

Modele de traiectorii aleatoare în studiul piețelor de capital

Autor: MORTICI C. CORINA MIHAELA

Academia de Studii Economice din București

Facultatea de Cibernetică, Statistică și Informatică Economică

Rezumat: Scopul prezentei lucrări este de a examina procesul de tip mers la întâmplare, prin intermediul a doi din cei mai importanți indici bursieri de pe BVB, și anume BET și BET Plus. Ipoteza mersului la întâmplare sugerează faptul că evoluția prețurilor acțiunilor are un caracter pur aleator, în consecință, profitul așteptat pentru un jucător al pieței este zero. Economisții care au studiat fenomenul au folosit procesele stochastice pentru a testa dacă ipoteza piețelor eficiente în sens slab este validă, această ipoteză fiind în general implicația manifestării unui proces de tip random walk. Când prețurile acțiunilor sau valorile indicilor bursieri nu prezintă fluctuații întâmplătoare, investitorii pot utiliza instrumente analitice pentru a studia prețurile perioadelor anterioare și a face predicții care să le aducă supraprofit. În ipoteza raționalității investitorilor și a neutralității la risc, o astfel de „versiune de EMH presupune că informația disponibilă din piață înaintea perioadei actuale nu ar trebui să fie utilă în predicția prețurilor viitoare”. Cu alte cuvinte, prețurile acțiunilor pe piața de capital nu sunt corelate una cu cealaltă; acest lucru înseamnă că modelul statistic al ipotezei piețelor eficiente este valid. În studiul întreprins, am efectuat teste precum: Augmented Dickey Fuller, Phillips Perron, KPSS, testul funcției de autocorelație, testul raportului de variație Lo MacKinley, modelarea fenomenului în sens GJR-GARCH(1,1), testul Wald Wolfowitz, testul Chow, testul Gujarati, modelarea fenomenului în sens ARCH(12). Au fost utilizate în acest sens date zilnice din perioada februarie 2016-noiembrie 2017. Aceste teste confirmă și susțin manifestarea fenomenului de mers la întâmplare în evoluția indicilor de pe Bursa de Valori București.

Abstract: The aim of this paper is to examine the random walk in two of the stock indexes of Bucharest Stock Exchange (BET and BET Plus). Random walk hypothesis states that stock prices move randomly; as a result, expected profit for the speculator is zero. Many economists believe that random walk can be applied to test the efficient market hypothesis in the weak level. Early literature used stochastic processes to test whether prices precluded everyone from easy profit and whether prices followed those processes or not. When stock prices do not fluctuate randomly, some investors can use past stock prices to gain abnormal return. Assuming rationality and risk neutrality, a version “of the efficient market hypothesis states that information observable to the market prior to week t should not help to predict the return during week t ”. In other words, stock returns are not correlated to one another; consequently, statistical model of the efficient market hypothesis holds and changes in returns are independent from one another. We employ different tests, such as: ADF, PP, autocorrelation function test, variance ratio test, GJR-GARCH modelling, Wald Wolfowitz test, Chow test, Gujarati test, Multiple Breakpoint test, ARCH modelling. Daily data on returns covered the period February 2016 – November 2017. These tests support the common results that the random walk theory is valid for the two indexes then the Bucharest Stock Market is weak-form efficient.

Cuvinte cheie: mers la întâmplare, proces stochastic, modificări structurale, piață de capital

Clasificare JEL: C46

Clasificare REL: 10B

PROCESELE DE TIP MERS LA ÎNTÂMPLARE ȘI IPOTEZA PIEȚELOR EFICIENTE

Ipoteza piețelor eficiente (EMH), cunoscută și sub numele de Teoria Mersului la Întâmplare, se referă la eficiența informațională de pe piețele de capital. În literatura de specialitate, termenul de piață eficientă este folosit pentru a explica relația de dependență între informația disponibilă și prețul acțiunilor. Aceste concepte au fost introduse și definite de Eugene Fama în anul 1970, în perspectiva căruia eficiența pe piețele financiare este dată de viteza de reacție și de ajustare a prețurilor la informația nou apărută pe piață.

În acest sens, în contextul unei piețe eficiente, prețurile acțiunilor din perioada curentă ar trebui să reflecte în totalitate informația relevantă pentru a face predicții cu privire la acestea, astfel încât să nu existe posibilitatea de a genera profit în exces folosind această informație. Prin urmare, factorii de decizie în ceea ce privește eficiența în sensul descris sunt: măsura în care informația este absorbită, timpul necesar în care este acumulată, dar și tipul de informație astfel încorporată.

1 Wooldridge, J., „Introductory Econometrics, A Modern Approach”, Pearson, 2009, pag. 385

Prețul unui activ reflectă valoarea actuală a venitului pe care acesta este speculat să îl genereze într-o perioadă imediat viitoare. Venitul așteptat este influențat de determinanți precum volatilitatea, lichiditatea, și riscul de faliment. În timp ce prețurile sunt determinate și stabilite în mod rațional, schimbările în preț se așteaptă a fi aleatoare și nepredictibile, deoarece informația nouă este impredictibilă prin natura sa. Așadar, prețurile pe piețele de capital sunt considerate a urma un proces de tip mers la întâmplare.

Mai concret, aruncarea unei monede sau selecția unei secvențe de numere dintr-un tabel cu numere aleatoare sunt numai câteva exemple de procese random walk. Revenind în cazul piețelor financiare, prețul din perioada curentă este independent și necorelat cu alte modele de evoluție a prețurilor din trecut.

Numim că o variabilă stochastică X urmează un proces de tip la întâmplare dacă:

$$X(t + 1) = \delta + X(t) + \varepsilon_{t+1} \quad (1)$$

unde δ este parametrul drift, iar eroarea de predicție ε_{t+1} este identic și independent distribuită.

1. Raționamentul lucrării

Scopul prezentei lucrări este de a analiza validitatea și corespondența cu piețele reale a conceptului teoretic de proces de mers la întâmplare. În acest sens, s-a considerat relevantă ilustrarea practică a cazului pieței de capital din România, mai exact, prin studierea evoluției a doi dintre cei mai importanți indici evaluați de Bursa de Valori București, și anume, BET și BET Plus, folosind date zilnice a valorii acestora pe o perioadă de doi ani. Caracterul aleator al acestor serii de timp este în strictă concordanță cu realitatea numai dacă piața este eficientă în sensul slab. Prin intermediul lucrării de față, ne propunem să studiem implicit și eficiența pieței de capital românească, contribuind cu rezultatele obținute la literatura existentă.

Evaluarea acțiunilor și a indicilor bursieri constituie o funcție importantă a piețelor financiare, întrucât conduce către formarea strategiilor de investiție a jucătorilor. Evaluarea valorii acțiunilor este de o importanță deosebită pentru a determina comportamentul piețelor, predicție posibilă numai dacă este știut tipul de eficiență a acestora. Luarea în considerare a informației disponibile este crucială pentru că poate duce la oportunități de arbitraj. Achiziționarea și vinderea simultană a unei acțiuni similare pe două piețe diferite ca rezultat al sesizării diferențelor în prețul acestora se conceptualizează sub forma ideii de arbitraj. Efectul arbitrajului joacă un rol esențial în eficiența unei piețe pentru că acest fenomen readuce prețurile către valoarea lor fundamentală.

Dacă piețele sunt eficiente în sens slab, nu este posibil ca jucătorii să cumpere o acțiune a cărei preț este subevaluat și să le vândă pe alte piețe, unde acestea sunt evaluate corect sau sunt supraestimate. Tocmai evenimentul descris face imposibil câștigul neuzitat al jucătorilor implicați.

În acest sens, întrebarea evidentă rămâne: de ce investitorii analizează piața cu instrumente sofisticate și consumatoare de timp dacă eforturile lor sunt zadarnice? Aceasta pare a fi principala preocupare a lucrării de față, întrucât investitorii în totalitate raționali nu ar juca, nu ar investi dacă nu ar avea șansa de a bate piața. Așa cum Lo, Mamaysky și Wang (2000) afirmă, “cu ajutorul tehnicilor statistice neparametrice sofisticate...[analizării] se bot bucura doar de o putere de predicție modestă”, de cele mai multe ori insuficientă pentru a juca bazându-se pe fundamentele acestor strategii complexe.

2. Date utilizate

Datele privind valorile de închidere ale indicilor bursieri BET și BET Plus din cadrul Bursei de Valori București din perioada 15 februarie 2016 – 2 noiembrie 2017 au fost analizate în studiul întreprins în această lucrare. Datele au fost colectate de pe website-ul oficial al Bursei de Valori București (www.bvb.ro) și prelucrate de autor în vederea efectuării testelor relevante în îndeplinirea scopului prezentat mai sus.

3. Limitări

Această lucrare este conturată de analiza și de testele efectuate pe seriile de timp BET și BET Plus asumând raționalitate și neutralitate la risc, și neluând în considerare costurile de tranzacție și alte taxe ce pot fi percepute pentru jucătorii pe piața de capital, în general, și pe piața de capital locală, în particular. Perioada analizată (435 de observații) a fost decizia subiectivă a autorului, și nu s-a încercat identificarea unei dimensiuni optime a timpului de observare a evoluțiilor indicilor.

4. Metodologie și aspecte teoretice

4.1. Efectul ARCH – Modelarea de tip ARCH a fenomenului

Seriile financiare sunt caracterizate, în mod frecvent, de volatilitate, fenomen ce este modelat prin procese de tipul ARCH. Efectul ARCH definește situația în care speculatorii pieței financiare estimează varianța dintr-o anumită perioadă prin informația apărută în perioada anterioară, și inclusă în model prin termenul ARCH. Acest context descrie conceptul cunoscut de “volatility clustering”, ceea ce înseamnă că perioadele de schimbări de magnitudine ridicată sunt urmate de perioade de fluctuații liniștite. Mai exact, schimbările semnificative în seriile financiare tind să se clusterizeze împreună, iar schimbările de magnitudine scăzută manifestă același comportament. Dacă în seriile de date ar fi prezent efectul ARCH, atunci se poate afirma că acestea sunt într-o anumită măsură predictibile și că răspund la speculațiile de pe piață (un exemplu cunoscut ar fi efectul de weekend).

Testând prezența acestui fenomen prin intermediul testului ARCH LM, precizăm că ipoteza nulă a testului este

$$H_0: \text{Nu există efect arch}$$

cu alternativa:

$$H_1: \text{Efectul arch este prezent în seria de date studiată.}$$

Regresia estimată va fi:

$$u_t^2 = \alpha_0 + \alpha_1 u_{t-1}^2 + \dots + \alpha_p u_{t-p}^2 + v_t, (2)$$

unde u_t reprezintă reziduurile din regresia inițială estimată prin metoda celor mai mici pătrate.

Statistica testului este $T * R^2$, cu T numărul de observații incluse în analiză, iar R^2 coeficientul de determinație al regresiei inițiale. Aceasta urmează o distribuție Chi-square cu p grade de libertate.

4.2. Clasificatorul Naiv Bayesian

Algoritmul Naiv Bayesian reprezintă o tehnică de clasificare bazată pe teorema lui Bayes, ce are la bază ipoteza de independență a predictorilor. Altfel spus, Clasificatorul Naiv Bayesian presupune faptul că prezența unei anume caracteristici într-o clasă este necorelată cu prezența oricărei alte caracteristici. De exemplu, un fruct este identificat ca fiind un măr dacă acesta este roșu, rotund, și are un diametru de aproximativ 5 cm. Deși aceste caracteristici depind una de cealaltă, sau, mai clar, existența fiecărei caracteristici este dependentă de existența celorlalte, toate aceste proprietăți contribuie, într-o manieră independentă, la probabilitatea ca acest fruct să fie denumit măr – și, de aceea, acest clasificator este denumit “naiv”.

Clasificatorul Naiv Bayesian este util chiar pentru seturi mari de date, și, în ciuda simplității ipotezelor pe care se bazează, este cunoscut faptul că oferă performanțe mai puternice chiar decât tehnicile de clasificare complexe, și, de aceea, a fost inclus în prezentul studiu.

Teorema lui Bayes prezintă un mod de a calcula probabilitatea posterioară $P(c/x)$ pe baza probabilității posterioare a clasei $P(c)$, a probabilității posterioare a predictorului $P(x)$ și a probabilității existenței unui predictor într-o anumită clasă dată $P(x/c)$, astfel:

$$P(c|x) = \frac{P(x|c) * P(c)}{P(x)} (3)$$

unde

$$P(c|X) = P(x_1|c) * P(x_2|c) * \dots * P(x_n|c) * P(c) (4)$$

4.3. Metoda celor mai apropiați K vecini

Metoda celor mai apropiați K vecini constituie un algoritm de clasificare ce reține toate observațiile din analiză și clasifică noile observații bazându-se, în general, pe o măsură a similitudinii, iar în majoritatea cazurilor, aceasta este o funcție de distanță. KNN a fost utilizat încă din anii 1970 în estimările statistice și în recunoașterea formelor, ca fiind o tehnică neparametrică.

Algoritmul presupune contextul că o nouă observație este clasificată pe baza votului majorității celor mai apropiați K vecini ai săi. O nouă instanță este asignată clasei celei mai întâlnite între vecinii acesteia stabiliți pe baza funcției de distanță. Cele mai utilizate funcții de distanță sunt: distanța euclidiană, distanța Manhattan și distanța Minkowski, pentru variabilele continue, iar pentru variabilele categoricale, cea mai potrivită distanță este distanța Hamming.

Decizia pentru stabilirea valorii optime pentru K este recomandată a fi luată prin inspectarea seriei de date. În general, o valoare ridicată a lui K conduce la rezultate mai precise, întrucât reduce zgomotele. Procedura de validare încrucișată reprezintă o altă metodă prin care se decide asupra valorii lui K. Pe baza studiilor empirice deja efectuate, pentru cele mai multe seturi de date, o valoare a lui K între 3 și 10 produce rezultate satisfăcătoare, și oricum, mai performante decât în cazul metodei 1NN.

4.4. Simularea de tip Monte-Carlo

Metoda de simulare Monte-Carlo este utilizată în vaste arii ale științei pentru a evalua din punct de vedere numeric așteptările cu privire la evoluția unei variabile, a unui indice al cărei comportament se recunoaște a avea caracter aleator. În acest caz, oricum, nicio funcție analitică nu poate descrie evoluția acestuia, iar decizia optimă rămâne aceea de a genera eșantioane aleatoare care să descrie cursul evoluției acestor variabile. Acuratețea estimărilor obținute prin metoda de simulare Monte-Carlo este invers proporțională cu numărul de extrageri. În studiul curent, tipul de simulare utilizată este denumită "time driven", în sensul că, pentru perioada analizată am construit scenariile unor evenimente ce pot conduce la fluctuațiile variabilelor în analiză.

Dacă un jucător de pe piața bursieră utilizează simularea de tip Monte-Carlo pe o perioadă trecută, pentru care cursul indicelui în analiză este cunoscut, acesta va conștientiza vastele traiectorii pe care acest indice le-ar fi putut urma, și deci, magnitudinea riscului pe care și l-ar fi asumat alegând o anumită strategie discretă pentru a obține profit. În strategiile viitoare, acesta va lua în considerare, cu siguranță, atât premiul la risc, cât și magnitudinea unei potențiale pierderi. Pierderea trebuie luată în calcul, și deține un rol crucial în calcularea profitului așteptat, pentru că, vizualizând numeroasele traiectorii pe care le poate urma un indice urmărit, se poate, evident, deduce probabilitatea ridicată a pierderii monetare adusă de acestea. Oricum, cu cât jucătorii sunt mai implicați în jocurile de pe piața de capital, cu atât riscul implicat este mai mare, iar importanța utilizării metodei Monte-Carlo devine un imperativ categoric. Simularea de tip Monte-Carlo este chiar superioară analizei de tip "What If", deoarece, în multe dintre cazuri, este foarte dificil a identifica sau a testa factorii determinanți fluctuațiilor unui indice de pe piața bursieră. Mai mult, în luarea unei decizii, este deosebit de importantă includerea unei vizualizări grafice a diverselor scenarii, pentru ca decidentul să devină conștient de probabilitatea asociată apariției fiecărei stări a naturii.

4.5. Mașini cu suport vectorial

Mașinile cu suport vectorial reprezintă un algoritm de învățare supervizată care poate fi folosit atât în regresii, dar mai ales, pentru clasificare. În cadrul acestei metode, fiecare observație este reprezentată într-un spațiu n-dimensional (unde n este numărul de stări ale variabilei categoricale), valoarea fiecărei stări fiind reprezentată pe câte o coordonată. Clasificarea se efectuează prin determinarea hiperplanului care separă aceste clase, în cazul analizei curente, este vorba, mai exact de separarea a două clase.

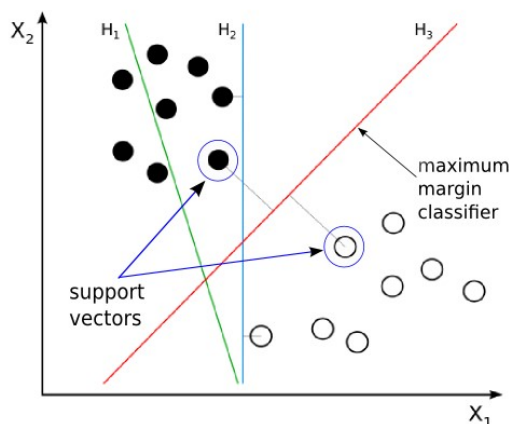


Fig. 1 – Reprezentare mașini cu suport vectorial

Sursă: <http://scikit-learn.org/stable/modules/svm.html>

Vectorii suport reprezintă, pur și simplu, coordonatele unei observații individuale, iar mașina cu suport vectorial constituie hiperplanul, frontiera care segregă cel mai bine cele două clase.

Pentru a stabili corect hiperplanul separator, trebuie să se decidă asupra formei nucleului. Pentru cele mai multe seturi de date, acesta se setează ca fiind liniar, deoarece este, în general, foarte probabil în contextul seriilor financiare a se găsi un plan separator liniar. De asemenea, coeficientul gamma reprezintă coeficientul asociat nucleului, astfel încât, cu cât acesta are o valoare mai ridicată, cu atât algoritmul va încerca să clasifice o nouă observație numai dacă aceasta prezintă o potrivire, o identificare aproape perfectă cu o anumită clasă. Dezavantajul unei valori mult prea ridicate a coeficientului gamma, este acela că unele observații nu vor putea fi clasificate.

5. Rezultate empirice

5.1. Modelarea în sens ARIMA și sezonalitatea

Modelarea evoluției indicilor bursieri ca un proces ARIMA este potrivită, întrucât, în general, în seriile de timp financiare, se regăsesc trenduri, sezonality, erori, șocuri, toți acești factori fiind luați în considerare într-o modelare de acest tip. Pentru efectuarea procedurii, am utilizat seriile de date staționarizate. Pentru perioada analizată, am inclus date lunare ale indicilor BET și BET Plus, pentru a putea surprinde fenomenul de sezonality, împreună cu influența asupra evoluției seriilor de timp. Mai mult decât atât, am considerat important să adăugăm variabile dummy corespunzătoare fiecărei luni a anului în analiză pentru a testa dacă acestea sunt semnificative din punct de vedere statistic. Deși, în principiu, trebuie să includem 12 factori sezonieri, am ales să excludem luna decembrie, deoarece, incluzând și termenul liber, s-ar fi produs capcana dummy.

Scopul final al analizei este acela de a previziona pe baza modelului ales valorile medii ale indicilor pe următoarele două luni.

Cel mai potrivit model a fost ales pe baza criteriului informațional Akaike, dintre cele 100 modele candidate estimate. Valoarea minimă (13.6714) corespunde celui mai performant model. Pe baza acestuia, valorile actuale și valorile previzionate sunt prezentate în tabelul de mai jos:

Tabelul 1

Valori reale vs. Valori previzionate

Luna	Valoarea medie reală	Valoarea medie previzionată
Octombrie	7946.01	7868.38
Noiembrie	7783.36	7955.372

Cum aceste valori diferă semnificativ de cele reale, și, mai mult, dacă în realitate de la luna octombrie la luna noiembrie, indicele BET prezintă, în medie, o scădere, iar prin modelare ARIMA, acesta a fost previzionat că va crește cu aproape 100 unitați, putem concluziona cu siguranță faptul că, nici măcar luând în considerare prezența factorilor sezonieri în analiză, evoluția Bursei de Valori București nu poate fi previzionată, urmând, deci, o evoluție de tip random walk.

5.2. Corelația Bursei de Valori București cu bursele de valori internaționale

Mai departe, considerăm important pentru studiul întreprins intensitatea legăturii dintre evoluția principalilor indici ai Bursei de Valori București și evoluția piețelor internaționale de capital. Pentru aceasta, este relevantă observarea fluctuațiilor indicelui bursier american S&P500, luând în considerare același orizont de timp. Standardizând datele, am obținut o corelație foarte puternică de 91.67% cu indicele BET și 92.29% cu BET Plus.

Inițial neluând în considerare factorii sezonieri, am încercat previziunea indicilor BET și BET Plus, pe baza propriei valori din ziua precedentă (luând în considerare impactul, dependența pe termen scurt, deci conceptul numit *volatility clustering*) și valorii din ziua precedentă a indicelui S&P500, în ipoteza că, incluzând o întârziere scurtă în timp, indicii studiați urmează, totuși, în mod fidel, evoluția principalelor burse de valori de pe plan internațional.

Prezentăm modelele estimate împreună cu cele mai importante proprietăți ale acestora mai jos:

$$DBET_t = 3.08 + 0.26 * DSP_{t-1} + 0.99 * DBET_{t-1} + \varepsilon_t$$

$$DBETPLUS_t = 1.27 + 0.26 * DSP_{t-1} + 0.99 * DBET_{t-1} + \varepsilon_t$$

Tabelul 2

Proprietățile statistice ale modelelor

Indice previzionat	Root Mean Square Error	F-statistic	Prob. asociată testului Fischer-Snedecor
BET	0.107	29463.05	0.000
BET Plus	0.108	32068.32	0.000

Deși modelele estimate sunt, evident, semnificative din punct de vedere statistic, previziunea celor doi indici pe baza acestora nu a fost performantă. Este evident că, utilizând acest tip de analiză tehnică a pieței, nu se poate obține supraprofit.

Se observă că, în foarte puține zile din perioada considerată valorile previzionate corespund cu valorile reale. Mai mult decât atât, în cazul ambilor indici, a fost validat în previziune trendul crescător al evoluției acestora de la începutul perioadei studiate până la începutul trimestrului trei al anului 2017. În rest, nici magnitudinea, dar nici sensul fluctuațiilor nu a fost previzionat corect.

Totuși, în scopul de a îmbunătăți această analiză, am admis, în mod explicit, existența trendului și am inclus în cele două modele efectul factorilor sezonieri pentru a încerca să explicăm precis fluctuațiile seriilor de date. Am exclus din analiză luna decembrie ca și factor de sezonialitate pentru a evita capcana dummy. Modelele obținute sunt următoarele:

$$DBET_t = 7.47 + 0.06 * DSP_{t-1} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBETPLUS_t = 3.49 + 0.04 * DSP_{t-1} + 0.08 * TR_t + 0.02 * mth1 + 0.01 * mth2 + 0.05 * mth3 + 0.09 * mth4 + 0.06 * mth5 + 0.01 * mth6 + 0.03 * mth7 + 0.02 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

Ambele modele sunt semnificative din punct de vedere statistic, însă, ceea ce este mai relevant în analiză constituie faptul că estimarea modelului a confirmat efectul semnificativ al factorilor sezonieri. Cu toate acestea, pe baza probabilității asociate testului t-Student, se constată faptul că efectele sezoniere ale lunii octombrie și noiembrie nu sunt semnificative. Pe baza acestor modele, rezultatele previziunii sunt reprezentate grafic în Anexa 2.

Graficele prezentate indică, cu siguranță, o calitate mult mai ridicată a previziunii, datorată, cu siguranță, includerii factorilor sezonieri, dar și a faptului că metoda aleasă este cea

statică, ceea ce înseamnă că valoarea de previzionat la o anumită iterație se bazează numai pe valorile reale trecute ale indicelui, și nu și pe valorile previzionate anterior. Astfel, în manieră cantitativă, concluzionăm că valorile previzionate ale indicelui BET în perioada analizată corespund în proporție de 96.64% cu valorile reale, iar valorile previzionate ale indicelui BET Plus sunt corelate cu valorile reale în proporție mai mare – de 96.91%.

Din rezultatele obținute, evoluția marilor jucători pe plan internațional, este cu siguranță, un determinant important în ceea ce privește evoluția Bursei de Valori București și dorim, în mod natural într-o analiză mai amănunțită să testăm dacă indicii BET și BET Plus sunt influențați de evoluția trecută pe termen scurt sau pe termen lung a indicelui de piață american S&P500. În acest raționament, am îmbunătățit modelul pentru a lua în calcul valorile pe ultimele 5 zile ale indicelui S&P500, obținând, astfel, următoarea estimatie:

$$DBET_t = 8.92 + 0.04 * DSP_{t-1} + 0.05 * DSP_{t-2} + 0.04 * DSP_{t-3} - 0.02 * DSP_{t-4} + 0.12 * DSP_{t-5} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBETPLUS_t = 7.18 + 0.03 * DSP_{t-1} + 0.04 * DSP_{t-2} + 0.04 * DSP_{t-3} - 0.02 * DSP_{t-4} + 0.11 * DSP_{t-5} + 0.01 * TR_t + 0.02 * mth1 + 0.06 * mth2 + 0.09 * mth3 + 0.09 * mth4 + 0.07 * mth5 + 0.05 * mth6 + 0.03 * mth7 + 0.04 * mth8 + 0.02 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

În cadrul acestui demers, observăm, pe baza testului Wald, faptul că acești coeficienți nou-introduși nu sunt semnificativ diferiți de 0 pentru ambele serii de timp. Valorile reale și valorile previzionate sunt ilustrate în Anexa 6.

Pentru indicele BET, coeficientul de corelație între valorile reale și valorile previzionate este de 96.57%, iar pentru indicele BET Plus, acesta are valoarea de 96.85%. În ambele cazuri, previziunile bazate pe ultima valoare din trecut sunt mai performante decât luarea în considerare a unei perioade mai lungi. Acest comportament se datorează, probabil, caracterului dinamic, în continuă schimbare a indicilor bursieri, a jucătorilor de pe piață, așa că, dacă unele șocuri apar, ele dispar rapid și nu are sens includerea mai multor zile din trecut în previzionarea indicilor din analiză.

În sfârșit, observând în mod evident că cel mai important determinant în previziune a reprezentat includerea fenomenului de sezonalitate, am adăugat modelului ce include factorii sezonieri, pe rând, un termen de tip medie mobilă, un termen autoregresiv și, în final, un termen AR și un termen MA simultan. Pe baza modelului inițial și a celor trei modele suplimentare, am realizat o *medie a predicțiilor* evoluției indicelui BET. Modelele obținute sunt semnificative din punct de vedere statistic, considerând validitatea testului Fischer-Snedecor, și iau următoarea formă:

$$DBET_t = 3.31 + 0.98 * DBET_{t-1} + 0.25 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 * mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 + 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 + 0.01 * mth10 + 0.01 * mth11 + \varepsilon_t$$

$$DBET_t = 6.43 + 0.85 * \varepsilon_{t-1} + 0.11 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 * mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 + 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 - 0.01 * mth10 - 0.01 * mth11 + \varepsilon_t$$

$$DBET_t = 3.45 + 0.98 * DBET_{t-1} + 0.01 * \varepsilon_{t-1} + 0.24 * DSP_{t-1} + 0.01 * TR_t + 0.01 * mth1 + 0.01 * mth2 + 0.01 * mth3 + 0.01 * mth4 + 0.01 * mth5 + 0.01 * mth6 + 0.01 * mth7 + 0.01 * mth8 + 0.01 * mth9 + 0.01 * mth10 + 0.01 * mth11 + \varepsilon_t$$

Reprezentarea simultană a acestor serii de timp, împreună cu previziunile obținute pe baza acestora sunt prezentate în Anexa 7.

Pentru o performanță mai ridicată, a previziunilor, acestea au fost elaborate prin metoda statică, deci nu au fost luate în considerare decât valorile reale. Media predicțiilor constituie, astfel, cea mai elaborată metodologie dintre cele testate în prezentul studiu, în vederea

previzionării indicilor BET și BET Plus. Media predicțiilor, împreună cu valorile reale sunt reprezentate în Anexa 8.

Deși modelul obținut prin Forecasting Average și bazat pe cele patru submodele este cel mai performant, diferențele între valorile reale și cele previzionate sunt totuși semnificative, neputându-se obține profit semnificativ prin aceste tehnici analitice de previziune. Valorile previzionate cresc atunci când și cele reale cresc, și scad când cele reale prezintă o evoluție negativă, dar magnitudinea fluctuației diferă, uneori semnificativ.

Mai mult, după cum am demonstrat, nu este mai relevantă luarea în considerare a ultimelor 5 zile ale indicelui S&P500 decât a zilei precedente în mod exclusiv, astfel că orice previziune obținută printr-o metodă analitică este oricum valabilă pe termen scurt, urmând ca, în general, și într-un timp foarte scurt, cursul evoluției să se schimbe semnificativ față de cel așteptat de speculatori. Din nou, am demonstrat că nu există oportunități profitabile pentru analiști. Nici măcar raportându-ne la piețele internaționale de capital, mai exact la cei mai importanți determinanți de pe plan global, nu putem previziona evoluția indicilor bursieri într-un sens relevant - în care să ne aducă profit semnificativ.

Într-o analiză mai amănunțită, cum calitatea previziunii a crescut în manieră considerabilă atunci când am luat în considerare efectul sezonier, concluzionăm că, pentru o previziune mai relevantă, este important să se studieze amănunțit și să se ia în considerare toți determinanții care există în piața de capital. Chiar și în acest caz ideal, șocurile vor rămâne elemente imposibil de previzionat. Oricum costurile acestor analize, dar și costurile de tranzacționare și alte costuri care apar pe piața financiară reală, determină ca aceste analize să nu aiba sens decisiv atunci când jucătorii doresc obținerea unui profit semnificativ.

5.3.Predicții prin învățare pe baza clasificatorului naiv bayesian și a metodei celor mai apropiați K vecini

Clasificatorul naiv bayesian reprezintă o tehnică de predicție prin intermediul căreia datele ce compun setul de antrenare (70% din setul total de date) sunt convertite într-un tabel de frecvențe. Se calculează probabilitatea asociată fiecărui eveniment (a vinde sau a păstra portofoliul în aceeași structură, ca și strategie optimă) pentru ca apoi, pe baza acestora, să se construiască tabelul de probabilitate. Folosind formula pentru probabilitatea posterioară propusă de Bayes, se calculează aceste probabilități pentru toate clasele, deci, probabilitatea posterioară pentru strategia de a vinde și probabilitatea posterioară pentru strategia de a păstra aceeași structură, pe baza valorilor indicilor BET Plus și S&P500. Fiind primul indice care a descris piața bursieră, am considerat relevant pentru stabilirea strategiei evoluția indicelui BET. O strategie simplă este următoarea: dacă acesta prezintă fluctuații pozitive, jucătorul va decide să aleagă evenimentul “Buy”, iar în caz contrar, el va alege evenimentul “Hold”. În realitate, considerând clasificatorul naiv bayesian, la fiecare iterație, va fi aleasă clasa cu probabilitatea posterioară cea mai ridicată. Pe baza acestor ipoteze, observațiile din cadrul setului de testare (30% din perioada inclusă în analiză) au previzionat strategia corectă în 62% din cazuri.

```
> round(sum(preds=="stockk.validation.sample$stock.STRATEGY, na.rm=TRUE)/length(stockk.validation.sample$stock.STRATEGY), digits=2)
[1] 0.62
```

Fig. 2: Rezultat performanță algoritm

Sursă: prelucrarea autorului

Mai mult, pe baza matricei de confuzie din Anexa 9, 80 observații au fost previzionate în clasa corectă “Buy”, și numai o singură observație a fost previzionată corect în clasa “Hold”. Totuși, coeficientul lui Cohen prezintă o valoare foarte scăzută, de 2%. Acest neajuns se poate corecta, de exemplu, încercând redefinirea variabilei categoriale (deci, a strategiei), adăugarea mai multor variabile în analiză sau chiar alegerea selectivă a determinanților care, în mod sigur influențează alegerea strategiei.

În ceea ce privește metoda celor mai apropiați k vecini, am utilizat aceleași seturi de testare și de antrenare, iar pentru ambele metodologii, am setat aceeași sămânță, pentru a asigura posibilitatea comparării performanțelor acestora. Metoda KNN face parte din procedurile de recunoaștere a formelor, astfel, pe baza observațiilor din setul de antrenare, pentru care considerăm cunoscută clasa reală, am calculat distanța euclidiană între observațiile din setul de

testare și cele din setul de antrenare. Stabilind inițial $k=3$, o nouă observație este clasificată pe baza simplei majorități a claselor celor mai apropiați trei vecini. Pentru $k=10$, am procedat în mod analog, iar rezultatele sunt furnizate în Anexa 10.

Cu siguranță, în ambele variante ale aplicării algoritmului KNN, acesta este mai performant decât utilizarea clasificatorului naiv bayesian. Deși acuratețea este mai mică, este important că valoarea coeficientului lui Cohen crește considerabil, astfel că, pe clase, raportul între valorile previzionate corect și toate valorile aparținând respectivei clase este echilibrat.

Cu toate acestea, se observă că luând în considerare numai trei vecini, calitatea algoritmului a fost mai ridicată decât alegând zece vecini. Acest lucru se poate datora, de exemplu, faptului că criteriul alegerii clasei pentru o observație neclasificată prin majoritatea simplă determină ca, alegerea unui număr impar de cei mai apropiați vecini să devină, cu siguranță, o decizie mai bună.

5.4. Simularea comportamentală a Bursei de Valori București prin Metoda Monte Carlo

Pentru stabilirea unei strategii optime pe piața de capital, un jucător trebuie să ia în calcul toate posibilitățile de traiectorii pe care le poate urma un indice bursier urmărit. În analiza de față, am ales urmărirea potențialelor fluctuații ale indicelui BET, fiind primul indice al bursei de valori românească, ce a evidențiat încă de la începutul BVB macrotendința celor mai lichide și active zece societăți tranzacționate. Am considerat important nu numai observarea valorilor trecute ale acestuia, ci, mai ales posibilitatea vastă a traiectoriilor pe care le-ar fi putut urma în contextul unor stări ale naturii potențiale.

Vizualizându-le în mod interactiv în Anexa 11, un decident al cărui scop este obținerea de profit de pe urma jocurilor de pe piața de capital, poate lua decizii într-o manieră rațională, deoarece în acest mod, va conștientiza probabilitățile asumării că un anumit eveniment, din spațiul larg al evenimentelor, va avea loc și îi va aduce câștig. Este crucială o astfel de presupunere, deoarece, într-un sistem complex, dinamic, așa cum este bursa de valori românească, dacă un alt eveniment va avea loc, acesta nu va aduce valoarea monetată așteptată egală cu cea rezultată din presupunerea făcută, sau, într-un caz pesimist, poate produce chiar pierderi semnificative.

În realitate, jucătorul trebuie să își propună o strategie, cuantificând, odată cu luarea deciziei, și efectele unei posibile erori în care scenariul nu va avea loc, prin luarea în considerare a posibilităților de apariție și a valorii monetare așteptate asociate a tuturor stărilor naturii potențial existente în piața de capital, la un moment dat.

5.5. Învățarea și predicția pe baza mașinilor cu suport vectorial

În ceea ce privește predicția pe baza mașinilor cu suport vectorial, am divizat întreg setul de date în două subseturi, după cum urmează: o treime din acesta reprezintă setul de testare, iar două treimi constituie setul de antrenare. Pentru antrenarea mașinii cu suport vectorial, am considerat nucleul în formă liniară, deoarece, în cvasitotalitatea cazurilor, această tipologie este potrivită pentru seriile de date financiare.

Pentru întreg setul de date, a fost adăugată o variabilă categorială *UPDOWN*, pentru care atunci când indicele BET prezintă fluctuații pozitive față de ziua precedentă, variabila ia valoarea *Up*, iar când se observă scăderi, variabila ia valoarea *Down*. Pe baza setului de antrenare, mașina cu suport vectorial a fost antrenată pentru a putea face predicții cu privire la sensul fluctuațiilor indicelui BET de la o zi la alta, pentru perioada inclusă în setul de testare. Ulterior, au fost comparate rezultatele obținute pe baza algoritmului de învățare supervizată cu fluctuațiile din realitate.

Rezultatele predicției, ilustrate în Anexa 12, coincid în proporție de 53.79% cu rezultatele reale, pentru cele 145 observații incluse în setul de testare. Deși algoritmul a efectuat predicții corecte în mai mult de jumătate din totalitatea cazurilor, acesta nu oferă garanție cu privire la existența unor oportunități profitabile.

Oricum, scopul acestei analize a fost de a proba posibilitatea de predicție a sensului fluctuațiilor indicelui bursier BET, și în niciun caz a magnitudinii acestor fluctuații. Această

deosebire trebuie semnalată pentru că un decident nu poate calcula valoarea monetară așteptată corespunzătoare luării unei decizii cu privire la o anumită strategie, chiar și dacă acesta ar ști cu siguranță dacă în ziua imediat următoare indicele va prezenta o evoluție pozitivă sau nu. Jucătorii de pe piața de capital ce doresc să obțină profit pe urma acestor strategii dețin, de regulă, un portofoliu vast de acțiuni, ce poate cuprinde atât comportamente de evoluție în sensul pieței cât și comportamente de evoluție în sens opus al acesteia, astfel că magnitudinea unei schimbări în cursul evoluției indicelui bursier BET este de importanță categorică. Putem concluziona deci faptul că predicțiile corecte în proporție de 53.79% ale sensului traiectoriei indicelui bursier BET nu aduce supraprofit pentru jucătorii din piață.

6. Concluzie

Provocarea pe care o lansează validitatea teoriei mersului la întâmplare pe piața de capital din România, demonstrată, de altfel, în manieră analitică în prezenta lucrare, către jucătorii de pe piețele financiare este conturată de contextul: dacă piețele sunt eficiente, atunci prețurile acțiunilor la orice moment de timp vor reprezenta chiar estimarea consistentă a valorii sale intrinseci. În acest sens, analiza fundamentală este utilă numai în condițiile în care analistul deține informație nouă, ce nu este încă disponibilă pe piață, deci nu a fost considerată în formarea prețurilor curente. Dacă analistul nu deține informație nouă sau contexte încă neexploatate în piață, atunci decizia optimă a sa ar trebui să fie alegerea acțiunilor în compunerea unui portofoliu sau în tranzacție printr-o procedură pur aleatoare.

În esență, testele efectuate nu au fost capabile să respingă ipoteza de descriere a evoluției prețurilor pe piața de capital sub forma unui proces de mers la întâmplare.

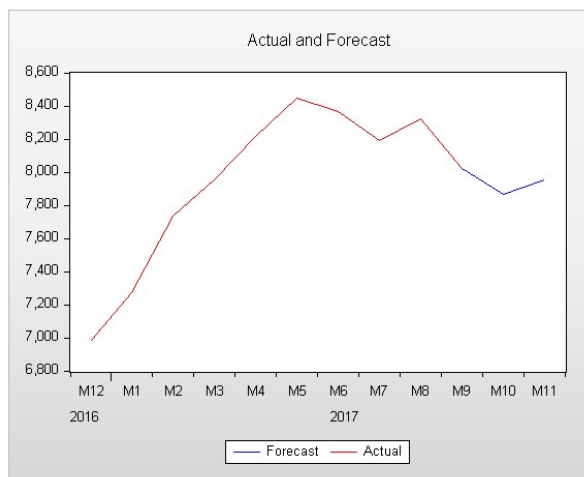
7. Recomandări pentru studii suplimentare

- un studiu al procesului de mers aleator poate fi întreprins în contextul EMH folosind date aparținând, de exemplu, Bursei de Valori Sibiu, în vederea suplimentării cadrului în care se desfășoară acțiunile pieței bursiere românești;
- același studiu poate fi elaborat utilizând alte instrumente de analiză pentru a îmbunătăți rezultatele și concluziile obținute în această lucrare;
- perioada analizată poate fi extinsă în vederea obținerii unor rezultate consistente.
-

Bibliografie

1. Wooldridge, J., „*Introductory Econometrics, A Modern Approach*”, Pearson, 2009
2. www.bvb.ro

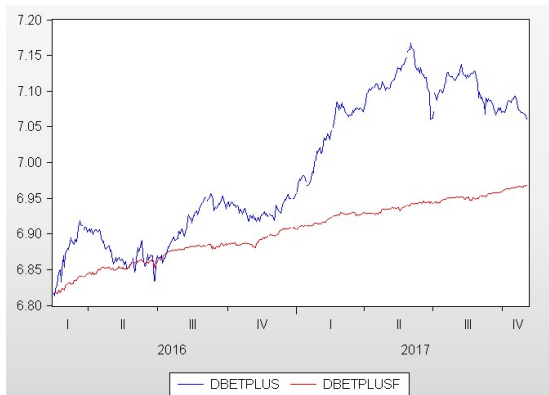
ANEXA 1 – PREVIZIUNI PE BAZA MODELĂRII ARIMA



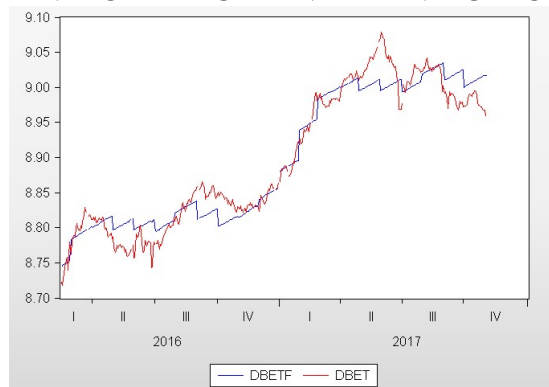
ANEXA 2 – GRAFICUL PREVIZIONĂRII INDICELUI BET PE BAZA DEPENDENȚEI DE VALOAREA DIN ZIUA PRECEDENTĂ ȘI VALORILE INDICELUI S&P500



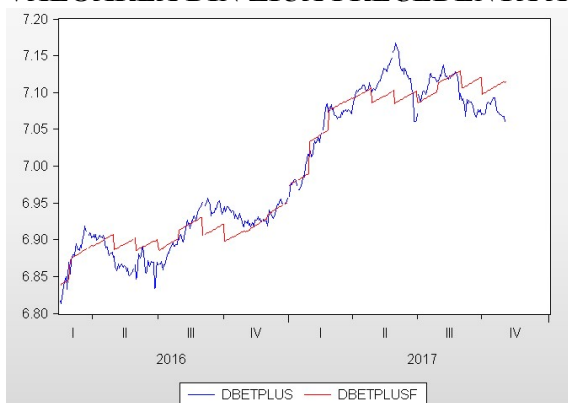
ANEXA 3 – GRAFICUL PREVIZIONĂRII INDICELUI BET PLUS



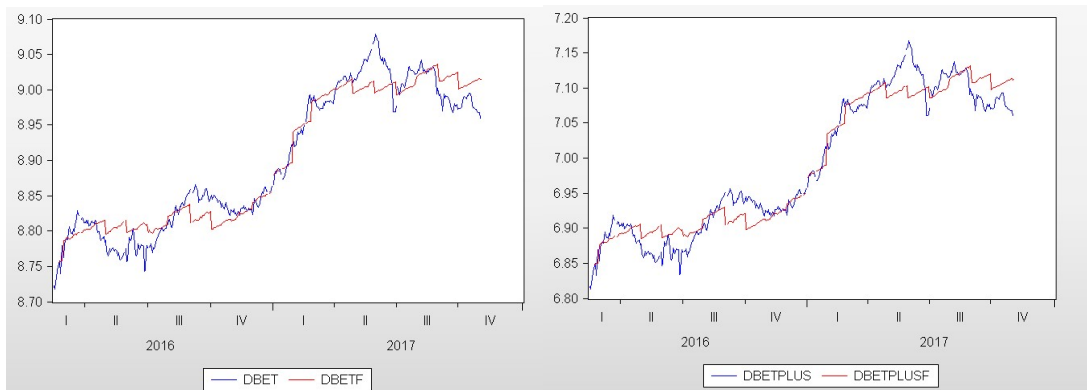
ANEXA 4– PREVIZIUNEA BET INCLUZÂND FACTORII SEZONIERI ȘI VALOAREA DIN ZIUA PRECEDENTĂ A INDICELUI S&P500



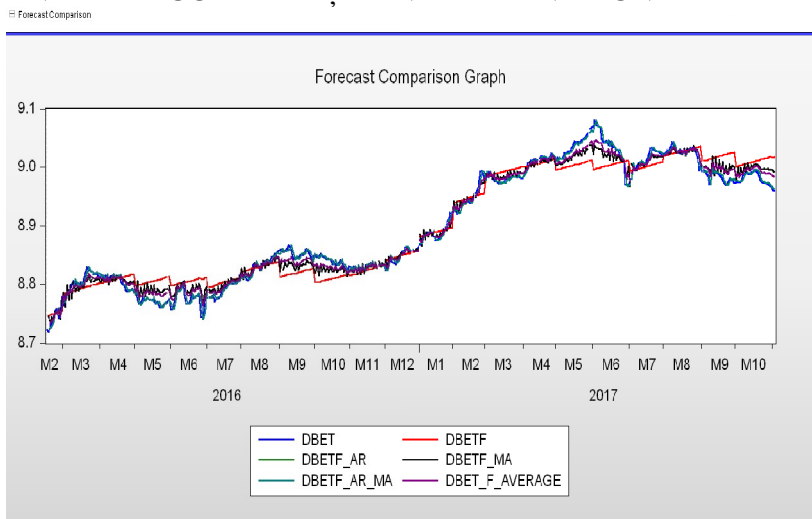
ANEXA 5 – PREVIZIUNEA BET PLUS INCLUZÂND FACTORII SEZONIERI ȘI VALOAREA DIN ZIUA PRECEDENTĂ A INDICELUI S&P500



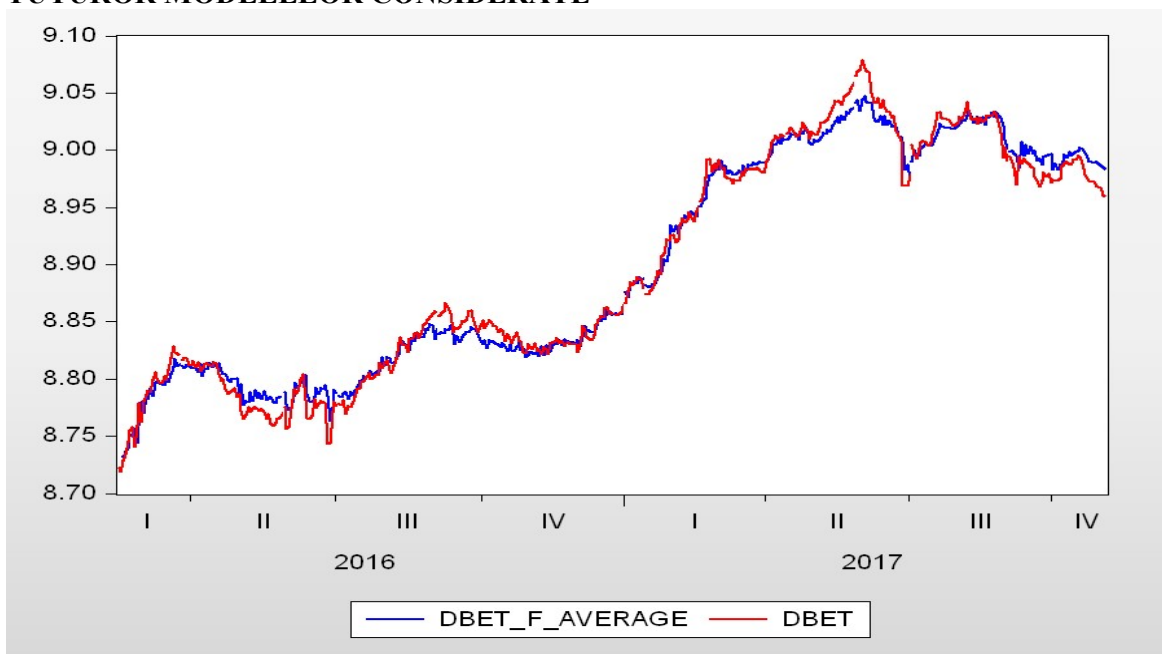
ANEXA 6 – PREVIZIUNEA INDICILOR BET ȘI BET PLUS PE BAZA VALORILOR S&P500 DIN ULTIMELE 5 ZILE



ANEXA 7 – COMPARAȚIE ÎNTRE PREVIZIUNI



ANEXA 8 – VALORI REALE VS MEDIA VALORILOR PREVZIONATE PE BAZA TUTUROR MODELELOR CONSIDERATE



ANEXA 9 – REZULTATELE PERFORMANȚEI ALGORITMULUI PE BAZA CLASIFICATORULUI NAIV BAYESIAN

```
Confusion Matrix and Statistics

precs Buy Hold
Buy 80 50
Hold 0 1

          Accuracy : 0.6183
          95% CI : (0.5294, 0.7018)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : 0.4669

          Kappa : 0.0238
  Mcnemar's Test P-Value : 4.219e-12

    Sensitivity : 1.00000
    Specificity : 0.01961
    Pos Pred Value : 0.61538
    Neg Pred Value : 1.00000
    Prevalence : 0.61069
    Detection Rate : 0.61069
    Detection Prevalence : 0.99237
    Balanced Accuracy : 0.50980

'Positive' Class : Buy
```

ANEXA 10 - PERFORMANȚELE ALGORITMULUI KNN (K=3, K=10)

```
Confusion Matrix and Statistics

precs Buy Hold
Buy 42 19
Hold 38 32

          Accuracy : 0.5649
          95% CI : (0.4755, 0.6512)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : 0.87753

          Kappa : 0.1428
  Mcnemar's Test P-Value : 0.01712

    Sensitivity : 0.5250
    Specificity : 0.6275
    Pos Pred Value : 0.6885
    Neg Pred Value : 0.4571
    Prevalence : 0.6107
    Detection Rate : 0.3206
    Detection Prevalence : 0.4656
    Balanced Accuracy : 0.5762

'Positive' Class : Buy
```

```
Confusion Matrix and Statistics

precs Buy Hold
Buy 44 24
Hold 36 27

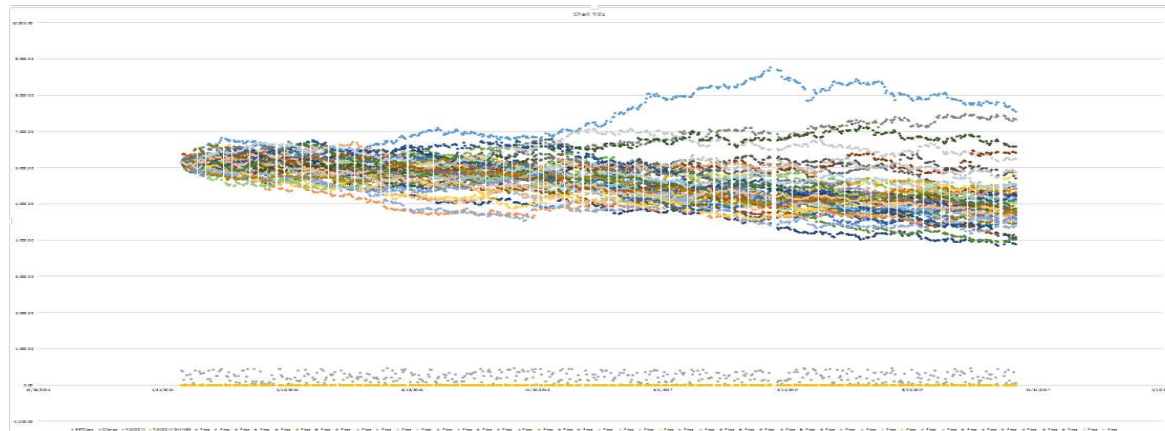
          Accuracy : 0.542
          95% CI : (0.4527, 0.6293)
    No Information Rate : 0.6107
    P-Value [Acc > NIR] : 0.9547

          Kappa : 0.0762
  Mcnemar's Test P-Value : 0.1556

    Sensitivity : 0.5500
    Specificity : 0.5294
    Pos Pred Value : 0.6471
    Neg Pred Value : 0.4286
    Prevalence : 0.6107
    Detection Rate : 0.3359
    Detection Prevalence : 0.5191
    Balanced Accuracy : 0.5397

'Positive' Class : Buy
```

ANEXA 11 – SIMULAREA DE TIP MONTE-CARLO A FLUCTUAȚIILOR BVB ÎN PERIOADA CONSIDERATĂ



ANEXA 12 – EVALUAREA PERFORMANTELOR ALGORITMULUI VSM (MAȘINI CU SUPT VECTORIAL)

```
[1] 145
```

```
> na.omit(settestare[,2])
```

```
 [1] Up   Up   Up   Up   Down Up   Down Down Up   Down Down Down Up   Up   Down
[16] Up   Down Up   Down Up   Up   Up   Up   Down Up   Up   Down Down Up   Up
[31] Down Up   Up   Up   Up   Down Down Up   Down Up   Down Down Up   Down Down
[46] Up   Up   Down Up   Up   Up   Down Up   Down Down Down Up   Down Down Up
[61] Up   Up   Down Down Up   Up   Up   Down Down Down Up   Down Down Down Down
[76] Down Down Down Up   Up   Up   Down Up   Up   Down Up   Down Down Down Up
[91] Up   Up   Down Up   Down Up   Down Down Up   Up   Up   Up   Up   Up   Up
[106] Down Down Up   Down Up   Up   Up   Up   Up   Up   Up   Down Down Down Up   Down
[121] Down Up   Down Up   Up   Down Down Up   Down Up   Up   Down Down Up   Down
[136] Up   Up   Down Down Down Up   Up   Up   Up   Up   Down
```

```
Levels: Down Up
```

```
> classAgreement(confuzie)
```

```
$diag
```

```
[1] 0.537931
```