



## ניתוח דוחות כספיים של בנקים – גישת מידול נושאים

### (Topic Modeling) לזיהוי סיכונים

#### אלון רביב

בית הספר למנהל עסקים  
אוניברסיטת בר אילן  
Alon.Raviv@biu.ac.il

#### אבינעם בלום

בית הספר למנהל עסקים  
אוניברסיטת בר אילן  
Avinoam.blum@biu.ac.il

#### תקציר

ניתוח טקסט באמצעות בינה מלאכותית נמצא בשימוש נרחב בשנים האחרונות בתחומים רבים כמו רפואה, מדע וביטחון. לאחרונה, החלו ליישם שיטות אילו גם בתחומי הכלכלה והמימון. בעבודה זו נסקור ונסביר את השיטות המובילות כיום לניתוח נושאים בספרות הכלכלית: LDA (Latent Dirichlet Allocation) ו-NMF (Non-Negative Matrix Factorization). בנוסף, נחקור את השוני בתגובה של הבנקים הישראלים למשבר הפיננסי של 2008 ולפרסום הנחיות באזל III. בהתייחס למידת ההומוגניות במערכת הבנקים, נמצא כי במרבית התקופות המערכת הבנקאית הגיבה באופן דומה, כאשר נצפתה התנהגות שונה בקרב בנק הפועלים בכל הקשור להשפעת המשבר על נושא ההון. לפי מודל ה-LDA, משקל הנושא עלה באופן מובהק עם פרוץ המשבר בעוד שביתר הבנקים לא חל שינוי מהותי. בתחום האשראי, רק בבנק לאומי נצפתה בתקופת המשבר, עליה מובהקת בשכיחות המונח "credit risk", בעוד שביתר הבנקים לא חל שינוי שכזה. ייתכן והשינויים הנצפים הם כי שני הבנקים המובילים בישראל, נאלצו לאמץ רגולציות עודפות ביחס ליתר הבנקים, בגין התפיסה של Too Big To Fail והחשש מתגובת שרשרת.

מילות מפתח: TBTF, Basel III, NMF, LDA

## הקדמה

בישראל<sup>1</sup>. המודלים הדומיננטיים במחקרים אלה כוללים מידול נושאים (Topic Modeling) וכלים מסורתיים לניתוח שכיחות המונחים: Document Term Matrix (DTM). ההחלטה להתמקד בניתוח הטקסט בדוחות הכספיים של הבנקים נתמכת במחקרים אחרונים המראים כי הסיכונים המדווחים בדוחות אילו מהווים מקור מידע חשוב למנהלי סיכונים (Campbell 2014). יתר על כן, מודלים מסוג LDA כפי שיושמו במחקר של בלום ורביב (2023), מאפשרים לזהות ולכמת את משקל נושאי הסיכון בדוחות הכספיים ובהתאם לזהות שינויים לאורך זמן (Datta et Bao 2014). במחקרם, בלום ורביב (2023) מנסים להתמודד עם שתי שאלות מרכזיות כיצד המשבר הפיננסי משפיע על תוכן הדוחות הכספיים של הבנקים בישראל? ואיך רפורמת באזל III משפיעה על תוכן הדוחות הכספיים? ברם, בעוד שעבודתם מתייחסת לתגובת כל המערכת הפיננסית לאירועים אלה, לא קיימת התייחסות לשונות בתגובת המוסדות הפיננסיים לאירועים המתוארים. לכן, במסגרת מחקר זה נבחן גם את השאלה הבאה: מהי מידת ההומוגניות בין הבנקים השונים בהתייחס לשאלות המחקר של בלום ורביב (2023), קרי, האם במערכת הבנקאית בישראל, ישנם בנקים שהתנהגו באופן שונה ביחס ליתר הבנקים? תשובה לשאלה זאת תאפשר בעתיד להסתכל על מאפיינים השונים של הבנקים כמו גודל, יציבות ואופי תיק האשראי בבואנו לנתח את תגובתם לאירועים שונים. בלום ורביב (2023), מצביעים על כך, שבתקופת המשבר הפיננסי של 2008, הבנקים המובילים בישראל שינו את המיקוד שלהם בדוחות הכספיים מהתמקדות בסיכונים שוק למעבר לסיכונים אשראי וסיכונים נזילות והגברת החשיבות של תרחישי קיצון. יתרה מזאת נמצא כי במהלך תקופת פרסום הנחיות באזל III (2010-2013) ותקופת יישום ההנחיות (2014-2018), נמשכה הירידה במשקלם של הנושאים הקשורים לסיכונים שוק וספציפית

בשנים האחרונות השימוש בבינה מלאכותית גובר בקצב התפתחות מעריכית, בעיקר בשל הגורמים הבאים: שיפורים ניכרים ביכולת המחשוב, התפתחות תחום הביג דאטה, התקדמות מתמדת באלגוריתמים, מעבר למחשוב בענן ומחשוב מבוזר/דמוקרטיזציה של המידע ולאחרונה הנגשת פלטפורמות כמו ChatGPT. גורמים אלו, בשילוב התפתחות מחקרית, תרמו לצמיחה ניכרת בתחום ה-AI. אחד מאותם תחומים שהתפתחו שייך לתחום ה-NLP, שכולל בתוכו ניתוח טקסט. התחום הנ"ל מורכב ממודלים שונים כאשר פריצת דרך נרשמה עם התפרסמות המאמר המהפכני של Blei, Ng and Jordan (2003) בשם "Latent Dirichlet Allocation", שעוסק בתחום של מידול נושאים (Topic Modeling). הרעיון מאחורי המודל שפיתחו, שניתן באופן ממוכן לסווג טקסט המורכב מאסופת מסמכים שונה (משפטים, מאמרים או תקציר דיונים לדוגמה) לנושאים שונים ולמצוא את משקל הנושאים השונים בכל מסמך, שמרכיב את הסט הגדול. מידע זה מאפשר לנו מבלי לקרוא את כל הטקסט להבין את התוכן שלו. מודל ה-LDA מאפשר ניתוח מידע על פני זמן, קרי, לזהות תבניות שונות בגין השינויים שנצפים במשקל הנושאים שהמודל מייצר לאורך זמן. כפועל יוצא ניתן לחקור מגמות שונות בתחומי מחקר שונים, לרבות זיהוי דפוסי התנהגות ועוד. למעשה מדובר בניתוח על סמך מודל לא מפקוח (unsupervised machine learning). יישום שיטות שכאלו מאפשר לכאורה להבין את תחום המחקר מזווית נוספת שעד כה נחקר בצורה מוגבלת.

במאמר זה נסקור מחקרים שהשתמשו במודלים לעיבוד שפה לניתוח הדוחות הכספיים של בנקים, בדומה לעבודתם של בלום ורביב (Blum and Raviv 2023), שחקרו את השפעות המשבר הפיננסי של 2008 ואת הנחיות באזל III על ניהול הסיכונים בארבעת הבנקים המובילים

במדגם. בנוסף, מדובר בבנק הקטן ביותר מתוך חמשת הבנקים הגדולים, ובהתאם ההשפעה אינה מהותית.

<sup>1</sup> הערה: הבנק החמישי בגודלו, הבנק הבינלאומי לא נכלל במדגם, משום שאחד החוקרים שימש בתפקידים בכירים שם מעל ל-20 שנה וכפועל יוצא הוחלט שלא להכליל זאת

## 2. שכיחות מונחים

השלב הראשון של המחקר התמקד בספירת שכיחות המילים שמרכיבות את המסמך (DTM - Document Term Matrix) כפי שמופיע בטבלה 1:

	T <sub>1</sub>	T <sub>2</sub>	T <sub>3</sub>	T <sub>4</sub>	T <sub>5</sub>	T <sub>6</sub>	T <sub>N</sub>
D <sub>1</sub>	W <sub>11</sub>	W <sub>12</sub>	W <sub>13</sub>	W <sub>14</sub>	W <sub>15</sub>	W <sub>16</sub>	W <sub>1N</sub>
D <sub>2</sub>	W <sub>21</sub>	W <sub>22</sub>	W <sub>23</sub>	W <sub>24</sub>	W <sub>25</sub>	W <sub>26</sub>	W <sub>2N</sub>
D <sub>M</sub>	W <sub>M1</sub>	W <sub>M2</sub>	W <sub>M3</sub>	W <sub>M4</sub>	W <sub>M5</sub>	W <sub>M6</sub>	W <sub>MN</sub>

טבלה 1. טבלת שכיחות מונחים DTM.

למשל אם לצורך ההמחשה, הקורפוס שלנו מורכב משלושת המסמכים הבאים:

1. Avinoam plays the violin, he enjoys playing music. Playing music is one of Avinoam's joys.
2. In the world of music, Avinoam's tunes are played with a violin. Avinoam's music can be joyful and calming. Avinoam cherishes it.
3. Avinoam's music is joyful. He cherishes every moment with the violin."

נמצא כי המונחים הייחודיים בטקסט הם: Avinoam, Violin, Playing, Music, Joy, Calming, Cherishes, במקרה זה מטרצת ה-

DTM תיראה כמו בטבלה 2:

Document/Term	Avinoam	Violin	Playing	Music	Joy	Calming	Cherishes
Document 1	2	1	3	2	2	0	0
Document 2	3	1	1	2	1	1	1
Document 3	1	1	0	1	1	0	1

טבלה 2. טבלת שכיחות מונחים DTM לפי הדוגמא שצורפה.

השלב השני התמקד בזיהוי שכיחות מונחים הקשורים לסיכון שנקחו מבאזל, גישה מפורחת זו

המושג VaR. מנגד, חל גידול במשקל סיכוני האשראי. לפיכך, המשבר הפיננסי של 2008 מסמן נקודת מפנה, שבה הבנקים החלו להתמקד ביתר שאת בתחום סיכוני אשראי וסיכוני נזילות, בעוד שהם התייחסו לסיכוני שוק כסיכון פוטנציאלי נמוך, זאת כי יתכן כי למדו לנהל את סיכוני השוק באופן יעיל ומבוקר. במסגרת מחקרנו נמצא כי קיימת התנהגות שונה של שני הבנקים הגדולים במערכת. תחילה, בנק לאומי מגדיל באופן מובהק את התייחסותו לסיכוני אשראי ותרחישי קיצון בזמן המשבר בעוד שבמערכת הבנקאית לא חלו שינויים מובהקים, ואחר כך עם פרסום באזל III, חלה עליה מובהקת במונח סיכוני אשראי בבנק הפועלים, בעוד שבבנק לאומי מזהים ירידה במונחים הקשורים לסיכוני שוק בתקופה זו. בשאר הבנקים במערכת לא חלו שינויים מובהקים. מניתוח התנהגות המערכת באופן פרטני בהתבסס על מודל ה-LDA, נמצא כי בנק הפועלים הינו הבנק היחיד שבתקופת המשבר, משקל נושא ה-"capital debt" גדל בצורה מובהקת. בנק זה, החזיק במכשירים פיננסיים מורכבים, שפגעו באופן ישיר בהלימות ההון שלו ביחס ליתר הבנקים המקומיים<sup>2</sup>. בנוסף נמצא כי בתקופה זו, בנק לאומי התאפיין בהפסדי אשראי גדולים ביחס למערכת בין השאר בגין החשיפה ל-Lehman Brothers<sup>3</sup> credit. יש לציין כי באותה תקופה, חלה עליה מובהקת בשכיחות המונח "risk credit" אך ורק בבנק זה. ייתכן ומידה מסוימת של הטרוגניות, שנצפתה בשני הבנקים הגדולים, נובעת מהחשיפה שלהם לשווקים בינלאומיים ומהחשש לנפילת אחד משני הבנקים הגדולים ביותר ו/או רגולציה עודפת שהושתתה על בנקים אילו, ששייכים במידה מסוימת לקטגוריה של Too Big To Fail.

<sup>3</sup> פרק ג' "סיכונים והלימות ההון" דוח בנק ישראל 2008

<sup>2</sup> פרק ג' "סיכונים והלימות ההון" דוח בנק ישראל 2007

כך מקבעים מראש את שני ההיפר-פרמטרים (קובעים אותם לפני הרצת המודל) הבאים:

**אלפא** הוא היפר-פרמטר לאמידת התפלגות הנושאים המרכיבים את המסמכים, קרי, מידת הריכוזיות של נושאים עלפני מסמכים שונים. ערך אלפא גבוה יותר גורם לכל מסמך להכיל תערובת של נושאים רבים, כלומר כל מסמך ישוּך למספר גדול יותר של נושאים. לעומת זאת, ערך אלפא נמוך יותר יגרום לכל מסמך להכיל פחות נושאים, וכתוצאה מכך ישנם פחות נושאים לכל מסמך (ריכוזיות גבוהה יותר).

**בטא** הוא היפר-פרמטר נוסף שמשפיע על התפלגות המילים המרכיבות את הנושאים. ערך בטא גבוה יותר גורם לכל נושא להכיל תערובת של מילים רבות, כלומר כל נושא משוּך למספר גדול יותר של מילים עם משקל נמוך יותר (התפלגות יותר אחידה). ערך בטא נמוך יותר יגרום לכל נושא להיות מורכב מקבוצה קטנה יותר של מילים, וכתוצאה מכך הקצאת פחות מילים לכל נושא (ריכוזיות גבוהה יותר).

הפלט של המודל מורכב ממשקלים שניתן לנרמל להסתברויות:

(א) ההסתברות שמסמך ספציפי יפיק נושא ספציפי: התפלגות הנושאים שמרכיבים את המסמכים.

(ב) ההסתברות שנושא ספציפי יפיק מילה ספציפית: התפלגות המילים שמרכיבות את הנושאים.

מידול נושאים המבוסס על LDA דורש בחירה אפריורית של מספר הנושאים כקלט של המודל. כדי להעריך את מהימנות ה-LDA משתמשים במדד ה-Coherency, אשר נותן הערכה בדבר איכות הנושא שיכול להיות מיוצג על ידי המכלול של מילים בתוכו (Chen, Stevens)

התהליך הגנרי לחישוב LDA הינו:

1. לכל נושא  $k \in \{1, \dots, L\}$ , יש לבחור

$\vec{\phi}_k \sim Dir(\beta)$  שמשקף את התפלגות

המילים לכל נושא

2. לכל מסמך  $d \in \{1, \dots, M\}$  יש לבצע

את השלבים הבאים:

יושמה על מנת לקבל אינדיקציה אודות חשיבות מונחים שונים שגלומים בטקסט. ההנחה הבסיסית במתודולוגיה זו, שמילים בעלות השכיחות הגבוהה ביותר (לאחר ניקוי הטקסט) הן המילים בעלות המשמעויות החשובות ביותר (Stemler, 2000). באמצעות מתודולוגיה זו, נוצרה מטריצת שכיחויות של מונחים בתחום הסיכונים הבנקאיים כפי שמופיעים בדרישות באזל III. מונחים אילו מתייחסים לסיכונים אשראי, נזילות, שוק, תרחישי קיצון ועוד. לבסוף, נעשה שימוש במבחן  $Z$ , כדי לבחון שינויים מובהקים בין התקופות שנדגמו. בדומה למחקר של בלום ורביב (2023), תקופת המדגם ששימשה למחקר כללה את השנים 2007-2018: לפי החלוקה הבאה: לפני תקופת המשבר (2007), במהלך המשבר (2008-2009), במהלך פרסום המלצות באזל III (2010-2013) ובמהלך יישום המלצות באזל III (2014-2018). זיהוי שינוי מובהק בתדירות המונח ביחס לתקופה שקדמה (השינוי המובהק יכול להיות חיובי וגם שלילי ביחס לתקופה שקדמה בין שתי תקופות מצביע על חשיבות אותו מונח ביחס לתקופה שקדמה).

### 3. מידול נושאים

מידול נושאים (topic modeling) שייד לתחום הלמידה הלא מופקחת (unsupervised learning). משפחת המודלים הללו משמשים לכריית טקסטים במטרה לגלות מבנים סמויים בטקסט המאפשרים לנו לזהות את הנושאים העיקריים שהטקסט מדבר עליהם, ואף לזהות שינויים לאורך זמן במשקל הנושאים השונים שחבויים בטקסט. באופן פרטני בלום ורביב (2023) השתמשו באלגוריתם LDA (Latent Dirichlet Allocation). מודל ה-LDA, שהוטמע מבוסס על Gibbs sampling והועדף על האלטרנטיבה של Variational Bayes sampling בשל הדיוק הרב יותר, למרות קצב המימוש האיטי יותר.

במודל ה-LDA מנתחים אוסף של מסמכים הנקראים יחדיו - corpus, כאשר מגדירים מראש את מספר הנושאים שיתקבלו כפלט מהמודל. כמו

#### 4. בחינת כמות הנושאים במידול נושאים

כפי שצוין אחת הסוגיות המרכזיות שצריך להתמודד עמם שמריצים LDA, זה כמות הנושאים שתשמש אותנו כקלט למודל. בכדי לקבוע כמות זו ישנן כמה חלופות, כאשר אחד המדדים המקובלים הוא לבצע אופטימיזציה של מידת הקוהרנטיות (coherency score), שנותן אומדן של עד כמה כל נושא יכול להיות מיוצג על ידי קבוצת המילים המשולבת שבתוכו (Röder et al., 2015; Stevens et al., 2013; Chen et al., 2012). ישנן שיטות שונות לאמידת ה Coherency score, כאשר מבחן הקוהרנטיות כולל מדידת הקשר או הקוהרנטיות של המילים המובילות בכל נושא. לפי תוצאות ההרצה שלנו יצא כי כמות הנושאים האופטימלית עומדת על חמישה נושאים (יותר מידי נושאים מייצרים מילים חופפות בנושאים שונים). הערה: ככל שהמספר גבוה יותר, אז התוצאה טובה יותר, 4, כאשר תוצאות מעל 65% משקפות תוצאות טובות. במודל שנבחר יצא כ-68%, המשקף תוצאה טובה מאוד.

#### 5. ממידול נושאים לדו"חות חשבונאיים

אחד המחקרים הראשונים בתחום של ניתוח טקסט של דוחות כספיים באמצעות מידול נושאים הוא מחקרם של באו ודאטה (Bao et Datta 2014). במחקר זה בחנו את נושאי הסיכון העיקריים, שקיימים בדוחות הכספיים של חברות מובילות באמצעות הדוחות הכספיים השנתיים (10k reports). המחקר שלהם מתמקד בהשוואת איכות המודל שפיתחו, שהנו וריאציה של מודל ה-LDA, ביחס למודל ה LDA המקורי (Latent Dirichlet Allocation), שנרחיב עליו בהמשך. לטענתם המודל שפיתחו משופר יותר ומייצר נושאים ברורים וקוהרנטיים יותר. במחקר של צ'ן ושו"ת (Chen et al 2017) משווים בין גישות שונות למידול נושאים באמצעות ניתוח הדוחות השנתיים של בנקים בארה"ב. מטרת

2.1. בחירת  $\theta_a \sim Dir(\alpha)$  שמשקפת

התפלגות הנושאים לכל מסמך

2.2. לכל מונח/מילה  $j \in \{1, \dots, N_d\}$  :

במסמך  $d$  יש לבצע את

השלבים הבאים :

2.2.1. בחירת הנושא

$$z_{a,j} \sim Multinomial(\theta_a)$$

2.2.2. בחירת מונח/מילה

$$w_{a,j} \sim Multinomial(\phi_{z_{a,j}})$$

3. מכאן שההסתברות שנושא  $k$  מייצר

מילה  $v$  במיקום  $j$  במסמך  $d$  הינה :

$$p(w_{a,j} = v | \alpha, \beta, K):$$

$$p(w_{a,j} = v | \phi_k) p(z_{a,j} = k | \theta_a) p(\phi_k | \beta) p(\theta_k | \alpha) d\theta_k d\phi_k$$

יש לבצע אינטגרציה של וקטור מרחב ההסתברויות באורך  $K(\theta)$  ואורך  $V(\phi)$ , כאשר מטרת ה LDA למקסם את ההסתברות  $p(\theta, Z | \alpha, \beta, K)$  ע"י בחירת  $\theta$  ו-  $Z$  בהינתן קורפוס  $W$  ופרמטרים  $\alpha$  ו-  $\beta$ . למרבה הצער בעיה זו אינה פתירה מיחשובית, כך ש  $\theta$  ו-  $\phi$ , שממקסמים את ההסתברות במשוואה לעיל, משוערכים באמצעות האלגוריתם של ה LDA. הטכניקה המדויקת שמטמיעים, משתנה כפונקציה של המודלים ותתי המודלים אשר נבחרים ושמוצעים בתהליך איטרטיבי (למשל Gibbs sampling ו/או Variational Bayes sampling).

באופן דומה לבחינת שינויים מובהקים שחלו בתקופות השונות כפי שיושם במטריצת ה DTM, ייושם גם ב- LDA מתודולוגיה זו באמצעות מבחני  $Z$ , רק שבמקרה דנן ההתמקדות היא בשינויים במשקל הנושאים השונים שה LDA מצא (לעומת שכיחות המונחים).

<sup>4</sup> יחד עם האמור לעיל, תוצאות שהן גבוהות מאוד, למשל מעל 75% חשודות כתוצאות שאינן אמינות

**טוקניזציה:** (Tokenization) הוא תהליך שבו הטקסט מחולק ליחידות קטנות יותר, הנקראות "טוקנים" או "מילים", כאשר כל טוקן הוא חלק יחיד בטקסט, והוא עשוי להיות מילה, מילת חיבור, רווח או סמל מיוחד.

**למטיזציה:** (Lemmatization) הוא תהליך שבו מורידים מהמילים סיומות ו/או תהליך שבו מביאים את המילה לצורת הבסיס (שורש) שלה בשפה. לדוגמה, מילה כמו "רציתי" תהפוך ל"רצה" ו"ספרים" למילה "ספר" (יחיד לעומת רבים). מטרת שלבי הכנת הנתונים, היא להסיר מונחים עם משמעות נמוכה ולמנוע ספירה כפולה של מילים בעלי אותו השורש. הדוחות הכספיים נלקחו מאתר בנק ישראל, שדווחו באנגלית, כפי שהוצגו לציבור המשקיעים בעולם ולחברות הדירוג. תחום ה-NLP בכל הקשור למידול נושאים טרם הגיע לבשלות בשפה העברית, כאשר בשנים האחרונות חלה התקדמות בנדון הן באקדמיה והן בתעשייה, לאור האמור המחקר התבסס על דוחות באנגלית.

## 7. מבחן Z

בדיקת השערות המחקר תבצע באמצעות מבחן Z לבדיקת השערות סטטיסטיות. כמו במחקר של בלום ורביב (2023), נעזרנו במבחן Z כדי לזהות שינויים מובהקים סטטיסטית בין התקופות. יש לציין כי כאשר בחנו את התדירות של מונחי הסיכון כפי שהוגדרו מראש ובמשקלות הנושאים שחישבנו באמצעות מודל ה-LDA, צצו דפוסים מסוימים. עם זאת, כדי לזהות שינויים מובהקים בתדירות שניתן להגדירם כתנודות "חריגות" לאורך זמן (בשונה מאירועים אקראיים), היה צורך למצוא מובהקות סטטיסטית. לכן, ערכנו מבחן Z כדי לזהות שינויים כאלה ברמת ביטחון של 99%.

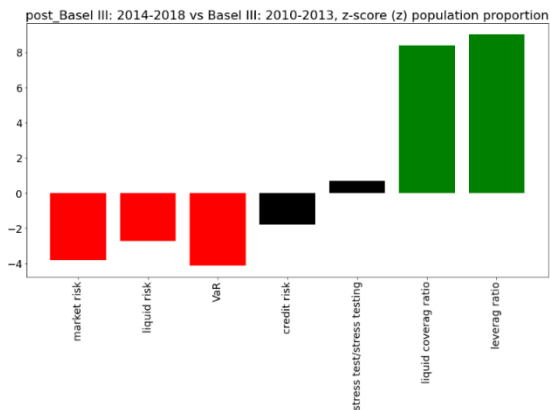
המחקר שלהם לאתר את המודל האפקטיבי ביותר לניתוח הטקסט בדוחות הכספיים של הבנקים בארה"ב ולזהות בנק כושל ביחס לבנק טוב על סמך מידול נושאים. קמפבל ושו"ת (Campbell et al 2014) ודיו ושו"ת (Dyer et al 2017), התמקדו בניית גילויים של גורמי הסיכון בדוחות הכספיים. קמפבל ושו"ת התמקדו בתכנים הקשורים לגורמי סיכון באמצעות ניתוח הדוחות הכספיים השנתיים (10k) ואת הקשר של גורמים אילו לרמות הסיכון של הפירמות. מחקרם של דיו ושו"ת התמקדו בניית גילויים בדוחות הכספיים השנתיים, כולל שינויים באורך הטקסט, בספציפיות ובתוכן המידע. סגל ושו"ת (Zeghal et al. 2018) התמקדו בבחינת החשיפה לניהול סיכונים הוליסטיים (ERM: Enterprise Risk Management). המחקר שלהם בחן את השפעת המשבר הפיננסי על החשיפות במונחי ERM, בדוחות השנתיים של בנקים גדולים בארה"ב, תוך התחשבות בגורמים כמו גודל הבנק, עצמאות הדירקטוריון, מידת הרווחיות וכן מינוף וגודל הדירקטוריון. במחקר של בלום ורביב (Blum and Raviv 2023) עושים שימוש בבחינת שינויים מובהקים במטריצת שכיחות מונחים (DTM) ובמידול נושאים (topic modeling) בתקופות שונות ע"מ לנתח את השפעתם של אירועים מכוונים כמו המשבר הפיננסי והנחיות באזל III על הטקסט בדוחות הכספיים של ארבעת הבנקים הגדולים בישראל. במסגרת מחקר זה זוהו מגמות ברורות בחשיבותם של סיכונים שונים לאורך זמן, כמו למשל ירידה במשקל סיכונים השוק בתקופת המשבר לעומת גידול במשקל סיכונים האשראי.

## 6. נתונים

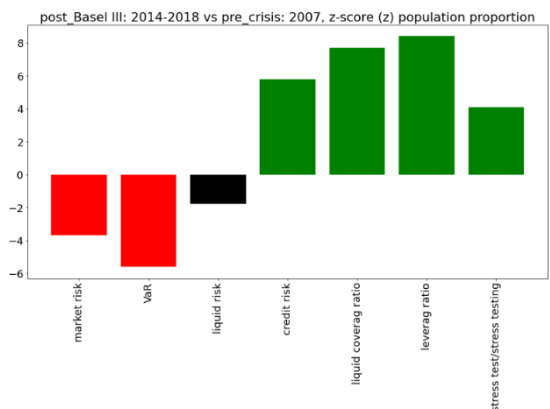
בדומה למחקר של בלום ורביב (2023), נאספו הדוחות הכספיים השנתיים של ארבעת הבנקים המובילים בישראל לשנים 2007–2018 (בנק הפועלים, בנק לאומי, בנק מזרחי-טפחות ובנק דיסקונט). לצורך הכנת הקורפוס לניתוחים סטטיסטיים: הוסרו מילות קישור, בוצע תהליך של טוקניזציה ולמטיזציה.

### 8. תוצאות בחינת שכיחות מונחים

בפרק זה אנו מציגים שינויים מובהקים<sup>5</sup> בשכיחות מונחים באופן סכמתי, שמאפשר לבדוד שינויים אילו בצורה נוחה בכל תקופה. השינויים חושבו ברמת מובהקות של 99% לפי מבחן Z, בין התקופות הבאות: תקופת המשבר (2008-2009) ביחס לתקופת טרום המשבר (2007), תקופת פרסום הנחיות באזל III (2010-2013) ביחס לתקופת המשבר, תקופת פרסום הנחיות באזל III (2014-2018) ביחס לתקופת פרסום הנחיות באזל III וכן תקופת פרסום הנחיות באזל III ביחס לתקופת טרום המשבר (תרשימים המצביעים על שינויים לאורך התקופות של שכיחות המונחים, תרשימים: 2-5).

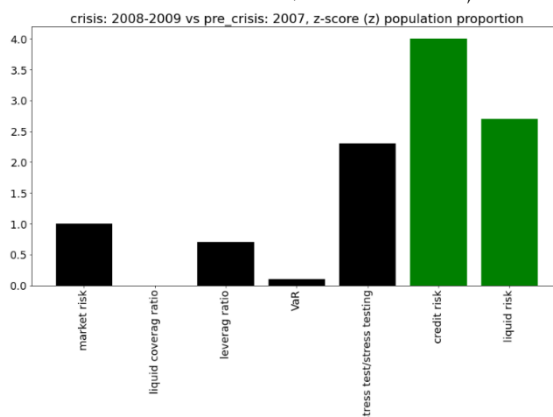


**תרשים 4** - עליה בשכיחות המונחים הקשורים לנזילות שמופיעים בהנחיות באזל III בתקופת יישום הנחיות באזל III ביחס לתקופת פרסום ההנחיות, ומנגד חלה ירידה בשכיחות המונחים הקשורים לסיכוני שוק לרבות הגדרה הישנה של סיכון הנזילות (חל שינוי בטרמינולוגיה לעניין המינוחים הקשורים לנזילות עם פרסום הנחיות באזל III). במקרה דנן עיקר השינוי הנו הירידה בשכיחות מונחים הקשורים לסיכוני שוק.

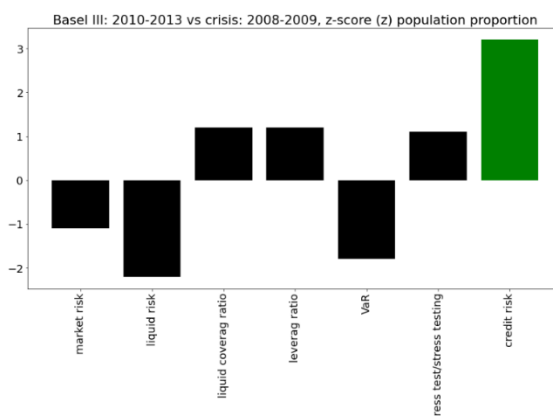


**תרשים 5** - עליה לאורך התקופה הנסקרת, בשכיחות המונחים: סיכוני אשראי, סיכוני נזילות ותרחישי קיצון ומנגד ירידה בשכיחות מונחים הקשורים לסיכוני שוק.

בדומה לממצאים מהמחקר של בלום ורביב (2023) נצפה כי חלה עליה מובהקת במונח "credit risk" לאורך התקופה. כמוכן המונחים "stress testing" ו-"liquidity risk", הפכו לשכיחים יותר, כאשר עם יישום הנחיות באזל III המונח השני הוחלף ב-"liquidity coverage ratio". מנגד אנו רואים, שמונחים הנוגעים לסיכוני שוק כמו "market risk" ו-"VaR" הפכו לפחות שכיחים במהלך התקופות שנדגמו. לסיכום ניתן להצביע על עליה בחשיבות תחום סיכוני האשראי, נזילות ותרחישי קיצון לעומת ירידה בחשיבות תחום סיכוני השוק.



**תרשים 2** - עליה בשכיחות המונחים: סיכוני אשראי וסיכוני נזילות בתקופת המשבר ביחס לתקופת טרום המשבר.

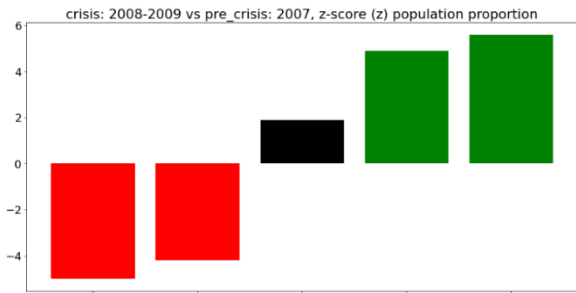


**תרשים 3** - המשך עליה בשכיחות המונח סיכוני אשראי עם פרסום הנחיות באזל III ביחס לתקופה המשבר

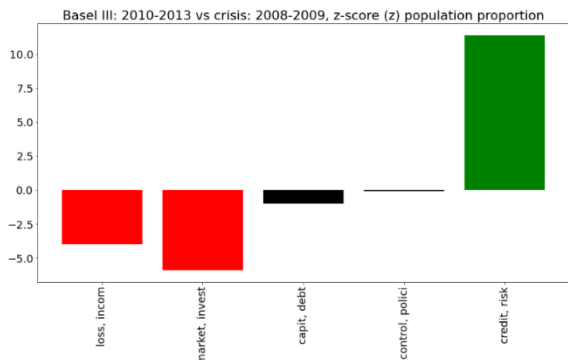
רמת מובהקות 99%, שינוי בשחור משקף שינוי שאינו מובהק ברמת מובהקות 99%.

<sup>5</sup> שינוי בצבע ירוק משקף עליה מובהקת לפי רמת מובהקות 99%, שינוי באדום משקף ירידה מובהקת לפי

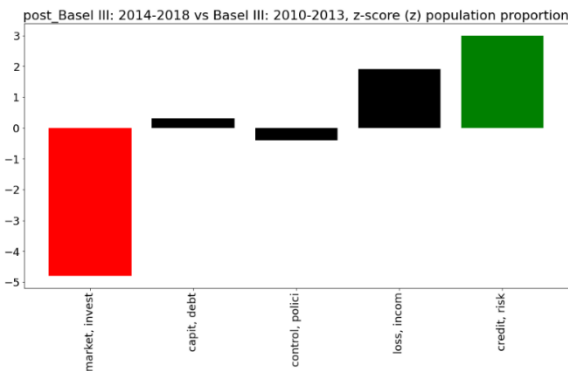
## 9. תוצאות מידול נושאים לפי חמישה נושאים



**תוצאות 6 -** בתקופת המשבר ביחס לתקופת טרום המשבר חלה עליה במשקל נושא סיכוני האשראי ונושא המדיניות והבקרה ומנגד חלה ירידה במשקל נושא סיכוני השוק וההשקעות ונושא הכנסות והפסדים.



**תוצאות 7 -** בתקופת פרסום הנחיות באזל III ביחס לתקופת המשבר מגמת העלייה במשקל נושא סיכוני האשראי המשיכה וכמוכן המשיכה מגמת הירידה בנושא סיכוני השוק וההשקעות ונושא הכנסות והפסדים.



**תוצאות 8 -** בתקופת יישום הנחיות באזל III ביחס לתקופת פרסום הנחיות באזל III, המשיכה מגמת העלייה במשקל נושא סיכוני האשראי וכמוכן המשיכה מגמת הירידה בנושא סיכוני השוק וההשקעות.

בהמשך לבדיקת כמות הנושאים האופטימלית לפי Coherency test הרצנו מודל LDA לפי חמישה נושאים. חמשת הנושאים לפי מודל ה LDA הם : market invest ~ סיכוני שוק והשקעות (משקל שלהם כ 25% מסך המילים שמרכיבות את הנושא), credit risk ~ סיכוני אשראי, loss income ~ הכנסות והפסדים, control ~ מדיניות ובקרה.

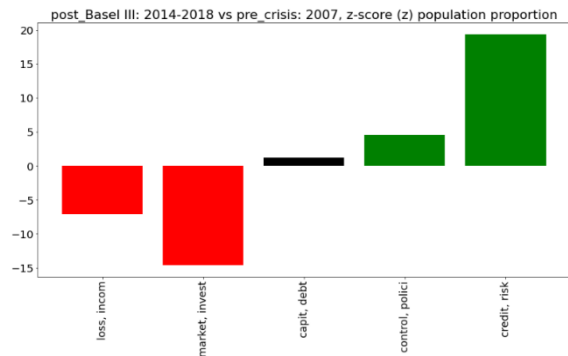
	topic 1	topic 2	topic 3	topic 4	topic 5				
market	13.70%	credit	37.40%	capital	20.30%	loss	16.0%	control	11.40%
invest	11.20%	risk	32.40%	debt	13.10%	income	14.70%	policy	6.50%
stock	2.20%	exposure	8.50%	supervisor	6.90%	profit	14.60%	mortgage	3.60%
global	1.60%	borrow	6.0%	basel	2.10%	provision	12.70%	liquidity	2.40%
labor	0.90%	liquid	1.40%	provision	1.60%	finance	9.0%	model	2.40%
crisis	0.80%	single	0.70%	motion	1.30%	revenue	2.30%	compliance	2.0%

**טבלה 3 -** המילים שמרכיבות את חמשת הנושאים ומשקלם (תרומתם היחסית לכל נושא). כפי שהוצע על ידי Steyver ו- Griffiths ויושמה במספר רב של מאמרים, למשל : Baker ו- Levy, בדקנו את השינוי במשקל הנושאים לאורך זמן על ידי חישוב המשקל הממוצע של כל נושא בכל המסמכים, שפורסמו בתקופה ספציפית. בניתוח לאורך זמן (תרשימים 6-9), ניתן לראות כי נושא סיכוני האשראי הפך להיות חשוב יותר לאורך התקופה, כאשר נושא סיכוני השוק הפך להיות פחות במיקוד של הבנקים כפי שהשתקף בדוחות הכספיים. בנוסף ישנו נושא חדש (הנושא החמישי) שהפך להיות חשוב יותר עם פרוץ המשבר וכן עם יישום הנחיות באזל III ביחס לתקופת טרום המשבר, נושא המדיניות והבקרות (policy control).



(2007), תקופת פרסום הנחיות באזל III (2010-2013) ביחס לתקופת המשבר, תקופת הטמעת המלצות באזל III (2014-2018) ביחס לתקופת פרסום המלצות באזל III. תוצאות מעל 2.58 (או מתחת ל-2.58) משקפות שינויים מובהקים ברמת מובהקות 99%.

ניתן לראות באופן כללי לפי הניתוח המופיע בטבלה 4 כי ברוב המקרים מידת ההומוגניות במערכת הבנקאית גבוהה, כאשר בשונה מהמערכת, עם פרוץ המשבר בנק לאומי הושפע באופן מובהק ביחס ליתר הבנקים בכל הקשור לעליה בחשיבות המונחים "סיכוני אשראי" ו"תרחישי קיצון", בעוד ביתר הבנקים לא נצפה שינוי מובהק. ניתן לנסות ולהסביר זאת בכך, שבבנק לאומי שנחשב לבנק השני בגודלו, נחשפו בין השאר ל-Lehman Brothers שפשט רגל במשבר<sup>6</sup> ובהתאם תיק האשראי נפגע. בתקופה של פרסום המלצות באזל III ביחס לתקופת המשבר, מרבית הבנקים התנהגו באופן דומה, למעט בנק הפועלים שבו עלתה חשיבות המונח "credit risk", כאשר חלה ירידה בשכיחות המונחים VaR ותרחישי קיצון בבנק לאומי. כלומר בהתייחס לשתי תקופות אילו ניתן לזהות כי הבנקים הגדולים הושפעו באופן שונה ביחס ליתר המערכת. ממצא זה מעניין ויכול לנבוע מהתנהלות שונה שנהוגה בבנקים אילו וכן מרגולציות עודפות בבנקים שעונים לקטגוריה של Too Big To Fail,



**תרשים 9** – מצביע על עליה לאורך התקופה הנסקרת, במשקל נושא סיכוני אשראי ובנושא מדיניות ובקרה לעומת ירידה בנושא סיכוני שוק והשקעות ובנושא רווח והפסד. תוצאות אילו בהתייחס לסיכוני אשראי וסיכוני שוק תואמים את התוצאות שעולות מניתוח שכיחות המונחים בתרשימים 2-5.

### 10. השוואת המערכת הבנקאית

במחקר של בלום ורביב (2023) בחנו את המערכת כמקשה אחת, בעבודה זו אנחנו מרחיבים את הניתוח ובוחנים גם את מידת ההומוגניות במערכת הבנקאית בכל הקשור להשפעת המשבר ובאזל III על הבנקים במימד של שכיחות מונחים ומידול נושאים. אנליזה זאת חשובה כדי להבין כיצד בנקים שהם מהותיים במערכת הגיבו לאירועים ולתקנות לעומת בנקים בינוניים וקטנים יחסית.

תוצאות השוואת המערכת הבנקאית לפי שכיחות

#### מונחים

בנק	מונחים	crisis: 2008-2009 vs pre_crisis: 2007	Basel III: 2010-2013 vs crisis: 2008-2009	post_Basel III: 2014-2018 vs Basel III: 2010-2013	post_Basel III: 2014-2018 vs pre_crisis: 2007
בנק פועלים	credit risk	1.2	4.3	-1	4.6
		2.8	-0.8	2.6	4.5
		2.3	1.2	-4.9	-0.8
		1.5	2.5	-0.9	3.4
בנק לאומי	market risk	0.3	0.6	-1.8	-0.8
		0.9	-2.7	-0.5	-2.1
		0.4	1.3	-4.1	-2.1
		0.6	-1	-1.7	-2.1
בנק דיסקונט	VaR	0	0.2	-0.4	-0.1
		0.5	-2.7	-1.8	-3.7
		-0.7	0.1	-3.2	-3.8
		0.6	-1	-2.8	-3.1
בנק מרחי טפחות	liquidity risk	0.8	-0.5	-2	-1.6
		2	-1.5	0	0.7
		1.4	-0.8	-3.7	-2.9
		1.1	-1	0.1	0.3
בנק פועלים	stress test/stress testing	-0.1	1.2	1.7	2.5
		3.1	-2.3	0.3	1.5
		0	3.3	-2.2	1.4
		0.4	0.8	1.9	3

**טבלה 4** - שכיחות מונחים:

הערה: אנו מחשבים את שכיחות מונחי הסיכון בארבע תקופות שונות: תקופת המשבר (2008-2009) ביחס לתקופת טרום המשבר

<sup>6</sup> דוח בנק ישראל משנת 2007 בפרק ג' "הסיכונים והלימות ההון"

תוצאות השוואת המערכת הבנקאית לפי מידול

LDA נושאים

בנק	שאל	crisis: 2008-2009 vs pre_crisis: 2007	Basel III: 2010-2013 vs crisis: 2008-2009	post_Basel III: 2014-2018 vs Basel III: 2010-2013	post_Basel III: 2014-2018 vs pre_crisis: 2007
בנק פועלים	capit, debt	3.9	-0.8	-0.6	2.8
בנק לאומי		0.5	-0.8	1.5	1.2
בנק דיסקונט		1	-1.1	-0.2	-0.2
בנק מרחי טפחות		-0.2	-0.5	0.3	-0.4
בנק פועלים	market, invest	-3.1	-1.5	-3.6	-8.1
בנק לאומי		-0.7	-3.3	-1.2	-5
בנק דיסקונט		-3.4	-1.4	-0.3	-4.9
בנק מרחי טפחות		-1.4	-2.8	-1.6	-5.7
בנק פועלים	credit, risk	0.6	5.3	2.9	8.3
בנק לאומי		4.5	5.8	-0.6	9.2
בנק דיסקונט		4.8	4	-1.6	6.8
בנק מרחי טפחות		3.6	4.3	0.9	8.7
בנק פועלים	loss, incom	-1.5	-3.3	0.7	-4.1
בנק לאומי		-4.2	-1.7	0	-5.6
בנק דיסקונט		-2.4	-0.5	2.8	0
בנק מרחי טפחות		-1.8	-1.2	1.8	-1.5
בנק פועלים	control, polici	1	0.6	1.2	2.7
בנק לאומי		1.1	0.5	0.8	2.3
בנק דיסקונט		0.9	-1.5	-1	-1.5
בנק מרחי טפחות		0.1	0.7	-2.3	-1.4

סיכוני האשראי. למעט בנק הפועלים יתר המערכת הגיבה באופן "אדיש" בתקופת המשבר, בכל הקשור לנושא ההון והחוב (capital debt), כאשר בבנק הפועלים חלה עליה מובהקת בתקופה זו ביחס לתקופת טרום המשבר. בנושא הכנסות והפסדים (loss and income) עם פרוץ המשבר חלה ירידה חדה בבנק לאומי בעוד שביתר המערכת לא חל כל שינוי מובהק, כאשר בתקופה שלאחר מכן חלה ירידה חדה גם בבנק הפועלים. יתרה מזו כאשר משווים בין תקופת יישום המלצות באזל III לתקופת טרום המשבר, רק בבנק הפועלים ובבנק לאומי חלה ירידה מובהקת בין התקופות, קרי, מזהים התנהגות דומה בקרב הבנקים הגדולים בתקופות עוקבות ביחס ליתר המערכת בכל הקשור לנושא זה. בנושא המדיניות והבקרות (policy control) לא ניתן לזהות שינויים בין הבנקים בתקופת המשבר ועם פרסום הנחיות באזל III, קרי, ההתנהגות הומוגנית בתקופות אילו. יחד עם האמור לעיל ניתן לראות, כאשר משווים בין תקופת יישום המלצות באזל III ביחס לטרום המשבר רק בבנק הפועלים ולאומי (שני הבנקים הגדולים במשק) מידת החשיבות עלתה בעוד שביתר הבנקים לא חל שינוי מובהק. באופן עקרוני קיימת מידה גבוהה של הומוגניות בחלק ניכר מהנושאים בעוד שבבנקים הגדולים ביחס ליתר הבנקים לעיתים מזהים התנהגות שונה.

**טבלה 5 - טבלה מצביעה על שינויים, שחלו במשקל חמשת הנושאים לפי מודל ה-LDA: market invest ~ סיכוני שוק והשקעות, credit risk ~ סיכוני אשראי, capital debt ~ הון וחבות, loss income ~ הכנסות והפסדים, policy control ~ מדיניות ובקרה לאורך התקופות השונות. הטבלה מאפשרת לנו להבין את ההשפעות של המשבר ובאזל III בתקופה הנסקרת על משקל הנושאים שצוינו. מה שמסומן בירוק משקף שוני ביחס ליתר המערכת בכל אחד מהסיכונים, כאשר הסימון בצהוב משקף דמיון. הערה: אנו מחשבים את משקל הנושאים בארבע תקופות שונות: תקופת המשבר (2008-2009) ביחס לתקופת טרום המשבר (2007), תקופת פרסום הנחיות באזל III (2010-2013) ביחס לתקופת המשבר, תקופת הטמעת המלצות באזל III (2014-2018) ביחס לתקופת פרסום המלצות באזל III. תוצאות מעל 2.58 (ואו מתחת ל-2.58) משקפות שינויים מובהקים ברמת מובהקות 99%.**

ניתן לראות באופן כללי לפי הניתוח המופיע בטבלה 5 כי בנושא סיכוני אשראי (credit risk) המערכת די הומוגנית למעט בנק הפועלים. בסיכוני אשראי המערכת הגיבה באופן מידי עם פרוץ המשבר הפיננסי, למעט בנק הפועלים (שהגיב תקופה לאחר מכן) בהתמקדות בתחום של סיכוני אשראי, כאשר המגמה המשיכה לאורך כל תקופת המדגם לכלל הבנקים. מנגד בתחום של סיכוני שוק והשקעות ניתן לראות כי עם פרוץ המשבר חלה ירידה מיידית במשקל נושא זה בבנק הפועלים ובבנק דיסקונט בעוד שבבנק מרחי ובבנק לאומי הירידה במשקל סיכוני השוק, באה לידי ביטוי בתקופה שלאחר מכן (תקופת פרסום הנחיות באזל III). יחד עם האמור לעיל, בכל הבנקים חלה ירידה במשקל סיכוני השוק בתקופת יישום המלצות באזל III ביחס לתקופת טרום המשבר לעומת עליה במשקל

Non-negative Matrix Factorization (NMF) היא טכניקת הפחתת ממדים ופירוק לגורמים המשמשת ללמידת מכונה וניתוח נתונים. הרעיון מאחורי ה NMF זה לפרק מטריצת נתונים נתונה לשתי מטריצות בעלות ממדיות נמוכות יותר כך שהמרחק (אוקלידי או פרובניוס) בין המטריצה המקורית לבין המכפלה של המטריצות החדשות שנוצרו יהיה מינימלי, תחת האילוץ שכל הערכים אינם שליליים. בהקשר של מידול נושאים, NMF משמש לגילוי נושאים החבויים באוסף של מסמכים כפי שגם מודל ה LDA עושה, כאשר מטריצה אחת מייצגת התפלגות מילים בתוך הנושאים והשנייה מייצגת את התפלגות הנושאים במסמכים, ובסופו של דבר מייצרת נושאים קוהרנטיים באמצעות ניתוח הטקסט. בשונה מה LDA שמבוסס על התפלגות Dirichlet ושיטות איטרטיביות שונות (למשל Gibbs sampling), כאן מדובר בפירוק מטריצות (גישה ליניארית, שמבוססת על אחת מהשתיים: מודל Generalized Kullback-Leibler Divergence לצורך זיהוי שוני בין התפלגויות שונות או Frobenius norm לצורך חישוב מרחקים אוקלידיים באמצעות SVD – לכן היא יחסית מהירה יותר. כחלק ממבחני האיכות, נהוג להריץ יותר ממודל אחד בכדי לחזק את התוצאות שיצאו. כפי שניתן לראות בטבלה 6, בדומה למחקר של בלום ורביב (2023), כי קיים דמיון רב בכל הנושאים במילים המובילות (המילים הראשונות בכל נושא הם בעלי המשקל הגדול ביותר ביחס ליתר המילים, ולכן ברגע שמוזהים דמיון במילים המובילות אזי ניתן לקבוע כי התוצאות איכותיות).

הסיבות לכך שבנושא סיכון אשראי המערכת הבנקאית בישראל היא די הומוגנית למעט בבנק הפועלים:

**הרגולציה:** בנק ישראל הטיל על הבנקים בישראל דרישות מחמירות בנושא סיכון אשראי. דרישות אלה, אשר הופעלו במסגרת יישום באזל III, הביאו לכך שהבנקים בישראל הגדילו את ההון העצמי שלהם כדי להקטין את הסיכונים שלהם. **התעשייה:** מערכת הבנקאות בישראל היא יחסית קטנה ויציבה. גורמים אלה הובילו לכך שהבנקים בישראל נוטים להיות פחות חשופים לסיכונים מאשר בנקים במדינות אחרות. באשר לבנק הפועלים, ניתן להעריך כי ההסבר להבדלים בתחום סיכוני האשראי ותחום ההון, הוא הגודל וההיקף של הבנק ומידת הגלובליות שלו. בנק הפועלים הוא הבנק הגדול בישראל, והוא מספק שירותים פיננסיים ללקוחות מגוונים. גודל זה והיקף הפעילות של הבנק הופכים אותו ליותר חשוף לסיכונים מאשר בנקים קטנים יותר. בנוסף, בנק הפועלים ובנק דיסקונט הם היחידים בישראל שיש להם פעילות משמעותית בחו"ל, כאשר בנק הפועלים שנתפס בזמנו כבנק המתוחכם ביותר, בחרו להשקיע כמו חלק ניכר מהבנקים המובילים בארה"ב ובאירופה, במוצרים הקשורים למשבר הסאב פריים, שהתגלו כמסוכנים, ובהתאם החוב של הבנק ויחס הלימות ההון נפגעו, כאשר חלה עליה מובהקת במשקל נושא זה במודל ה-LDA.

### 11. השוואת תוצאות הנושאים לפי LDA לעומת NMF

Topic	LDA#5	NMF#2	LDA#2	NMF#3	LDA#1	NMF#4	LDA#4	NMF#5	LDA#3
financial	control	credit	credit	risk	market	Income	loss	capital	capital
control	policy	loss	risk	market	invest	Profit	income	market	debt
market	mortgage	provision	exposure	exposure	stock	Loss	profit	profit	supervisor
invest	liquidity	debt	borrow	control	global	Expense	provision	supervisor	Basel
exposure	model	borrow	liquidity	policy	labor	provision	financial	invest	provision

טבלה 6 - השוואה בין תוצאות מודל LDA לתוצאות מודל NMF במידול נושאים. ניתן לראות כי יש דמיון רב בתוצאות.

**12. סיכום**

במחקר זה סקרנו מודלים המאפשרים ניתוח דוחות כספיים באמצעות ניתוח הטקסט לצורך זיהוי דפוסי התנהגות בתחום ניהול סיכונים. התמקדנו במחקרם של בלוס ורביב (2023), שבו בחנו את השפעות המשבר הפיננסי ובאזל III על תחום ניהול הסיכונים בארבעת הבנקים המובילים בישראל. במסגרת המחקר בדומה למחקר של בלוס ורביב (2023), זוהו מגמות סטטיסטיות מובהקות בתדירות של מונחי סיכון שונים ובמשקל הנושאים שחולצו ע"י מודל ה-LDA, כפי שהשתקפו בדוחות הכספיים. המחקר התמקד בזיהוי שינויים מובהקים שהתרחשו בארבע תקופות: תקופת טרום המשבר 2007, תקופת המשבר 2008–2009, תקופת פרסום באזל III 2010–2013 ותקופת יישום באזל III 2014–2018. המחקר הצביע על כך כי השימוש בכלים מתקדמים בתחום ה-AI ובפרט בתחום ה-NLP הוסיפו ממד נוסף בהבנת השפעות המשבר הפיננסי של 2008 והנחיות באזל III על אופן ניהול הסיכונים בבנקים ישראלים מובילים. בלוס ורביב (2023) מצאו כי הבנקים שינו את המיקוד שלהם מ"סיכון שוק" וספציפית מ-"VaR" להתמקדות ב-"סיכון אשראי". מגמות אלה המשיכו גם בתקופת פרסום ויישום באזל III. אף על פי שהמשבר הוביל לעלייה מיידי בתדירות המונח "סיכון נזילות", ניתן היה לזהות נסיגה במגמה זו בתקופה שלאחריו. במסגרת המחקר התמקדנו בבחינת מידת ההומוגניות במערכת הבנקאית בישראל, כאשר הממצאים מצביעים על כך כי במרבית הנושאים ישנה מידה גבוהה של הומוגניות בין הבנקים, יחד עם האמור שני הבנקים הגדולים הגיבו אחרת למשבר בהתייחס למונחי סיכון ובהתייחס לנושאים

מסוימים ביחס ליתר הבנקים. לפי בחינת השינויים המובהקים במונחי הסיכון, נמצא כי עם פרוץ המשבר בנק לאומי התמקד יותר בתחום סיכוני האשראי ותרחישי קיצון ביחס ליתר המערכת, בעוד שעם פרסום באזל III, בנק הפועלים התמקד יותר בתחום של סיכוני אשראי בעוד שבתקופה זו בנק לאומי הפחית את מידת ההתמקדות שלו בתחום סיכוני השוק וה-VaR. לפי מודל ה-LDA בנק הפועלים התמקד בעת תקופת המשבר בנושא החוב וההון. ייתכן והבדלים אילו נובעים מכך שבגין זאת שבנק הפועלים נתפס כבנק הגדול ביותר עם חשיפה גבוהה לשווקים בינלאומיים לרבות השקעה ישירה ברכיבים מסוכנים בתחום אג"ח מגובה משכנתא<sup>7</sup>, ביחס למרבית הבנקים המקומיים, בהתאם החוב ויחס הלימות ההון שלו נפגעו וכפועל יוצא הופעלו עליו רגולציות נוספות. ייתכן גם בגין החשש לאפקט הדבקה אזי הייתה רגישות יתר בכל הקשור למדיניות ניהול הסיכונים שם. בבנק השני בגודלו, שכיחות המונח "סיכוני אשראי" עלה באופן מובהק, כאשר בבנק זה נחשפו בעת תקופת המשבר באופן ישיר להשקעות בעיתיות גבוהות לבנקים בארה"ב שכשלו (למשל: Lehman Brothers). המחקר חושף את ציבור הקוראים, למתודולוגיות חדשות מבוססות AI בתחום ה-NLP לבחינת הדוחות הכספיים מנקודת מבט של ניתוח טקסטואלי. כלים אילו מאפשרים לבדוד השפעות של אירועים ספציפיים כפי שהוצגו במחקר זה, שקשה לבדוד אותם במודלים המסורתיים ובהתאם מציפים מידע חדש להבנת תופעות כלכליות ופיננסיות שונות שעד כה לא נחשפנו להם. **ממצאים אלו עשויים להיות בעלי השלכות משמעותיות עבור חוקרים, רגולטורים וגופים פיננסיים.**

<sup>7</sup> ראה נספח: דוח בנק ישראל חשיפות בנקים למשבר

## רשימה ביבליוגרפית

- Al-Harbi, A., 2017. Determinants of banks' liquidity: evidence from OIC countries. *J. Econ. Adm. Sci.* 33 (2), 164–177.
- Bao, Y., Datta, A., 2014. Simultaneously discovering and quantifying risk types from textual risk disclosures. *Manag. Sci.* 60 (6), 1371–1391.
- Baker, H.K., Kumar, S., Goyal, K., Sharma, A., 2021. International review of financial analysis: a retrospective evaluation between 1992 and 2020. *Int. Rev. Financ. Anal.* 78, 101946.
- Basel Committee on Banking Supervision (2018). Disclosure requirements. [https://www.bis.org/basel\\_framework/chapter/DIS/10.htm](https://www.bis.org/basel_framework/chapter/DIS/10.htm).
- Blei, D.M., Ng, A.Y., Jordan, M.I., 2003. Latent Dirichlet Allocation. *J. Mach. Learn. Res.* 3, 993–1022.
- Blei, D.M., 2012. Probabilistic topic models. *Communications. ACM* 55 (4), 77–84.
- Blum, A., Raviv A., 2023, [The effects of the financial crisis and Basel III on banks' risk disclosure: A textual analysis](#), *Finance Research Letters* Volume 53, May 2023, 103680.
- Campbell, J.L., Chen, H., Dhaliwal, D.S., Lu, H.M., Steele, L.B., 2014. The information content of mandatory risk factor disclosures in corporate filings. *Rev. Account. Stud.* 19 (1), 396–455.
- Chen, Y., Rabbani, R., Gupta, A., Zaki, M., 2017. IEEE Symposium, Comparative text analytics via topic modeling in banking.
- Chen, Z., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., Ghosh, R., 2013. Exploiting domain knowledge in aspect extraction. In: *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 1655–1667.
- Dyer, T., Lang M., Lawrence, L., The evolution of 10-K textual disclosure: Evidence from Latent Dirichlet Allocation. *Journal of Accounting and Economics* Volume 64, Issues 2–3, November 2017, Pages 221-245.
- Griffiths, T.L., Steyvers, M., 2004. Finding scientific topics. *Proc. Natl. Acad. Sci. USA* 101 (suppl. 1), 5228–5235.
- Levy, D., Mayer, T., Raviv, A., 2022. Economists in the 2008 financial crisis: slow to see, fast to act. *J. Financ. Stab.* 60, 100986.
- Stemler, S., 2000. An overview of content analysis. *Pract. Assess. Res. Eval.* 7 (1), 17.
- Stevens, K., Kegelmeyer, P., Andrzejewski, D., & Buttler, D. (2012). Exploring topic coherence over many models and many topics. In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* (pp. 952–961).

Zeghal, D., El Aoun, M., 2016. The effect of the 2007/2008 financial crisis on enterprise risk management disclosure of top US banks. J. Mod.

פרק ג' "סיכונים והלימות ההון" דוח בנק ישראל 2007 עמוד 115

פרק ג' "סיכונים והלימות ההון" דוח בנק ישראל 2008 עמוד 123

## נספחים

## נספח: פרק ג' "סיכונים והלימות ההון" דוח בנק ישראל 2007 עמוד 115

השפעת משבר הסאב פריים על הלימות ההון ניכרה בעיקר ב"בנק פועלים": בדצמבר 2007 רשם הבנק הפסדים קבועים של 1.176 מיליארדי ש"ח, הנזקים לדוח הרווח וההפסד, שבעטיים גידולו של הרווח היה מתון מגידולם של נכסי הסיכון. נוסף על כך רשמו חמשת הבנקים הגדולים במחצית השנייה של 2007 הפסדים המסוגלים כהפסדים זמניים בסך 961 מיליוני ש"ח, אשר נזקפים כירידה בשווי ההגון של ההון העצמי (תיבה ג'-1). בהפסדים אלו, למרות סיווגם כהפסדים זמניים בלבד ואף שאינם מופחתים בדוח הרווח וההפסד, גלום פוטנציאל להפוך בעתיד להפסדים קבועים.

## נספח: פרק ג' "סיכונים והלימות ההון" דוח בנק ישראל 2008 עמוד 123

(2) התפתחות הרכב יתרת תיק ניירות הערך והסיכון בו מפרוץ המשבר ועד סוף שנת 2008 המשבר הפיננסי, שהחל במחצית השנייה של 2007, הביא בשלב הראשון לצניחת ערכם של ניירות הערך המגובים בנכסים – אג"ח מגובות משכנתאות (MBS) ואג"ח מגובות נכסים שאינם משכנתאות (ABS) – ולכן ירד ערכם של ניירות אלו, ירידה שחלקה נזקף לקרן ההון, וחלקה – לחשבון הרווח והפסד. התפתחות זו השפיעה במיוחד על "בנק הפועלים", שרווחיותו נפגעה ברביע האחרון של 2007 בשל ירידת הערך של ניירות ערך מגובי נכסים.

ברביע השני של שנת 2008 החשיפה של הבנקים לאג"ח שאינן ממשלתיות, המזוהות עם סיכון גבוה, יחסית להשקעות באג"ח ממשלתיות, ירדה ירידה מהותית ושבה לרמה של סוף שנת 2005. הירידה נבעה מהחלטת דירקטוריון "בנק הפועלים" לממש את תיק ה-MBS המוחזק על ידי סניף ניו-יורק<sup>30</sup>; היקף ההשקעה ביום המכירה עמד על 3.4 מיליארדי דולרים, ובגינה רשם הבנק הפסד ניכר של כ-3.1 מיליארדי ש"ח.

ברביע השלישי של 2008, בעקבות ההחרפה במשבר הפיננסי העולמי, שהביאה לפשיטת רגל של גופים פיננסיים ומדינות, הייתה עלייה ניכרת של פרמיות הסיכון, שבאה לידי ביטוי בעלייה חדה של מירווחי האג"ח בכלל השווקים, הן בחו"ל והן בארץ. התפתחות זו הביאה להתממשות סיכונים בנכסים שעד כה נתפסו כלא מסוכנים – השקעות בניירות ערך של בנקים ומוסדות פיננסיים זרים שאינם מגובי נכסים. השקעות אלו, מרביתן בתיק הזמין למכירה ברכיב "אג"ח אחרות", רשמו במחצית השנייה של השנה הן "ירידות ערך בעלות אופי אחר מזמני", שנזקפו לדוח הרווח וההפסד, והייתה להן השפעה ניכרת על רווחיות הבנקים, והן ירידות ערך "בעלות אופי זמני", שנזקפו לקרן ההון<sup>31</sup>. יש להדגיש שקיימת שונות בין הבנקים ברמת החשיפה לבנקים ומוסדות פיננסיים זרים, וכתוצאה ישירה מכך – גם בהפסדים. עיקר ההפסדים נרשמו "בבנק לאומי", בשל החשיפה ל-Lehman Brothers ו-Washington Mutual. העלייה החדה של מירווחי האג"ח בארץ הביאה גם לירידת ערך של אג"ח קונצרניות<sup>32</sup> שהונפקו בישראל, אך הודות להיקף המצומצם יחסית של החשיפה, כ-5 מיליארדי ש"ח, השפעה זו על הסיכון של תיק ניירות ערך כולו מועטה