategije na učni množici pripomore k večji uspešnosti na testni množici.
2. Če je Fi = ½, je bila v veliki večini kombinacij strategija, ki je bila najbolj uspešna na učni množici enaka mediani uspešnosti na testni množici. To pomeni, da se zgodovinsko preverjanje prekomerno prilega do te mere, da postopek izbora optimalne strategije na učni množici ne doda vrednosti.
3. Če je Fi >> ½, je bila v veliki večini kombinacij strategija, ki je bila najboljša na učni množici slabša od mediane uspešnosti na testni množici. To pomeni, da je prekomerno prileganje tako veliko, da izbor optimalne strategije na učni množici povzroči slabšo pričakovano uspešnost na testni množici, kot bi bila, če bi naključno izbrali eno izmed strategij.

//

**5. Rezultati**

Metodo sem uporabila na več različnih kombinacijah strategij, ki sem jih prej opisala. Tukaj so verjetnosti prekomernega prileganja za nekatere kombinacije:

Če pogledamo grafe, kjer je prikazan kapital portfelja za kombinacije teh strategij, se zdi rezultat dokaj logičen. Visoko verjetnost prekomernega prileganja sem dobila, ko se kapitali portfelja pri danih strategijah prepletajo čez celo obdobje – enkrat ima ena strategija višjo vrednost kapitala, enkrat druga. Nizko verjetnost prekomernega prileganja pa sem dobila takrat, ko ima ena strategija izrazito večji kapital od ostalih strategij tekom celotnega obdobja.