**פרויקט חיזוי S&P 500**

**גילעד גלבוע**

**מבוא:**

מדד S&P 500 הוא אחד המדדים המרכזיים והמשפיעים ביותר בעולם ההשקעות, המייצג את ביצועיהן של 500 החברות הגדולות הנסחרות בבורסות האמריקאיות. המדד נחשב לברומטר עיקרי של הכלכלה האמריקאית ושל שוק ההון העולמי, ומשמש ככלי מרכזי להערכת מגמות כלכליות והחלטות השקעה, בין היתר משום שהוא מייצג כ – 80% משווי השוק הכולל של החברות הנסחרות בבורסה האמריקאית, לאור כל זאת, זוכה המדד לפופולריות עצומה בקרב משקיעים פרטיים, מוסדיים ואנליסטים.

בעידן שבו שוק ההון מושפע מגורמים כלכליים רבים וממקורות מידע שונים, הבנת הכוחות המשפיעים על ביצועי מדדים מרכזיים כמו S&P 500 הפכה לקריטית. בעבודה זו נבחן את הדרכים שבהן ניתן לחזות את תשואות המדד באמצעות מודלי הסקה בייסיאנים, תוך שילוב בין גורמים כלכליים ומידע עכשווי, במטרה לשפר את יכולת הניבוי והבנת השוק.

**נתונים:**

* קובץ csv המכיל מידע על ערך מדד S&P 500 בעל עמודת DATE ועמודות open, close

מקור: Kaggle .

* קובץ csv המכיל מידע על ההשפעות המאקרו כלכליות (תמ"ג, אינפלציה, אבטלה וריבית) בעל עמודת DATE

הקובץ נבנה על ידי באמצעות חיבור עמודות רלוונטיות מקבצים שונים שנלקחו מאתרים רשמיים של רשויות כלכליות אמריקאיות (FRED, BIS, OECD )

* קובץ csv המכיל מידע על מדד S&P 500 וכמות החדשות הנוגעות אליו עם העמודות (הרלוונטיות): מס' כתבות חיוביות ושליליות על המדד, מס' ניתוחי אנליסטים בחדשות על המדד.

מקור: Kaggle .

**שאלת מחקר 1:**

**כיצד מדדים מקרו כלכליים משפיעים על תשואת S&P 500?**

שיטת המחקר: בדיקת השערות

**השערת האפס:**  למדדים המקרו כלכליים (תמ"ג, שיעור האינפלציה, שיעור האבטלה וריבית) אין השפעה משמעותית על תשואות מדד S&P 500 .

**השערה אלטרנטיבית:** למדדים המקרו כלכליים יש השפעה משמעותית על תשואות מדד S&P 500 .

המודל: מודל רגרסיה לינארית בייסיאני.

המודל יבחן את הקשר בין תשואות המדד באחוזים (yield percents) לבין המדדים המקרו כלכליים באמצעות רגרסיה לינארית.

הנחות המודל:

הנחת הבסיס של המודל תהיה ששיעור התשואה החודשי של מדד S&P 500 מתפלג נורמלי:

Yield\_percents

פריורים:

* בחרנו בכל הפרמטרים על מנת לאפשר את ההנחה שאין נטייה ברורה להשפעה חיובית או שלילית למדדים המקרו כלכליים על מדד S&P 500.
* בחרנו למקדם החופשי את השונות על מנת לאפשר מרחב אפשרויות רחב לפרמטרים.
* בחרנו לפרמטרים את השונות המאפשרת למודל להיות גמיש מספיק לתפוס את ההשפעה האפשרית של כל אחד מהמשתנים, ומייצרת פריור לא אינפורמטיבי מדי.
* בחרנו לפרמטר משום שהוא nuisance parameter ובחירה זו אינה אינפורמטיבית ומאפשרת טווח רחב יחסית של ערכים, אך עדיין משמרת את האפשרות שהשגיאות לא יהיו גדולות מדי.

תמונה שמכילה טקסט, תרשים, קו, עלילה

התיאור נוצר באופן אוטומטי

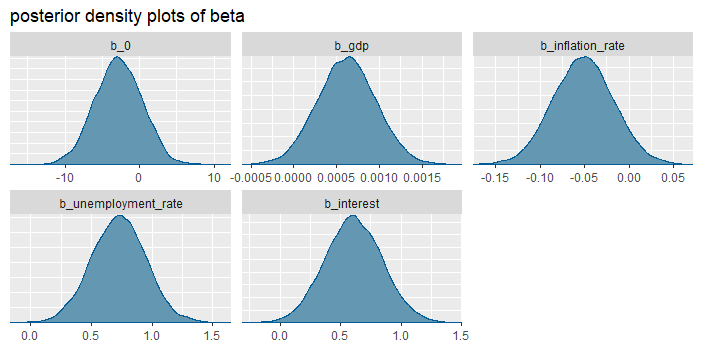
התפלגות פוסטריורית:

MCMC שימש ליצירת דגימות מההתפלגות הפוסטריורית של המקדמים. הגדרות הדגימה:

* 4 שרשראות
* 5000 איטרציות בכל אחת
* 1000 איטרציות (מתוך ה5000) שימשו כ"חימום"
* Adapt\_delta = 0.95 (המלצה מוצלחת שקיבלתי להגדלת הess )

נבחן את ההתפלגות הפוסטריורית של המקדמים כדי לבדוק האם הן משפיעות על תשואת מדד S&P 500

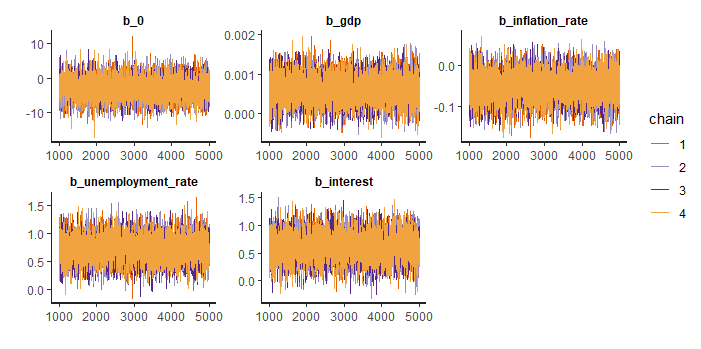
* ניתן לראות את תוצאות הדגימות המורחבות בקוד המצורף



הנתונים המרכזיים מוצגים בטבלה:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **beta** | **mean** | **se\_mean** | **sd** | **2.50%** | **50%** | **97.50%** | **ess** | **Rhat** |
| b\_0 | -2.83783 | 0.0672484 | 3.136694 | -8.9702 | -2.85335 | 3.154498 | 2175.605 | 1.002392 |
| b\_gdp | 0.000619 | 4.53E-06 | 0.000341 | -4.28E-05 | 0.000621 | 0.001288 | 5654.566 | 1.00013 |
| b\_inflation\_rate | -0.05157 | 0.0005155 | 0.033865 | -0.1175 | -0.05156 | 0.014521 | 4315.486 | 1.000371 |
| b\_unemployment\_rate | 0.725527 | 0.0033326 | 0.222955 | 0.288945 | 0.727274 | 1.15981 | 4475.773 | 1.001457 |
| b\_interest | 0.607342 | 0.0031695 | 0.235909 | 0.145593 | 0.606894 | 1.071457 | 5539.972 | 1.000964 |

נוודא עקבות תקינות של הבטאות שלנו:

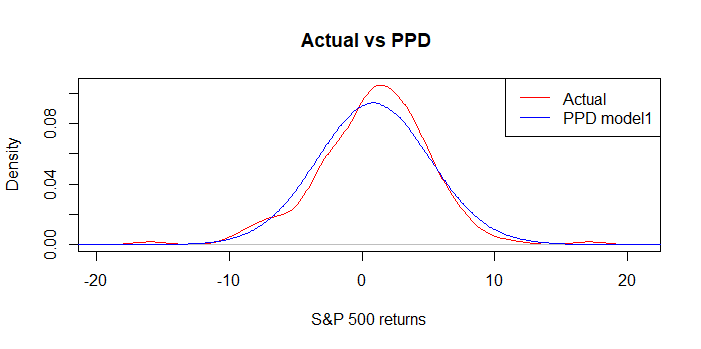


כעת, בעבור כל ערך מאקרו כלכלי, נבדוק האם ההשפעה שלו על המדד שונה מאפס בהסתברות של 95%.

מסקנות כלליות:

* באופן כללי ניתן לראות כי המודל מצליח לחזות בקירוב את תשואת מדד S&P 500 ומייצר התפלגות דומה מאוד להתפלגות הנתונים שלנו.
* ניתן לראות כי כל אחד מהמשתנים התורמים לתשואת S&P 500 מראה אפקט שונה.
* שיעור האבטלה והריבית משפיעים בצורה משמעותית על השוק, כאשר עלייה בהם מובילה לעלייה בתשואת S&P 500.
* שיעור האינפלציה מצביע בממוצע על השפעה שלילית על תשואת השוק, ככל הנראה בשל העלויות הנוספות שנובעות מאינפלציה גבוהה.
* התמ"ג משפיע באופן חיובי בממוצע אך מינורי על התשואה, מה שיכול להצביע על כך שצמיחה כלכלית משתקפת בשוק ההון, אך עם השפעה מוגבלת.

תוצאות בדיקת ההשערות:

* בהתאם לתוצאות המודל, אנו יכולים לבחון האם לדחות או לקבל את השערות האפס עבור כל אחד מהמשתנים הנכללים במודל. בהשערת האפס, אנחנו מניחים כי לכל אחד מהמשתנים (תמ"ג, שיעור אינפלציה, שיעור אבטלה וריבית) אין השפעה מובהקת על תשואת ה-S&P 500, כלומר שההשפעה שלהם היא אפסית.
* לפי תוצאות המודל, ניתן לראות כי עבור שיעור האבטלה והריבית, הטווח הבדוק של ערכי ההשפעה אינו מכיל את האפס, מה שמצביע על כך שהשפעתם על השוק היא מובהקת. לכן, אנו דוחים את השערת האפס עבור משתנים אלו ומקבלים את ההשערה כי יש להם השפעה מובהקת על תשואת השוק.
* ****לעומת זאת, עבור משתני התמ"ג ושיעור האינפלציה, אף כי הטווח הבדוק מראה מגמות של השפעה, אין בכך כדי לשלול לחלוטין את האפשרות שאין להם השפעה מובהקת על תשואת השוק, שכן האפס נמצא בתוך הטווח הבדוק. לכן, במקרים אלו אנו מתקשים לדחות את השערת האפס, ומציעים כי ייתכן וההשפעה שלהם על השוק היא פחות מובהקת או אף אינה קיימת.

**שאלת מחקר 2:**

**האם המודל הבייסיאני שלנו יכול לחזות שינוי בתשואות מדד S&P 500 בין תקופה של צמיחה כלכלית משמעותית לבין תקופה של חוסר התקדמות?**

שיטת המחקר: posterior predictive distribution

בחלק זה נבחן האם המודל הבייסיאני שבנינו מסוגל להבחין ולהציג שינויים בתשואות מדד S&P 500 בין 2 תקופות כלכליות מוגדרות:

* תקופה של "צמיחה כלכלית משמעותית": תקופה שבה התמ"ג עלה ב1% או יותר ואחוזי האינפלציה או האבטלה אינם חלק מ20% העליונים מתוך כל הנתונים הקיימים בידינו.
* תקופה של "חוסר התקדמות": תקופה שבה התמ"ג עלה בפחות מ1% או ירד, או שאחוז האינפלציה או האבטלה גבוה בצורה חריגה (לפי הפרמטר המצוין לעיל)

מודל ונתונים:

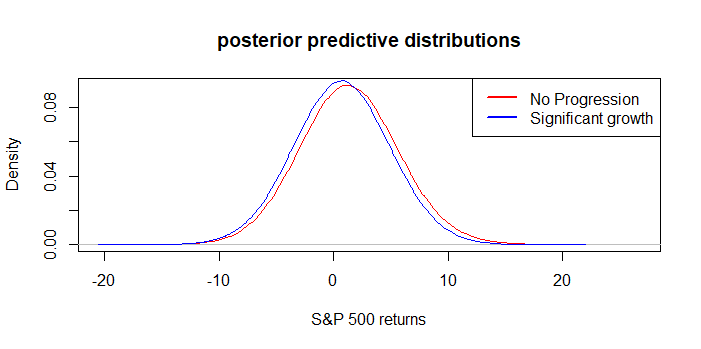
נעשה שימוש באותו מודל רגרסיה לינארית בייסיאני ובאותם נתונים כלכליים שנעשה בהם שימוש בשאלה 1, במטרה לבחון את ההשפעה של שלבים שונים של מחזור כלכלי על תשואות מדד S&P 500.

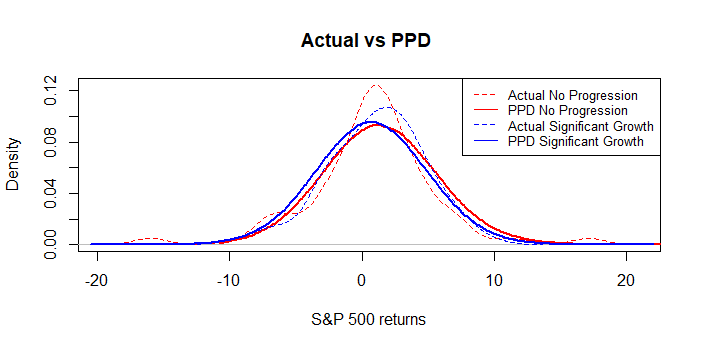
התפלגות פוסטריורית חזויה (PPD):

MCMC שימש ליצירת דגימות מההתפלגות הפוסטריורית של התשואות בשתי התקופות הכלכליות. הגדרות הדגימה:

* 4 שרשראות
* 5000 איטרציות בכל אחת
* 1000 איטרציות (מתוך ה5000) שימשו כ"חימום"
* Adapt\_delta = 0.95

נבחן את ההתפלגות הפוסטריורית החזויה של התשואות כדי לבדוק האם המודל יכול להבחין בין תקופות של צמיחה כלכלית לבין תקופות של חוסר התקדמות.

* ניתן לראות את תוצאות הדגימות המורחבות בקוד המצורף

נציג גם את ההתפלגות הפוסטריורית החזויה אל מול התצפיות (הנתונים האמיתיים):

הגרפים המצורפים מראים את ההתפלגות החזויה הפוסטריורית לתשואות מדד S&P 500 עבור תקופות של "צמיחה משמעותית" לעומת "חוסר התקדמות", וכן השוואה להתפלגות התשואות האמיתיות.

מסקנות כלליות:

* המודל הצליח לחזות בקירוב את ההתפלגות הכללית של תשואות מדד S&P 500 בשתי התקופות שנבדקו (צמיחה כלכלית משמעותית וחוסר התקדמות), אך ישנם פערים באיזור התצפיות השכיחות יותר.
* בהשוואה בין תוצאות ההתפלגות הפוסטריורית החזויה בשתי התקופות, נראה כי המודל מייצר שני גרפים כמעט זהים, מה שמעיד על כך שהמודל מתקשה להבחין בין מצבי המחזור הכלכלי.
* לאור הממצאים, ניתן להסיק כי המודל אינו יכול לחזות באופן מדויק שינויים במדד S&P 500 על בסיס מצב המחזור הכלכלי.

תוצאות:

בהתאם לממצאים, המודל הבייסיאני שנבנה הצליח להציג התפלגות פוסטריורית כללית לתשואות מדדS&P 500 התואמת בקירוב להתפלגות הנתונים האמיתיים בשתי התקופות שהוגדרו: תקופה של "צמיחה כלכלית משמעותית" ותקופה של "חוסר התקדמות". עם זאת, כאשר משווים בין התפלגות התשואות הפוסטריורית בשתי התקופות, נמצא כי המודל ייצר שני גרפים כמעט זהים. הדבר מעיד על כך שהמודל מתקשה להבחין בין מצבים שונים במחזור הכלכלי, ומצביע על כך שהיכולת שלו לחזות שינויים בתשואות המדד על בסיס המצב הכלכלי הכללי מוגבלת.

בשל כך, ניתן להסיק כי המודל הנוכחי אינו מתאים באופן מיטבי להבחנה בין תקופות של "צמיחה כלכלית משמעותית" לבין תקופות של "חוסר התקדמות". נדרש לשפר את המודל על ידי הוספת משתנים נוספים או שימוש בגישות אחרות, כמו רגרסיה לא לינארית, שיכולות לנבא בצורה מדויקת יותר את השינויים בתשואות במהלך תקופות כלכליות שונות.

**שאלת מחקר 3:**

**האם עדיף להעריך את תשואת S&P 500 בתדירות חודשית ע"פ מדדים מקרו כלכליים או בתדירות יומיומית ע"פ דיווחים והערכות של ערוצי חדשות?**

שיטת המחקר:השוואה בין מודלים

המודלים:

* מודל 1: מודל (חודשי) לפי מדדים מקרו – כלכליים: אותו מודל רגרסיה לינארי שנבנה בשאלה 1.
* מודל 3: מודל (יומי) לפי ניתוחי אנליסטים והערכות של ערוצי חדשות:

המודל יבחן את הקשר בין תשואות המדד באחוזים (yield percents) לבין דיווחים והערכות של ערוצי חדשות באמצעות רגרסיה לינארית.

הנחות המודל:

הנחת הבסיס של המודל תהיה - שיעור התשואה החודשי של מדד S&P 500 מתפלג נורמלי (כמו במודל משאלה 1):

Daily\_yield

פריורים:

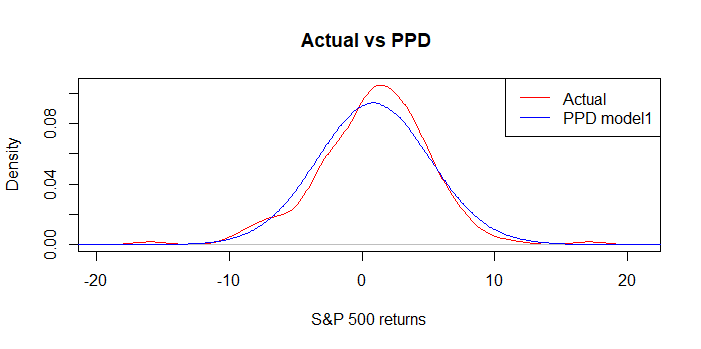
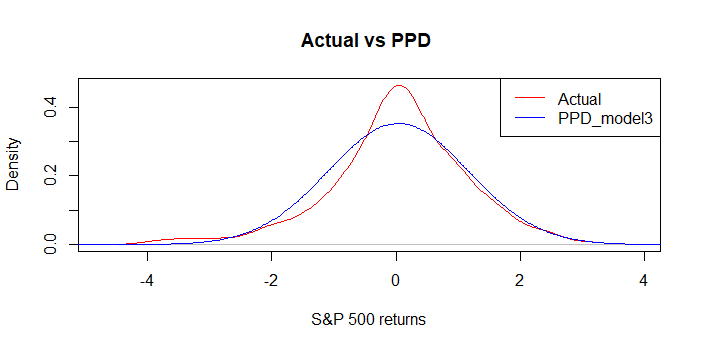
* המחשבה על בחירת הפריורים נעשתה בדיוק כמו במודל הראשון.

התפלגות פוסטריורית:

MCMC שימש ליצירת דגימות מההתפלגות הפוסטריורית של המקדמים לשני המודלים.

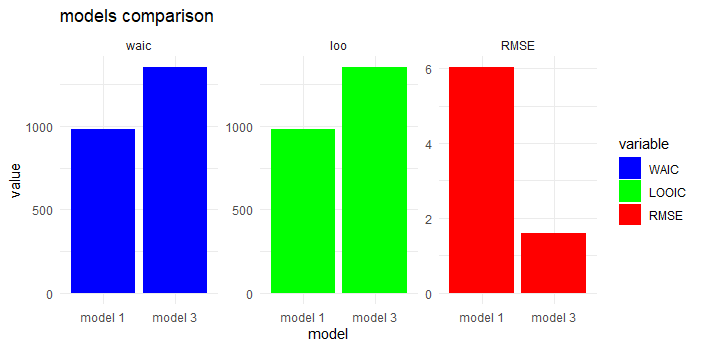
הגדרות הדגימה היו שוות וזהות להגדרות הדגימה בשאלה מספר 1.

נבחן את ההתפלגות הפוסטריורית החזויה של המודלים אל מול הנתונים:



|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **beta** | **mean** | **se\_mean** | **sd** | **2.50%** | **50%** | **97.50%** | **ess** | **Rhat** |
| b\_0 | 0.293261 | 0.0096023 | 0.352069 | -0.39339 | 0.300763 | 0.96643 | 1344.317 | 1.001665 |
| b\_positive | 4.43E-05 | 8.45E-07 | 9.05E-05 | -0.00013 | 4.37E-05 | 0.000221 | 11472.12 | 0.999918 |
| b\_negative | -0.00011 | 3.29E-06 | 0.000181 | -0.00046 | -0.00011 | 0.000247 | 3020.362 | 1.001372 |
| b\_analyst | -2.85E-05 | 5.45E-07 | 4.31E-05 | -0.00011 | -2.86E-05 | 5.59E-05 | 6248.554 | 1.000102 |

השוואת המודלים:

המודלים הושוו באמצעות 3 מדדים: WAIC, LOOIC, RMSE(root mean squared error)

מסקנות כלליות:

* מודל 1 (חודשי, מתבסס על מדדים מקרו כלכליים) מציג ערכים WAIC ו- LOO נמוכים יותר, מה שמעיד על התאמה טובה יותר לנתונים בהשוואה למודל 3 (יומי, מתבסס על כמות ניתוחים אנליסטים והערכות).
* מודל 3 מציג ערך RMSE נמוך יותר, מה שמעיד על ניבוי מדויק יותר של תשואות יומיות ויכול להציג מסקנה הפוכה.

אף על פי כן, יש לקחת בחשבון:

1. ה RMSE רגיש לסטיות גדולות בנקודות מסוימות.
2. מס' הנתונים שניתן למודל 3 גדול יותר (יומי, לא חודשי).

נקודות אלו יכולות להסביר את ההבדל בערך הRMSE בין שני המודלים.

* מבחינה ויזואלית, כאשר משווים את התחזיות המיוצרות על ידי שני המודלים למול הנתונים בפועל, נראה שלמודל 1 יש התאמה טובה יותר להתפלגות הכללית של הנתונים, אך גם בעניין זה נדרשת משנה זהירות, משום שפרופורציות הצירים בשני הגרפים שונות.

תוצאות:

הבחירה בין המודלים תלויה במטרת המחקר:

אם המטרה היא ניבוי ארוך טווח או חיזוי מדד S&P 500, המודל החודשי התלוי במדדים מאקרו כלכליים עשוי להיות הבחירה המועדפת, מכיוון שהוא מציג התאמה טובה יותר לנתונים הכוללים כפי שנמדד על ידי WAIC ו- LOO .אך אם המטרה היא לנבא שינויים קצרים או התנהגויות של המדד לטווח קצר, עדיין ניתן לשקול את המודל היומי המתבסס על כמות ניתוחים אנליסטים והערכות, משום שניתן לראות כי המודל מצליח לחזות בקירוב את תשואת מדד S&P 500 ומייצר התפלגות דומה באופן יחסי להתפלגות הנתונים.

**סיכום העבודה:**

במסגרת העבודה בחנו את יכולות המודלים הבייסיאניים לניבוי תשואות מדד S&P 500 תוך שימוש במדדים מקרו-כלכליים ונתונים עכשוויים כמו דיווחי חדשות וניתוחי אנליסטים. באמצעות שילוב של כלים מתקדמים להסקה בייסיאנית, הצלחנו לפתח מודלים המספקים תובנות משמעותיות לגבי האופן שבו משתנים כלכליים משפיעים על תשואות המדד, והראינו כיצד ניתן להשתמש במודלים אלו כדי להבין ולחזות את התנהגות המדד בתנאים כלכליים משתנים.

המודלים שפיתחנו הצליחו לזהות קשרים מרכזיים בין המדדים הכלכליים לבין ביצועי המדד, ולהציע דרכים מבוססות נתונים לניבוי תשואות המדד לאורך זמן. העבודה הדגימה את הפוטנציאל של גישות בייסיאניות להעניק תמונה מקיפה ומדויקת של השוק, תוך שמירה על גמישות והתאמה למגוון רחב של מצבים כלכליים.

במהלך העבודה זיהינו מספר אתגרים והזדמנויות לשיפור המודלים. אחד האתגרים המרכזיים היה הקושי של המודל הראשון להבחין בצורה מובהקת בין תקופות כלכליות שונות, כמו תקופות של צמיחה משמעותית לעומת תקופות של חוסר התקדמות. בנוסף, במודל האחרון, ראינו פערים משמעותיים יחסית בין התחזיות לנתונים באיזורים השכיחים יותר של המודל. פערים אלו יכולים לנבוע ממורכבות הנתונים, איכות המדדים (אין לנו אינדיקציה על איכות החדשות וההערכות בנתונים), או פשטות המודל, מה שמצביע על הצורך לשפר את המודל כדי ללכוד בצורה מדויקת יותר את הדינמיקות המורכבות של השוק היומי.

למרות האתגרים, העבודה מהווה בעיניי צעד משמעותי בהבנת התנהגות מדד S&P 500 והקשרים הכלכליים המורכבים המשפיעים עליו. התובנות שהתקבלו מדגישות את החשיבות של פיתוח מודלים מתקדמים יותר, שיכולים לשפר את הדיוק והאמינות של התחזיות, ובכך לסייע למשקיעים ואנליסטים בקבלת החלטות מושכלות בשוק ההון.