15/03/2018

<u>מבחן בית - סדנת מומחים</u>

88-654 פרופ' ראובן כהן ופרופ' ג'רמי שיף

סטודנט: יובל סמואלוב ת.ז. 204025258

<u>רקע</u>

במסגרת קורס סדנת מומחים, למדנו כיצד ניתן לנתח נתוני מסחר של מניות בצורה מתקדמת יותר, וע"ב הניתוחים, ניתן לקבל החלטות או לקבוע אסטרטגיות מסחר מושכלות יותר. להלן פתרון למבחן בית לשאלות הבאות:

1. <u>מקדמי מתאם עצמיים של מניית X:</u> (Auto-Correlation Coefficients)

✓ בשאלה נתוני מסחר בבורסת NYSE ל-8 מניות[1] למשך כל דקת מסחר במשך השנים 2011-2013.

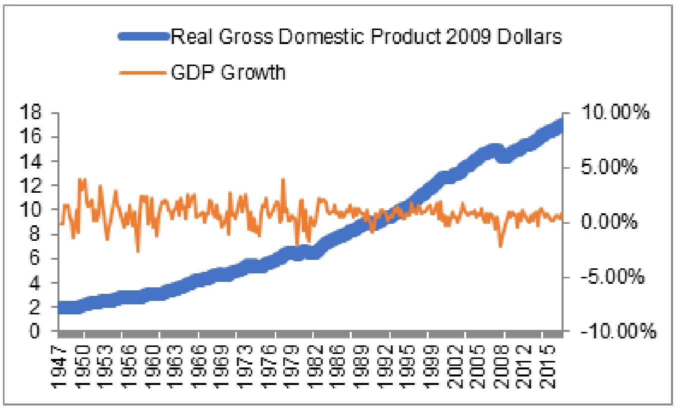
✓ נתוני מסחר בכל דקה, מ-30 :09 ועד 00 :16 בנטרול חגים[2] (בעלי השפעה על מסחר), עבור מניית FORD, לאור נפח מסחר ממוצע מקסימלי לדקה של המניה, ביחס לאחרות שניתנו.

✓ **הבעיה:**

א. היעדר סטציונריות וקיומו של מתאם סדרתי עם נתוני העבר.

(1) **מתאם סדרתי** - לאור אופי המסחר (שער סגירה של יום X מהווה שער פתיחה של יום X+1) ברור שיש פיגור בנתונים, קרי המשתנה המסביר לתשואת המניה היום הוא תשואת המניה אתמול. לכן, לא נכון לאמוד רגרסיה לינארית רגילה לנתוני השאלה. השאלה הנשאלת אפוא, היא **כמה התקופות יש ללכת אחורה (Lags) בכדי להסביר את תשואת המניה.**

(2) **סטציונריות** – להלן דוגמא פשוטה מפרמטרים מאקרו כלכליים, בכדי להסביר את המורכבות שבדבר. שליפה מאתר ה-FED, של GDP – תמ"ג (תוצר המקומי הגולמי) וצמיחה בארה"ב.



כפי שניתן לראות, התוצר (כחול) **אינו סטציונרי**. התוחלת שלו משתנה עם הזמן, ואינה מתכנסת סביב ממוצע קבוע. לעומת זאת, הצמיחה (כתום) מאוד תנודתית, אך מתכנסת סביב ממוצע קבוע בטווח הארוך (בקנה מידה זה, לצורך המחשה בלבד). המשמעות, כידוע בכלכלה, היא ש**צמיחה** לרוב **סטציונרית**, בעוד כי **התוצר** מתנהג כ**לא סטציונרי** בטווח הארוך. משפט גאוס מרקוב (מוכיח כי אומדי הריבועים הפחותים הם הכי יעילים - נדרש לרגרסיה לינארית) אינו מתקיים, ונדרשת העמקה ובחינת סדר המתאם הסדרתי שבנתונים, לצורך ביטול תופעה זו.

✓ **הפתרון:**

א. אנו נשתמש בפונקציית Python המחשבת AR. ונצפה גם בגרף המציג את דרגת הפיגור – Lag כפונקציה של המתאם), Correlogram, מול מובהקות של 95% כפי שנלמד (2 חלקי שורש גודל המדגם)

---

[1] Conoco Philips, Chevron, **Ford**, General Motors, Coca-Cola, Occidental Petroleum, Pepsi, ExxonMobil

[2] Chrisms 'eve, Thanksgiving, Independence-Day

ב. <u>חישוב מקדם המתאם הפנימי :</u>

להלן התוצאות עבור התרחישים המבוקשים :

אופציה	צורת חישוב	תוצאה	דרגת פיגור מובהקת	רמת מובהקות
א	תשואות יומיות (3 שנים)	אין מתאם	**ללא**	**95%**
ב'	**ריבועי תשואות יומיות (3 שנים)**	**יש מתאם**	**1,2,3,7**	
ג'	**תשואה לדקה (2011, 2013)**	**יש מתאם**	**3,4,6,7,12,13, 15, 18**	
ד'	**ריבועי תשואה לדקה (2011, 2013)**	**יש מתאם**	**כל הדרגות עד 20 כולל**	

כל הממצאים בפירוט נרחב, בנספח א'.

ג. <u>תוצאות בעלות משמעות סטטיסטית :</u>

בדומה להרצאות, ניכר כי **ריבועי התשואות אכן מתואמות פנימית**. ניתן להשתמש בתנועת העבר (בכפוף לתוצאות שהתקבלו) כמשתנה מסביר לאמידת התשואה העתידית. כלומר, **ברמת מובהקות של 95% יש קשר מובהק בין ריבועי התשואה היום, לריבועי התשואה אתמול.**

לדוגמא, ב-3 שנים האחרונות, ניתן לקחת כמשתנה מסביר את ריבועי התשואות היומיות ל-1,2,3 ו-7 ימים אחורה.

ד. <u>יחס בין מקדמי ההתאמה לתשואות :</u>

ע"ב הנתונים, היחס שביניהם הוא נמוך מאוד (שלילי) = 0.087

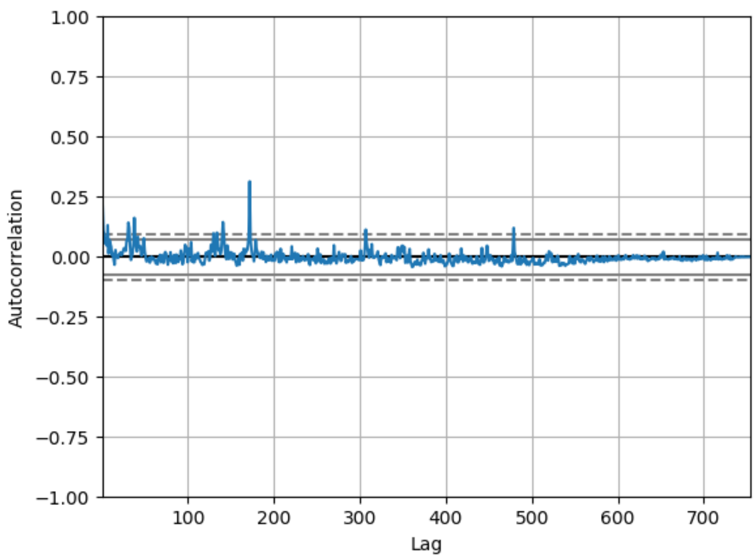
ניתן להסיק, כי ברמת המובהקות המבוקשת, תנועה יומית בלתי תלויה בתנועה בכל דקה ודקה.

✓ **לסיכום:**

ניתן לראות שריבועי תשואות נתוני המסחר, מתנהגות כסדרות עיתיות (Time Series), ועל כן בבואנו לנתחם יש להתייחס אליהם בהתאם למודל הסטטיסטים המתאימים (GARCH). קיים קושי לאמוד ולבצע תחזיות למניות, לאור התנודתיות הגדולה שבמסחר בכל דקה.

לאור הממצאים מעלה, אציג את הנתון המעניין מהניתוח לעיל.

להלן גרף הקורלציות הפנימיות (Correlogram) של ריבועי תשואות היומיות של המנייה (קו כחול) :



ציר Y – גובה הקורלציה.

ציר X – דרגת הפיגור (מס' תקופות אחורה שיש לקחת שמשתנה מסביר

קו שחור – ספ החלטה ברמת מובהקות 95%. (2/(N)^0.5)

מבחן סטטיסטי : למעשה, ניתן לראות את האנומליות בגרף. נקרא לכל נק' i אנומלית אם עבורה הגרף הכחול מעל ספ ההחלטה – (הקו השחור) אשר הן המבוקשות בשאלה זו. ניכר כי עבור דרגות פיגור מסוימות נוספות, מעבר ל20 התקופות שהתבקשנו לבחון, קיימת מובהקות וניתן לקחת כמשתנה מסביר לא רק את ריבועי התשואות של תקופות ואחורה i=1,2,3,7) כמפורט לעיל), אלא כל i. למעשה, מתוך חצי השנה האחרונה של המסחר (גרוסו-מודו) ניתן לזהות תקופות אשר תורמות סטטיסטית לחיזוי המניה.

2. <u>זיהוי אנומליות באמצעות K-Means</u>: (אלגוריתם ללמידה לא מפוקחת Unsupervised Learning)

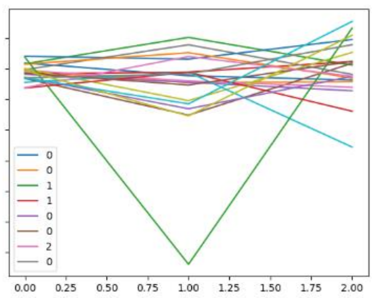
<u>רקע</u>

- האלגוריתם הוא ממשפחות אלגוריתמים **ללמידת המכונה.**
- באופן כללי, המטרה היא ע״ב סט אימון X (x_train) וסט תגיות (y_train) לאפשר לאלגוריתם ללמוד/לאפיין מתוך סט היפותזות (H), פונקציה (fø(x) = y^) שתאפשר לחזות אומדן אופטימלי לתיוג y, ע״י מזעור השגיאה/פונקציית ההפסד שהוא הפער בין האומד החזוי לאומד האמיתי (loss(y,y^ לנתונים שהאלגוריתם לא נחשף אליהם.
- **בפועל, האלגוריתם מחפש ø כך שפונקציית ההפסד תגיע למינימום ((loss(y, fø(x**
  כאשר y התיוג האמיתי (label) , ו(fø(x זו התחזית לתיוג.
- **K-means** מסווג/מתייג את המספרים לפי k מרכזים בהתאם לבחירת המשתמש.
  א. ראשית, הוא ממקם ניחוש על הגרף, ובודק את המרחקים האוקלידיים (עבור דו ממדי) בין הנק׳ שנבחרה לכל הנקודות.
  ב. לאחר מכן, האלגוריתם משפר מיקום במידה וניתן למזער את המרחקים הממוצעים מכל הנקודות במס׳ איטרציות, כך שככל שנמצא מרחק קרוב יותר, הוא משפר את מיקומו.
  ג. בסופו של דבר, האלגוריתם מצליח למצוא k מרכזים שונים לנתונים.
- הנתונים כוקטור התשואות היומיות מגודל 8*20, (8מניות עבור 20 ימים 01/2011)עליהם ייושם האלגוריתם.[3]
  מניות : ConocoPhilips,Chevron,Ford,GeneralMotors,CocaCola,OccidentalPetroleum,Pepsi,ExxonMobile

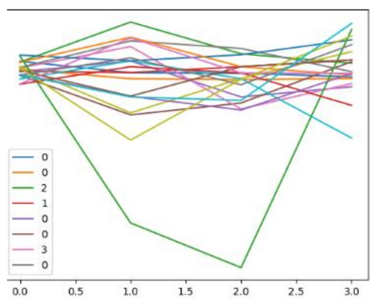
<u>ממצאים</u>

א. נתוני המניות חולקו ל-K קבוצות. להלן גרף הסיווג :

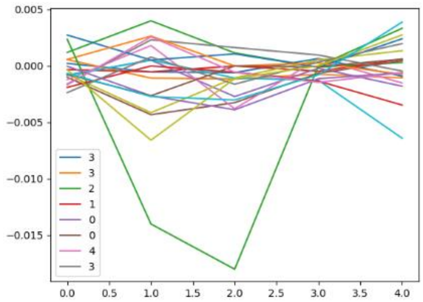
K=3



K=4

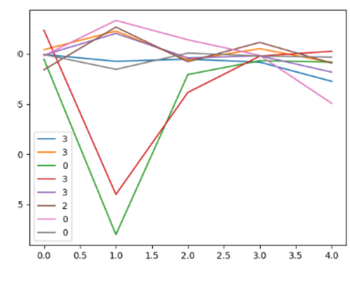


K=5

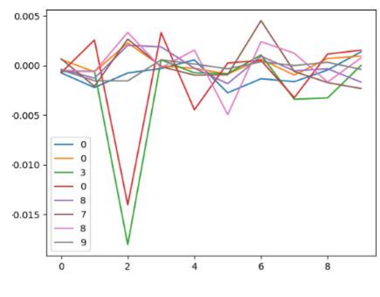


ב. נתוני הימים חולקו לK קבוצות, ע״י שיחלוף של המטריצה (Transpose) :

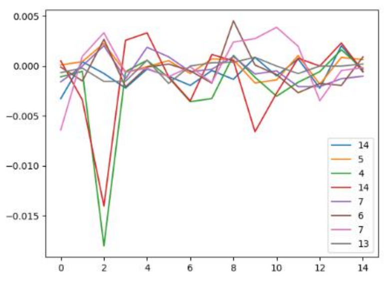
K=5



K=10



K=15



✓ המקרא בצד שמאל הוא תוצאות הסיווג לנתונים שבגרף. כפי שניתן לראות, חילקתי ע״י אלגוריתם הסיווג Kmeans את תשואות המניות בחודש ינואר 2011 ואת המניות למס׳ שונה של קבוצות.

✓ בכל האיטרציות השגיאה הייתה מאוד נמוכה (ככל שבחרתי K קטן יותר, השגיאה גדולה יותר, סדר גודל השגיאה 5-^10)

✓ <u>לפי מניות</u> - קיימת מנייה חריגה אשר בצורה קונסיסטנטית סווגה בקבוצה נפרדת מיתר הקבוצות (קו ורוד). סה״כ בחודש של מסחר זוהה סיווג אחד כחריג משמעותית (לפחות ויזואלית), מניית פפסי (Pepsi).

✓ <u>לפי ימים</u> – קיימים שני ימי מסחר חריגים משמעותית, שמסווג כירוק ואדום, לאור תשואות חריגות שליליות.

[3] Kmeans function from Sklearn module, python 3.

3. **אסטרטגיות מסחר ע"ב מקדם המתאם** : **(Pair Trading using Pearson correlation Similarity)**

א. עבור התשואות היומיות, להלן מקדם המתאם לכל זוג מניות :

pearson's Corr	Conoco	Chevron	Ford	GM	Coca Cola	OP	Pepsi	Exxon Mobile
Conoco	1.00	0.64	0.47	0.42	0.11	0.56	0.32	0.61
Chevron	0.64	1.00	0.57	0.50	0.15	0.73	0.41	0.83
Ford	0.47	0.57	1.00	0.72	0.11	0.53	0.31	0.54
GM	0.42	0.50	0.72	1.00	0.13	0.51	0.29	0.49
Coca Cola	0.11	0.15	0.11	0.13	1.00	0.13	0.22	0.17
OP	0.56	0.73	0.53	0.51	0.13	1.00	0.37	0.69
Pepsi	0.32	0.41	0.31	0.29	0.22	0.37	1.00	0.45
Exxon Mobile	0.61	0.83	0.54	0.49	0.17	0.69	0.45	1.00

ניתן לראות שמקדם המתאם המקסימלי הוא עבור מניות Chevron, המקסימליות :

- ✓ **0.83 = ExxonMobil**
- ✓ 0.73 = OP

חברת קוקה קולה מתואמת הכי פחות עם יתר המניות שבעבודה זו.

ב. מסחר בזוגות – Pair Trading :

עבור מניות Chevron ו- ExxonMobil נקנה אחת, נמכור את השנייה, תחת התנאים הבאים :

לאור המתאם הגבוה, ניתן לבצע אסטרטגיה אשר תנצל את המתאם שבין תנועות התשואות.

למעשה, ניתן לשחק עם הקניות והרכישות של המניות (לונג, ושורט) ובכך למקסם תשואה, עבור המניות הללו בשנת 2011 **קיבלתי תשואה של 24.04%, עם 247 פעולות (קוד PYTON)**

✓ יש לשים לב כי הנ"ל נכון בעולם ללא עמלות. במידה וקיימות עמלות קניה ומכירה, יש לקבוע סף פעולה לאור הכמות הגדולה של העסקאות הנדרשות לקבלת תשואה זו. קרי, תשואה אשר מעליה אכן נמכור או נקנה (Tradeoff) מול גובה העמלה בכדי למנוע הפסדים מהאסטרטגיה), בחרתי כרך את ממוצע ההפרשים בין התשואות.

ג. ניישם את האסטרטגיה שלעיל על שנת 2012.

רווח מוערך (ללא עמלות או מס) : 56.74% , עם 749 פעולות.

ככל שמקדם המתאם גבוה יותר, הרווח מהאסטרטגיה הזו בטווח הארוך יהיה גבוה יותר.

מצד שני, ככל שהמתאם גבוה יותר, סביר שיהיו פחות אפשרויות לארביטראז'ים בין המניות.

4. **אגרות חוב ללא קופון והחלפות:** (Swaps Rates & Zero Coupon Bonds)

א. ע"ב הנתונים, חישבתי את הריביות השנתיות להחלפה לעוד 5 שנים ועוד 10 שנים:

הנתונים חושבו על ריבית רציפה (r אשר נתונה בשאלה), ולאחר מכן ע"ב הנוסחה הבאה:

$$c = q * \frac{(1-DF_n)}{\sum_i^n DF_i}$$

C = Annual Swap Rates שיעור ריבית החלפה

Q = עבור 5 שנים ועבור 10 שנים

DF = Discount Factor, מקדם ההפחתה

הריביות משמשות החלפות תזרימי מזומנים (נכס/התחייבות) בין חברות אשר רוצות לשנות את פילוח אחזקותיהן לגידור סיכונים מפני שינויי ריבית במשק.

<table>
<tr><th>Years to maturity</th><th>Zero-Coupon Bonds</th><th>Discount Factor</th><th colspan="2">Annual swap rates</th></tr>
<tr><td>1</td><td>2.02%</td><td>0.980</td><td rowspan="5">5 years</td><td rowspan="5">2.661%</td></tr>
<tr><td>2</td><td>2.23%</td><td>0.956</td></tr>
<tr><td>3</td><td>2.40%</td><td>0.931</td></tr>
<tr><td>4</td><td>2.54%</td><td>0.903</td></tr>
<tr><td>5</td><td>2.64%</td><td>0.876</td></tr>
<tr><td>6</td><td>2.72%</td><td>0.849</td><td rowspan="5">10 years</td><td rowspan="5">2.924%</td></tr>
<tr><td>7</td><td>2.79%</td><td>0.823</td></tr>
<tr><td>8</td><td>2.84%</td><td>0.797</td></tr>
<tr><td>9</td><td>2.88%</td><td>0.772</td></tr>
<tr><td>10</td><td>2.91%</td><td>0.748</td></tr>
</table>

ב. להלן שערי אג"ח קופון אפס: למעשה, זו פונקציה הופכית לסעיף א' שלעיל.

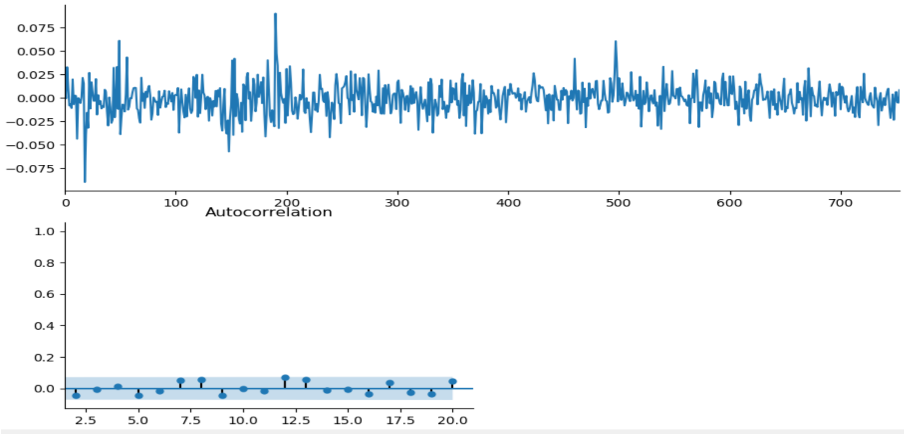
חישבתי ידנית את מקדם ההפחתה עבור שנה 1, ולאחר מכן על בסיסו חושבו הריביות הרציפות באמצעות ln הנתונים.

Years to maturity	Zero-Coupon Bonds	Discount Factor	SR
1	**-0.39%**	1.0039	-0.39
2	**-0.15%**	1.0030	-0.15
3	**0.12%**	0.9964	0.12
4	**0.36%**	0.9856	0.36
5	**0.59%**	0.9707	0.59
6	**0.81%**	0.9527	0.80
7	**0.99%**	0.9329	0.98
8	**1.15%**	0.9123	1.13
9	**1.29%**	0.8902	1.27
10	**1.41%**	0.8687	1.38

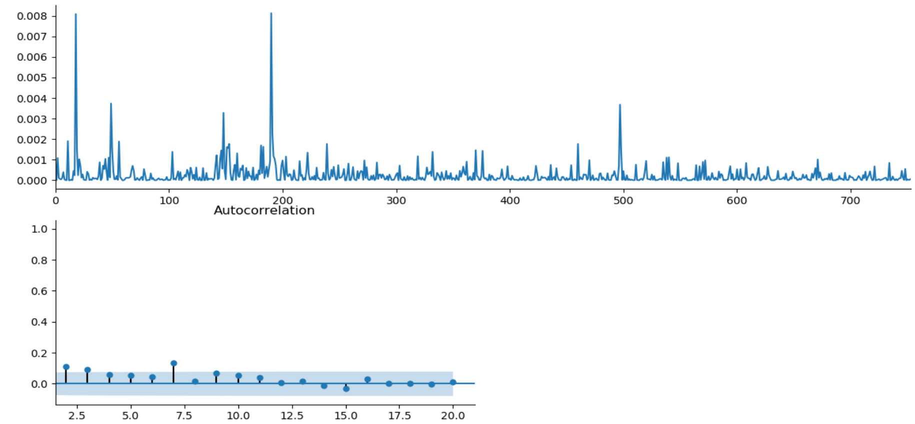
התשואות מגלמות עליה במקסימלית בשערי ריביות בשנה השנייה לשלישית של כ- 0.3%. בשל המעבר מריביות שליליות לחיוביות. ניתן לראות שעל רקע הריביות הנמוכות במשק, בשבדיה כבר נסחרות אג"ח המגלמות תשואה שלילית.

נספח א' – גרף תשואות ומקדם מתאם פנימי

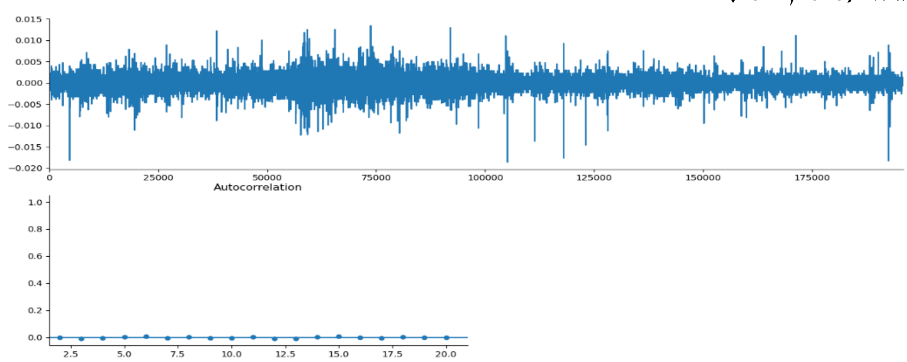
1. תשואות 3 שנים:



2. ריבועי תשואות 3 שנים:



3. תשואות יומיות נבחרות (2011,2013):



4. ריבועי תשואות יומיות נבחרות: