

סדרת ניירות עבודה
WORKING PAPER SERIES

מס' 55 No.

אמידת תחזיות מרחביות לשיעורי פשיעה בישראל

Estimation of temporal - spatial forecasts
for crime rates in Israel

יורי גובמן*, ניר פוגל*, דמיטרי רומנוב*, בשורה רגב**, שי עמרם**

Yury Gubman*, Nir Fogel*, Dmitri Romanov*

Besora Regev**, Shai Amram**

כסלו תשע"א, נובמבר 2010 November

* הלשכה המרכזית לסטטיסטיקה
** משטרת ישראל, אגף תכנון וארגון

* Central Bureau of Statistics

** Organization & Planning Department, Israeli Police

L H B E N

הוצאת הלשכה המרכזית לסטטיסטיקה, רח' כנפי נשרים 66, פינת רח' בקי,

ת"ד 34525, ירושלים 91342

טל': 02-6592666; פקס: 02-6521340

אתר הלמ"ס באינטרנט: www.cbs.gov.il

דואר אלקטרוני: info@cbs.gov.il

הלשכה המרכזית לסטטיסטיקה (הלמ"ס) מעודדת מחקר המבוסס על נתוני הלמ"ס. פרסומי תוצאות מחקרים אלו אינם פרסומים רשמיים של הלמ"ס, והם לא עברו את הביקורת שעוברים פרסומים רשמיים של הלמ"ס. הדעות והמסקנות המתבטאות בפרסומים אלו, כולל בפרסום זה, הן של המחברים עצמם ואינן משקפות בהכרח את הדעות והמסקנות של הלמ"ס. פרסום מחדש של העבודה, כולה או מקצתה, טעון אישור מוקדם של המחברים.

רחוב כנפי נשרים 66 פינת רחוב בקי, גבעת שאול, ת"ד 34525, ירושלים 95464 טלפון: 02-6592666, פקס' 02-6521340

דואר אלקטרוני: info@cbs.gov.il כתובת האתר: www.cbs.gov.il

תקציר

בחירת שיטות לצרכי אמידת התחזיות לשיעורי פשיעה מסוגים שונים, לפי אזורים גיאוגרפיים, מהווה אתגר מחקרי חשוב, כאשר יש לקחת בחשבון מספר רב של משתנים חברתיים-כלכליים המשפיעים על רמת הפשיעה. כמו כן, רמת הפשיעה מושפעת מפעילות משטרתית היוצרת הרתעה, ולכן יש לכלול במודל לחיזוי את מגוון מאפייני השיטור. התחזית נבנית עבור יחידות גיאוגראפיות שונות: ישוב/רשות מקומית ושטח המוגדר כתחום פעילותה של תחנת המשטרה, כאשר שטח זה יכול לכלול יישוב אחד או מספר יישובים, ומהווה בדרך כלל יחידה גיאוגראפית מספיק גדולה. בניית תחזית מרחבית מעמידה אתגר נוסף - אמידת התלות המרחבית הקיימת בין שיעורי הפשיעה באזורים שונים, ושינוי במבנה התלות המרחבית לאורך זמן. תחזית מרחבית עשויה להיות שימושית במיוחד למשטרת ישראל במטרה לייעל את תכנון הפעילות המשטרתית לתוך בינוני וארוך, ולסייע לדרג הפיקודי של משטרת ישראל לקבוע פריסה אופטימאלית של שוטרים בין תחנות המשטרה.

נבחנה השפעה של מספר גורמים המשפיעים על שיעורי הפשיעה מסוגים: אלימות ועבירות רכוש עיקריות. נאמדה ההשפעה של הקצאת כוחות שיטור בין תחנות, לפי מקצוע, תוך ניטרול האנדוגניות הידועה בין השיטור לפשיעה בתקופת הבסיס, הנובעת מכך שהקצאת כוח שיטור לפי אזורים תלויה ברמת הפשיעה באותה נקודת הזמן.

בין החידושים העיקריים במאמר: אמידת מודלים לתחזיות מרחביות לטווח קצר, בינוני וארוך, עבור שיעורי פשיעה מסוגים שונים, ברזולוציות גיאוגראפיות שונות; שימוש במודלים מרחביים ובמשתנים מרחביים לצרכי חיזוי שיעורי הפשיעה, תוך אמידת סטיות התקן לתחזיות; השוואה בין מגוון רחב של מודלים פרמטריים לחיזוי שיעורי הפשיעה בטווחי זמן שונים. בנוסף, נעשה שימוש בשיטה בלתי ישירה לאמידת תחזיות של יחידות גיאוגראפיות מצרפיות, כאשר היתרון של השיטה הבלתי ישירה נעוץ ביכולת התחשבות בשוני במאפייני האוכלוסייה וברמות הפשיעה בין היישובים המרכיבים את התחנה. נאמדו והשוו סטיות תקן לתחזיות המתקבלות מכל אחת מהשיטות הנ"ל.

עבור תחזיות לטווח קצר (רבע) נעשה שימוש במתודולוגיה לניתוח סדרות עתיות הכוללת אמידת מודל ARIMA, ניכוי עונתיות והתאמת הסדרות ללוח העברי - אמידת השפעת החגים העבריים וימי הפעילות הנהוגים במשק ישראלי.

מילות מפתח: חיזוי פשיעה, מודל מרחבי, אינדיקטורים מובילים, חיזוי סדרה עתית

תודתנו נתונה ל :

- פרופ' מיכאל בינשטוק, האוניברסיטה העברית בירושלים - על רעיונות בנושא אמידה מרחבית ;
- ד"ר לריסה פלישמן, האוניברסיטה העברית בירושלים, הלשכה המרכזית לסטטיסטיקה - על הערות מתודולוגיות בנושא נתונים מרחביים ואפשרויות אמידת התלות המרחבית ;

תוכן עניינים

7	1. כללי.....
7	1.1 מבוא.....
7	1.2 חידושים מרכזיים במחקר הנוכחי.....
8	2. סקירת שיטות לחיזוי פשיעה.....
11	3. מודלים להסבר וחיזוי פשיעה: מתודולוגיה.....
11	3.1 אמידת תלות מרחבית.....
13	3.2 מודל OLS אוטורגרסיבי.....
15	3.3 מודלים מרחביים.....
15	3.3.1 מודלים SER.....
18	3.3.2 מודל SGLS.....
18	3.4 מודלים להסבר שיעורי פשיעה - ניתוח גמישויות.....
19	3.5 אמידת תחזיות שיעורי הפשיעה וסטיות תקן לתחזיות.....
20	3.6 שיטה בלתי ישירה - חישוב תחזיות ואמידת סטיות תקן.....
21	3.7 מדדי טיב החיזוי.....
22	3.8 אמידת תחזית רבעונית של שיעורי הפשיעה - ניתוח סדרה עתית.....
22	3.8.1 תהליך ניכוי עונתיות ואמידת השפעת חגים וימי פעילות.....
23	3.8.2 בניית תחזיות על סמך מרכיבי הסדרה החזויים.....
24	4. נתונים וסטטיסטיקה תיאורית.....
24	4.1 בסיסי נתונים.....
26	4.2 סטטיסטיקה תיאורית של נתוני פשיעה בישראל.....
27	4.2.1 ישובים יהודים ומעורבים.....
27	4.2.2 ישובים ערביים.....
28	4.2.3 תחנות משטרה.....
28	5 תוצאות אמידה.....
28	5.1 מודלים להסבר שיעורי פשיעה.....
29	5.1.1 מודלים להסבר שיעורי פשיעה ברמת ישוב/מועצה.....
30	5.1.2 מודלים להסבר שיעורי פשיעה ברמת תחנות משטרה.....
31	5.2 מודלים לחיזוי - בחירת המודל הטוב ביותר והתחזיות המתקבלות.....
33	5.3 תחזיות ממודל לסדרה עתית.....
34	6. דיון ומסקנות.....
37	7. רשימה ביבליוגרפית.....
39	רשימת נספחים.....

1. כללי

1.1 מבוא

תחזית פשיעה נחוצה לתכנון פריסת כוחות השיטור במסגרת מגבלות כוח אדם, בהתאם לזיהוי אזורים בעלי פוטנציאל גבוה להתגברות הפשיעה. בניית תחזית פשיעה מהווה אתגר רציני בתחום חיזוי תופעות חברתיות בגלל ריבוי גורמים המשפיעים על תופעת הפשיעה ובהינתן מבנה התלות של התצפיות. כך, בנוסף למגוון גורמים חברתיים, כלכליים ודמוגרפיים הקשורים לפשיעה, קיים קשר מרחבי, דהיינו כאשר שיעורי הפשיעה הנצפים באזור מסוים תלויים בשיעורי הפשיעה באזורים סמוכים. יתרה מזאת, רמת הפשיעה בהווה מתואמת עם שיעורי הפשיעה בעבר באותו האזור או/ו באזורים השכנים. כמו כן, באמידת תחזיות פשיעה יש לקחת בחשבון מספר רב של גורמים המשקפים פעילות של מערכת אכיפת החוק: פעילות משטרתית (מספר תיקים שנפתחו, אחוז גילויים בתיקים אלה, כוח שיטור באזור וכו') ופעילות מערכת המשפט (מספר הרשעות, עונשים שנגזרו וכו'). מלבד המטרה לעמוד על הקשר בין הגורמים השונים ורמת הפשיעה בכל אזור נתון, התחזיות אמורות להיות רלוונטיות לפעילות המשטרתית. לכן יש צורך לאמוד אותן ברזולוציה מרחבית בעלת עניין תפעולי למשטרה ובאופן זמן התואם את צרכי התכנון שלה - עד 3 שנים. בנוסף, יש להפריד בין סוגי הפשיעה השונים. אתגר מתודולוגי נוסף הוא האנדוגניות של נתוני השיטור והפשיעה בתקופת הבסיס - המשטרה מתגברת את כוחותיה כתגובה להתגברות הפשיעה באזורים מסוימים, וכך נוצר מתאם חיובי בין שני הגורמים על פני זמן. אי-התחשבות בתופעה זו עלולה לשבש את ההסקה מהמודלים להסבר תופעת הפשיעה ולזיהוי תרומת השיטור לצמצום הפשיעה.

במסגרת הפרויקט המשותף של הלשכה המרכזית לסטטיסטיקה (הלמ"ס) ומשטרת ישראל הוקם בסיס נתונים מזווג הכולל, בנוסף לנתוני עבריינות הנאספים במשטרה באופן שוטף, מגוון מאפיינים חברתיים-כלכליים ודמוגרפיים ברזולוציות של רשות מקומית ותחנת משטרה. במסגרת עבודה זו, נאמדו מודלים להסבר שיעורי פשיעה מסוגים: אלימות ועבירות רכוש, עבור שנים 2003 – 2009. על בסיס מודלים אלה הופקו תחזיות לשיעורי פשיעה משני הסוגים הנ"ל עבור שנים 2010 – 2012.

1.2 חידושים מרכזיים במחקר הנוכחי

מטרת העבודה הנוכחית היא אמידת תחזיות לשיעורי פשיעה מסוגים אלימות ועבירות רכוש ברזולוציה של רשות מקומית ותחנת משטרה בכיסוי ארצי. תחזיות אלו מיועדות לשימוש משטרת ישראל בתכנון פריסת כוחות שיטור. בספרות, חיזוי שיעורי הפשיעה לצרכי המשטרה נערך עבור יחידות גיאוגרפיות בודדות בלבד, ומבחינה זו העבודה הנוכחית מהווה חידוש מתודולוגי.

בין שאר החידושים במחקר הנוכחי ניתן למנות:

1. זיהוי הגורמים המשפיעים על שיעורי הפשיעה מסוגים אלימות ועבירות רכוש, תוך הבחנה בין המודלים להסבר למודלים לחיזוי הפשיעה;
2. אמידת מודלים לתחזיות מרחביות לטווח קצר, בינוני וארוך, עבור שיעורי פשיעה מסוגים שונים, ברזולוציות גיאוגרפיות שונות;
3. השוואה בין מגוון רחב של מודלים פרמטריים ובחירת המודל לחיזוי הטוב ביותר, במונחי דיוק, יציבות ועמידות לתצפיות חריגות;
4. שימוש במודלים מרחביים ובמשתנים מרחביים לצרכי חיזוי שיעורי הפשיעה, תוך אמידת סטיות התקן לתחזיות נאמדות;

5. שימוש בשיטה בלתי ישירה לאמידת תחזיות של יחידות גיאוגרפיות מצרפיות (כמו שטח שיפוט של תחנת משטרה) ואמידת סטיות תקן לאומדנים המתקבלים;

6. אמידת תחזיות רבעוניות לשיעורי הפשיעה תוך שימוש במתודולוגיה לניתוח סדרות עתיות הכוללת אמידת מודל ARIMA, ניכוי עונתיות והתאמת הסדרות ללוח העברי - אמידת השפעת החגים העבריים וימי הפעילות הנהוגים במשק ישראלי.

המאמר מאורגן כדלקמן: בפרק 2 נסקרת הספרות הקיימת בנושאי תחזיות פשיעה וניתוח נתונים תלויי זמן ומרחב. בפרק 3 נפרשת מתודולוגיה של המחקר הנוכחי: נדונים מודלים לחיזוי שיעורי פשיעה ברמה שנתית תוך השוואה בין שיטות שונות לבניית תחזיות; מוצגת מתודולוגיה לאמידת תחזיות פשיעה לטווח קצר (רבע) באמצעות ניתוח סדרות עתיות תוך התאמתן ללוח העברי הנהוג בישראל. פרק 4 מתאר את בסיס נתוני הפשיעה ששימש לבניית מודלים לחיזוי שיעורי הפשיעה ולהפקת תחזיות. בפרק 5 מוצגים אומדני המודלים ומדדי טיב התחזיות המתקבלות, ופרק 6 דן בעיקרי המסקנות.

2. סקירת שיטות לחיזוי פשיעה

לאורך שנים פותחו מספר גישות לבניית תחזיות פשיעה, כאשר ניתן לחלק את השיטות המוצעות לשתי קבוצות מרכזיות: (1) אמידת תחזיות ממודלים פרמטריים תוך שימוש בכלים סטטיסטיים ואקונומטריים מתקדמים, לרבות שימוש במודלים לתלות בזמן ובמרחב; (2) בניית תחזיות באמצעות שיטות של DATA MINING, כולל אמידת משוואת רגרסיה OLS ממימד גבוה, שיטות neural networks, "עצי רגרסיה" וכד'. לרוב, המודל נבחר לפי שני שיקולים: טיב החיזוי ודרישות הלקוח לגבי רזולוציה מרחבית ותקופת החיזוי.

Olligschlaeger (1997) משווה מספר שיטות לאמידת תחזיות שנתיות וחודשיות לרמת הפשיעה מסוג סחר בסמים בעיר פיטסבורג, ארה"ב. העיר חולקה לתאי האמידה, ובכל אחד מהם נאמד מודל OLS לחיזוי פשיעה, תוך שימוש במשתנים מסבירים עבור אזורים סמוכים לתא, בפיגור של שנה אחת. לאמידת תחזית בתדירות חודשית, נעשה שימוש בשיטה weighted spatial adaptive filtering. לפי שיטה זו, התחזיות נאמדות לפי אלגוריתם איטרטיבי, בו וקטור של מקדמי המשתנים המסבירים מתעדכן בכל איטרציה, ומקדמים סופיים מחושבים כממוצע משוקלל של מקדמי המשתנים המקבילים באזורים הסמוכים לאזור האמידה. במאמר מוצעת שיטה לאמידת משקלות ולבחירת פונקציית התיקון לביצוע איטרציות. התחזיות מופקות לכל אזורי העיר פיטסבורג. טיב התחזיות נבדק תוך שימוש בסימולציות של נתוני הפשיעה מספר רב של פעמים, תוך הכנסת רעש מקרי.

שימוש בשיטת neural networks מיועד בעיקר לזיהוי "נקודות חמות" (hot-spots), אזורים בהם רמת הפשיעה מסוגים שונים גבוהה במיוחד, או לזיהוי אזורים בעלי סיכוי גבוה להתגברות הפשיעה בטווח החיזוי. האלגוריתם שפותח הותאם למערכות GIS (מערכת מידע גיאוגרפית) הנמצאות בידי המשטרה.

המחבר מדווח, כי שימוש בשיטות weighted spatial adaptive filtering ו-neural networks שיפר את טיב התחזיות ביחס למודלים OLS. יש לציין, כי שיטות אלו פותחו לאמידת תחזיות באזורים קטנים, תוך התחשבות בתלות מרחבית ובמתאם סדרתי.

מאמר של Gorr, Olligschlaeger and Thompson (2000), הוא המשך למחקרו של Olligschlaeger (1997). על מנת להעלות את טיב תחזיות חודשיות, קיים צורך להתחשב בעונתיות בפשיעה. מוצעת שיטה לאמידת גורמים עונתיים בסחר בסמים של פיטסבורג, תוך שימוש במודל של פירוק הסדרה העתית למרכיבים: מגמה בסיסית, עונתיות ואי-סדירות באמצעות מודל רגרסיה פרמטרית. יש לציין, כי פירוק

הסדרה נעשה על ידי הכנסת משתנים מסבירים מתאימים למודל רגרסיה. שיטה זו ישימה בנתונים הזמינים בתדירות גבוהה (חודשית או רבעונית). יש לציין, כי לא נעשה שימוש במודלים לפירוק וניתוח סדרות עתיות, כגון מודל ARIMA, שיטות לפירוק עונתי וניכוי עונתיות כמו X-12-ARIMA וכד'.

במאמר של Morenoff and Sampson (1997) מוצג ניתוח נתוני פשיעה בעיר צ'יקגו, ארה"ב, בשנים 1970-1990, כאשר רזולוציה גיאוגראפית הוגדרה ברמת שכונה. במאמר נעשה שימוש ברגרסיה OLS וניתוח רב-משתני למטרות הסבר רמת הפשיעה מסוג אלימות כפונקציה של מגוון משתנים חברתיים-כלכליים ודמוגרפיים, וחושב מקדם לתלות מרחבית על סמך ניתוח שאריות. על מנת לאמוד את השפעות השינויים בערכי משתנים מסבירים, המודל כלל הפרשים בין ערכי המשתנים המסבירים בשתי שנים עוקבות. יצוין, כי הגורם המרחבי לא נותח כחלק אינטגרלי של המודל להסבר רמות הפשיעה.

במאמר של Anselin, Cohen, Cook, Gorr and Tita (2000) נדון בסיס תיאורטי לקיום תלות מרחבית בנתוני פשיעה. מאז שנות ה-90 של מאה ה-20, חלה התפתחות בגישות להסבר תופעת הפשיעה, כאשר למרחב ולזמן ניתן מקום מרכזי במודלים תיאורטיים, קרימינולוגיים ואקונומטריים כאחד, המסבירים החלטת הפרט לבצע פשע. בחלק המתאר מדידת תלות בזמן ובמרחב, נסקרו שיטות מבוססות GIS, ניתוח "נקודות חמות" (hot spots), החלקת Kernel של נתונים תלויי מרחב. מוצע שימוש במגוון מדדים לתלות מרחבית, ביניהם מדד Moran's I (Anselin, 1988). בנוסף, נסקרו שיטות ניתוח מרחבי פרמטרי לנתוני פשיעה. המאמר מסכם, כי לא ניתן להתעלם מהתלות המרחבית בנתוני הפשיעה, אך מודלים מרחביים לנתונים אלה אינם מפותחים די הצורך. יודגש, כי המאמר דן במודלים תיאורטיים היכולים להסביר תופעת הפשיעה כחלק מהחלטות הפרט; הוא לא דן בהבדלים הקיימים בין רזולוציות גיאוגראפיות שונות ובין סוגי פשיעה שונים. סוגיית אמידת תחזיות הפשיעה אף היא לא נדונה.

Corcoran, Wilson and Ware (2003) מציעים אלגוריתם לבניית תחזיות הפשיעה לטווח קצר בתאי שטח קטנים. במקום לנתח פשיעה לפי אזורים המוגדרים אל ידי משטרה (תחנת משטרה, רובע, יישוב), מוצע לקבוע תאי חיזוי, על סמך ניתוח מקרי פשיעה, לפי מקום ביצוע פשע. לאחר חלוקת שטח העיר לשכבות הומוגניות ביחס לנתוני פשיעה וזיהוי "נקודות חמות" - מקומות עם מספר רב של מקרי פשיעה מסוג מסוים, לכל שכבה מותאם מודל להסבר הפשיעה. המשתנים המסבירים כוללים את נתוני הפשיעה מאותו הסוג באזורים סמוכים, כאשר "אזור סמוך" מוגדר כנמצא ברדיוס מסוים מגבולות תא החיזוי. ערך הרדיוס מוגדר מראש על ידי החוקרים. על סמך מודלים אלה, לכל שכבה נאמדת תחזית אירועי הפשיעה מסוג נתון.

בעבודה זו נעשה שימוש בשיטת artificial neural network. חלוקה לשכבות הומוגניות עשויה להעלות את טיב התחזיות, היות ובכל שכבה גיאוגראפית המודל להסבר נתוני הפשיעה עשוי להיות שונה, בהתאם לנתונים. כמו כן, חלוקה לשכבות מאפשרת הקטנת מימד הבעיה בכל אחד מהשכבות. מחברי המאמר מניחים, כי הפשיעה באזור מסוים משפיעה על הפשיעה באזורים הנמצאים ברדיוס שנקבע, והשפעה זו הופכת להיות זניחה עם הגדלת המרחק. יש לציין, כי המחקר נעשה על נתוני סחר בסמים ברחובות העיר דהיינו מדובר במודל ברזולוציה גיאוגראפית גבוהה מאוד.

Liu and Brown (2003) אומדים את צפיפויות וצפיפויות מותנות של נתוני פשיעה לשם הפקת תחזית לטווח קצר מאוד (מספר ימים או שבועות). השיטה המוצעת מבוססת על החיפוש אחרי "הנקודות החמות" של הפשיעה במרחב נתון וניתוח צפיפות מרחבית של הפשיעה מסוג מסוים. שיטה זו נבדקה בנתוני עבירות רכוש בעיר ריצ'מונד, ארה"ב.

סקירה מקיפה של שיטות לחיזוי פשיעה מוצגת במאמר של Gorr and Harries (2003). המחברים מבדילים בין שני סוגי התחזיות: לטווח מיידי (מספר ימים), קצר (מספר שבועות או חודשים), בינוני (שנה-שנתיים) וארוך (מעל שנתיים). מסקנת המאמר היא, ששימוש בשיטות המבוססות על neural networks וחלוקה לשכבות גיאוגרפיות מוצדק יותר לאמידת תחזיות לטווח קצר, בעוד שלאמידת תחזיות לטווח בינוני וארוך יש להעדיף את מודלים של רגרסיה ושיטות לחיזוי סדרות עתיות.

בעבודה של Schubert (2004), נעשה שימוש בשיטות של סטטיסטיקה תיאוריות לזיהוי "נקודות חמות" של הפשיעה בעיר דאלאס, ארה"ב. השיטה המוצעת מיועדת לשימוש במערכות GIS של המשטרה. שיטה זו לא נותנת אפיון הפשיעה כפונקציה של משתנים מסבירים חברתיים-כלכליים.

Oatley, Ewart and Zeleznikow (2006) משווים שיטות שונות לחיזוי הפשיעה, תוך שימוש בנתוני הפשיעה בעיר קליוולנד, אנגליה. המחברים דנים בשיטות עיבוד בסיסי נתונים גדולים והם מחלקים שיטות אלו למספר סוגים. ראשית, שיטות ויזואליות הינן שיטות הנתמכות על ידי מערכות GIS, ומפיקות בעיקר סטטיסטיקה תיאורית. בין השיטות שנדונו - אמידת התפלגות הפשיעה באמצעות החלקת Kernel, איתור "נקודות חמות" של פשיעה בזמן ובמרחב, בניית תרשימים תיאוריים. שנית, שיטות סטטיסטיות הכוללות ניתוח רב-משתני (מרכיבים ראשיים) הנועד לצמצום מימד הבעיה, שיטת "עץ רגרסיה" - חלוקת בסיס הנתונים לשכבות הומוגניות ביחס לסט נתון של משתנים מסבירים ועוד. שלישית, שיטות Data mining - שימוש ב-neural networks (אופטימיזציה של פונקציית מטרה בצורה איטרטיבית לפי כללי ההחלטה מוגדרים מראש). כמו כן, נדונה שיטת חיזוי המבוססת על השילוב של הגישה הבייסיאנית עם האלגוריתם של neural networks. כל השיטות הנדונות במאמר זה אינן פרמטריות, מה שלא מאפשר כימות ההשפעה של גורמים שונים על הפשיעה בתקופה הנוכחית או/על הפשיעה העתידית.

Brown and Smith (2005) אומדים מודל רגרסיה מולטינומית לוגיסטית להסבר וחיזוי של אירועי טרור, תוך ניתוח אירועי הטרור בישראל. במודל זה, התלות המרחבית נלקחה בחשבון על מנת לזהות אזורים המועדפים על ידי המחבלים לביצוע עתידים של פעולות טרור. השיטה מבוססת על הגישה של "מקסום פונקצית תועלת מקרית", במונחי התועלת של המחבלים מביצוע אירוע טרור באזור מסוים כפונקציה של מגוון משתני רקע. האמידה מתבצעת תוך שימוש בשיטת הנראות המרבית (Maximum Likelihood).

יישום שיטות סטטיסטיות לתחזיות הפשיעה בנתוני הערים רוצ'סטר ופיטסבורג, ארה"ב, מוצג בעבודה של Cohen (2006). כל השיטות המוצעות במאמר זה מבוססות על ניתוח סדרות עתיות, מקוריות ומנוכות עונתיות. ה"שיטות הנאיביות" מבוססות על השלכת אחוז שינוי היסטורי לתקופת החיזוי. לפי שיטות "החלקה מעריכית" (exponential smoothing) המגמה הבסיסית של הסדרה נאמדת באמצעות הפעלת ממוצעים נעים עם משקלות הולכים וקטנים בקצב מעריכי עם התרחקות בזמן מהנתון העדכני. לפי שיטה זו, התחזית נבנית תחת הנחה שהמגמה שנאמדה בתקופה האחרונה תימשך גם בעתיד. בהמשך למאמר Cohen, Gorr and Olligschlaeger (2005), נעשה שימוש בשיטת האינדיקטורים המובילים (Leading Indicators) לחיזוי סדרה עתית לצרכי חיזוי הפשיעה. לפי שיטה זו, משתנים מסבירים בפיגור משמשים מעין אינדיקטור לחיזוי פשיעה בעתיד. בשלב ראשון, נאמד מודל רגרסיה OLS הכולל גם משתנים מסבירים בפיגור. בשלב שני, מניחים כי מבנה התלות הנאמד בנתונים היסטוריים נשמר לאורך זמן. לכן, התחזית מתקבלת על ידי הצבת ערכים עכשוויים למשתנים בפיגור תוך שימוש במקדמי הרגרסיה שנאמדו בשלב ראשון. הכנסת משתני מרחק ומשתני הפשיעה באזורים הסמוכים מאפשרת התחשבות בהשפעות מרחביות בנתוני הפשיעה. בנוסף, מוצע יישום השיטה של neural networks בסדרות עתיות לצרכי חיזוי

הפשיעה. תקופת החיזוי במאמר זה הינו חודש אחד. למעט השיטה של אינדיקטורים מובילים, השיטות הנדונות במאמר זה אינן מאפשרות בניית תחזית הפשיעה כפונקציה של מאפיינים חברתיים-כלכליים ודמוגרפיים של האזור. התחזיות נבנו ברזולוציות גיאוגרפיות גבוהות (רובע, תת-רובע), והופקו מדדי טיב לדיוק התחזיות.

במאמר של Mitchell, Brown and Conklin (2007), נדונה השיטה של Spatial discrete choice model המזהה אזורי פשיעה עתידיים, תוך אמידת מודל רגרסיה ליניארית המסביר את רמת הפשיעה הנוכחית כפונקציה של משתני רקע שונים. המחברים מניחים, כי מקרי פשיעה מסוג מסוים יתרחשו באזור בעל אותם המאפיינים כמו האזורים בהם התרחשו מקרי פשע בעבר. לכן, בהינתן נתוני הרקע בתקופה הנוכחית, ניתן לחזות פשיעה בעתיד על ידי ניתוח התפתחות של משתנים חברתיים-כלכליים ודמוגרפיים. המאמר עוסק בהפקת תחזיות לטווח של שבועות; במקרה של אמידת תחזיות לטווח ארוך יותר, יש לבדוק את תוקף ההנחה הזאת. המחברים פיתחו תוכנה מיוחדת להפקת תחזיות פשיעה לטווח קצר.

3. מודלים להסבר וחיזוי פשיעה: מתודולוגיה

לצרכי חיזוי שיעורי הפשיעה בעבודה הנוכחית, הוגדרו 2 רזולוציות גיאוגרפיות: שטח רשות מקומית ושטח תחנת משטרה. רשות מקומית יכולה לכלול ישוב אחד (עירייה או מועצה מקומית) או מספר ישובים קטנים יחסית (מועצה אזורית). שטח תחנת משטרה נקבע על ידי המשטרה לפי הצרכים התפעוליים שלה, ושטח זה יכול לכלול ישוב אחד או מספר ישובים. יצוין, כי כל שטח מדינת ישראל מכוסה בכל אחת משתי החלוקות, כאשר רשות מקומית או תחנת משטרה מהוות תא אמידה. בפתח חקירת נתוני הפשיעה יש לבחון האם קיימת תלות מרחבית בשתי הרזולוציות המרחביות שהוגדרו, בכל אחד מסוגי הפשיעה. כאשר קיימת תלות מרחבית מובהקת בנתונים, הנחת OLS אודות אי-תלות התצפיות במשתנה המוסבר מופרת, ולכן ייתכנו הטיות בפרמטרים הנאמדים ובשונותם (Anselin, et al. (2000)). במקרה זה, נשתמש במודלים מרחביים המתוארים בסעיף 3.3 להלן.

3.1 אמידת תלות מרחבית

המדד הנפוץ לבדיקת קיום התלות המרחבית הינו Moran's I (Anselin (1988)). המדד, עבור המשתנה הנחקר X , ניתן על ידי:

$$I = \frac{N \sum_i \sum_j w_{ij} (X_i - \bar{X})(X_j - \bar{X})}{\sum_i \sum_j w_{ij} \sum_i (X_i - \bar{X})^2} \quad (1)$$

ושונות המדד ניתנת על ידי:

$$\begin{aligned} \text{Var}(I) &= \frac{NS_4 - S_3S_5}{(N-1)(N-2)(N-3)(\sum_i \sum_j w_{ij})^2}, \\ S_1 &= 0.5 \sum_i \sum_j (w_{ij} + w_{ji})^2, \\ S_2 &= \sum_i (\sum_j w_{ij} + \sum_j w_{ji})^2, \\ S_3 &= \frac{N^{-1} \sum_i (X_i - \bar{X})^4}{(N^{-1} \sum_i (X_i - \bar{X})^2)^2}, \\ S_4 &= (N^2 - 3N + 3)S_1 - NS_2 + 3(\sum_i \sum_j w_{ij})^2, \\ S_5 &= S_1 - 2NS_1 + 6(\sum_i \sum_j w_{ij})^2 \end{aligned} \quad (2)$$

בנוסחאות (1) ו-(2), N מסמן סך הכל מספר התצפיות ו- w_{ij} - משקל מרחבי המתאר מבנה התלות המרחבית בין X_i ל- X_j .

לכל ישוב/מועצה i הוגדרה קבוצה S_i - 10 ישובים/מועצות הקרובים ביותר ל- i , ולכל אחד מהם חושב המרחק עד לישוב/מועצה i d_{ij} (לפי מרחק אווירי עד למרכז של היחידה הגיאוגראפית). הנוסחה למשקל מרחבי ניתנת על ידי:

$$w_{ij} = \frac{(1/d_{ij})}{\sum_{j \in S_i} (1/d_{ij})} \quad (3)$$

בנוסחה (3) מניחים כי התלות המרחבית הולכת ופוחתת עם הגדלת המרחק בין הישובים. יש לסייג ולומר, כי תוקף הנחה זו עשוי להיחלש עם הגברת הנגישות בין אזורים שונים בארץ על ידי בניית כבישים ומסילות רכבת וכו'. גודל קבוצה S_i נקבע על סמך ניתוח אמפירי, כאשר הכללת יחידות גיאוגרפיות מרוחקות יותר לקבוצה לא מוסיפה מידע משמעותי על מבנה התלות המרחבית. קיים קושי להגדיר מרחק בין יחידות גיאוגראפיות גדולות כמו שטח תחנות משטרה, המכילות בדרך כלל מספר ישובים/מועצות. כך, ניתן להגדיר "מרכז היחידה הגיאוגראפית" באופנים שונים (מרכז לפי מרחקים מהגבול, הישוב הגדול ביותר, מרכז פעילות כלכלית הגדול ביותר וכו'). לכן, לכל תחנת משטרה i הוגדרה קבוצה T_i הכוללת את כל תחנות המשטרה הגובלות עם i , ולכל זוג תחנות חושב אורך הגבול המשותף r_{ij} . במקרה של תחנות משטרה, המשקל הגיאוגראפי ניתן על ידי:

$$u_{ij} = \frac{r_{ij}}{\sum_{j \in T_i} r_{ij}} \quad (4)$$

ההנחה ב- (4) - התלות המרחבית תהיה חזקה יותר ככל שהגבול המשותף בין שתי תחנות ארוך יותר. על סמך (3) ו-(4), קל לראות, כי $\sum_{j \in S_i} w_{ij} = 1$, $\sum_{j \in T_i} u_{ij} = 1$, מה שמפשט נוסחאות החישוב של מדד Moran's I

ושל שונותו. ניתן להוכיח, כי Moran's I שואף בהתפלגות להתפלגות נורמאלית (Anselin, 1988), עם

תוחלת $E(I) = -\frac{1}{N-1}$, ושונות כפי שמופיע ב- (2). אז, ניתן לחשב p-value לבדיקת מובהקות של מדדי

התלות המרחבית. לוחות 1א ו-1ב מציגים את תוצאות האמידה ברזולוציה מרחבית של ישובים/מועצות ושל תחנות המשטרה, בהתאמה, לשני סוגי הפשיעה הנחקרים בעבודה זו (הודגשו מקרים בהם נמצאה תלות מרחבית מובהקת ברמת מובהקות 5% במשתנה "שיעורי הפשיעה", המוגדר כמספר תיקי עבירות מסוג פשיעה נתון ל-1000 תושבים):

לוח 1א: רזולוציה של ישובים/מועצות				
שנה	עבירות אלימות		עבירות רכוש	
	Moran's I	P-value	Moran's I	P-value
2003	0.0317	0.1438	0.2042	<0.0001
2004	0.0071	0.6414	0.2224	<0.0001
2005	0.0094	0.5759	0.2319	<0.0001
2006	0.0108	0.5378	0.2304	<0.0001
2007	0.0028	0.7709	0.2350	<0.0001
2008	0.0001	0.8546	0.2568	<0.0001
2009	-0.0028	0.4754	0.1830	<0.0001

לוח 1ב: רזולוציה של תחנות משטרה				
שנה	עבירות אלימות		עבירות רכוש	
	Moran's I	P-value	Moran's I	P-value
2003	0.0233	0.6760	0.0592	0.4380
2004	-0.0087	0.9296	0.0879	0.2777
2005	0.1707	0.0562	0.0767	0.3296
2006	0.1200	0.1471	0.0778	0.3237
2007	0.1422	0.1052	0.2988	0.0014
2008	0.1156	0.1768	0.3207	0.0006
2009	1.1148	0.1148	0.5396	<0.0001

ניתן לראות, כי בשיעורי פשיעה מסוג עבירות רכוש עיקריות ברזולוציה של ישובים/מועצות נמצאה תלות מרחבית מובהקת ויציבה על פני כל תקופת הניתוח. ברמת תחנת משטרה, בנתוני שיעורי פשיעה מסוג עבירות רכוש עיקריות, התלות המרחבית נמצאה מובהקת בשנים 2007 עד 2009, אך בשנים מוקדמות יותר בנתוני הפשיעה מסוג עבירות רכוש עיקריות מדד Moran's I אינו מובהק. בנתוני שיעורי פשיעה מסוג אלימות, לא נמצאה תלות מרחבית מובהקת באף אחת מהשנים ובאף אחת משתי הרזולוציות הגיאוגרפיות. מובהקות המדד Moran's I לתלות המרחבית מצביעה על כך, שהשונויות המשותפת בין תצפיות עבור יחידות גיאוגרפיות סמוכות i, j אינה שווה לאפס. לכן, קיים צורך בבניית מודל מרחבי לנתוני עבירות רכוש עיקריות, לפחות ברמת ישוב/מועצה.

3.2. מודל OLS אוטורגרסיבי

במטרה להסביר ולחזות תופעת הפשיעה מסוג k ביחידה גיאוגרפית i בשנה t , נשתמש במודל הבא:

$$y_{i,k,t} = \log(Y_{i,k,t}) = \lambda \log(Y_{i,k,t-r}) + \gamma P_{i,k,t-r} + \delta S_{i,t-r} + \beta_1 X_{i,t-r} + \beta_2 d(X_{i,t-r}) + \varepsilon_{i,k,t} \quad (5)$$

כאשר $Y_{i,k,t}$, מסמן שיעורי הפשיעה בשנה t מסוגים k, r - תקופת הפיגור במשתנים מסבירים, $P_{i,k,t-r}$ - אוסף משתנים המבטאים כוח שיטור ופעילות משטרתית, $S_{i,k,t-r}$ - אוסף משתנים מרחביים (מרחק מתל-אביב ושיעורי פשיעה באזורים שכנים, לפי הגדרות של קבוצות השכנים S_i לישובים/מועצות ו- T_i לתחנות משטרה בסעיף הקודם), $X_{i,t-r}$ - אוסף משתנים חברתיים-כלכליים ודמוגרפיים המאפיינים את אוכלוסיית היחידה הגיאוגרפית i בשנה $t-r$. בנוסף, $d(X_{i,t-r})$ מסמן שינויים שחלו במשתנים $X_{i,t-r}$ משנה $t-r-1$ עד לשנה $t-r$: $d(X_{i,t-r}) = X_{i,t-r} - X_{i,t-r-1}$. מסמן שאריות בלתי תלויות, בעלות תוחלת אפס ושונות קבועה σ^2 .

טרנספורמאציית ה- \log (טבעי) המופעלת על משתנה "שיעורי הפשיעה" מאפשרת לקבל משתנה מוסבר המתפלג נורמאלי בקירוב. המשתנה המרחבי "שיעורי הפשיעה ביחידות גיאוגראפיות שכנות" הוכנס למודל לאחר טרנספורמציית ה- \log משיקולי עקביות.

יצוין כי חלק מאפייני כוח שיטור ופעילות משטרתית הינם ספציפיים לסוג הפשיעה (למשל, אחוז גילויים בתיקי עבירות מסוג רלוונטי או שיעור עבריינים המתגוררים ביחידה גיאוגראפית שהורשעו בעבירות מסוג רלוונטי), ואילו חלק אחר, כגון מספר שוטרים בתחנה והתפלגות השוטרים לפי מקצועות השיטור רלוונטי לכל סוגי הפשיעה. במחקר הנוכחי, נאמדו מודלים בפיגור של עד שלוש שנים: $r = 1, 2, 3$, ולכך שלוש סיבות:

1. בנתוני פשיעה, שימוש במשתנים מסבירים עם פיגור בזמן נותן פתרון לאנדוגניות נצפית בנתונים ומאפשרת לאמוד השפעות של גורמים שונים, ובפרט גורמי שיטור, על שיעורי הפשיעה. שימוש במשתני שיטור בפיגור מאפשר לנטרל את השפעת תגבור כוחות השיטור כתגובה להתגברות הפשיעה על מקדמי המתאם בין משתני שיטור לבין שיעורי פשיעה, מה שמאפשר לאמוד תרומת השיטור להורדת שיעורי הפשיעה.

2. שיטת האינדקאטורים המובילים מצריכה אמידת מודל רגרסיה עם פיגור השווה לתקופת החיזוי. בעקבות מספר תוצאות בספרות המראות עדיפות של שיטת חיזוי זו על פני השיטות הפרמטריות המתחרות לאמידת תחזיות לטווחי זמן בינוני וארוך (למשל Cohen, Gorr and Olligschlaeger (2005)), שיטת האינדקאטורים המובילים תהווה בסיס בבניית המודלים לחיזוי הפשיעה במחקר הנוכחי.

3. רוב המשתנים החברתיים-כלכליים והדמוגרפיים זמינים בלמי"ס בפיגור של שנה ויותר, בגלל מגבלות קליטה וטיוב הנתונים המנהליים.

נסמן ב- Z את מטריצת המשתנים המסבירים ממשוואה (5). יש לאמוד את הווקטור $\beta = [\lambda \ \gamma \ \delta \ \beta_1 \ \beta_2]^T$ של מקדמי המשתנים הכלולים ב- Z וכן את הפרמטר σ^2 . תחת הנחת נורמאליות במודל OLS, התפלגות המשתנה המוסבר בכתוב ווקטורי הינה: $y \sim N(Z\beta, \sigma^2 I)$. אמידת מודל (5) נעשית בשיטת OLS תוך שימוש באלגוריתם איטרטיבי של stepwise selection, בו נבחר סט משתנים מסבירים כך שכל מקדמי המשתנים מובהקים ברמת מובהקות α לפחות. בעבודה הנוכחית, נקבע $\alpha = 0.1$. באמידת תחזיות, קיים יתרון חשוב לשיטת stepwise selection: שימוש אך ורק במשתנים שהשפעתם מובהקת מביא להקטנת סטיות התקן של תחזיות ומעלה יציבותן.

נאמדו שני מודלים רגרסיה OLS עבור הרזולוציה המרחבית של ישובים/מועצות: OLS1 - מודל (5) הנאמד על סמך כל התצפיות עבור כל הישובים/מועצות בארץ, ו-OLS2 - מודל (5) הנאמד למגזר היהודי ולמגזר הערבי בנפרד. מודל OLS2, בניגוד ל-OLS1, אינו מניח זהות הגורמים המשפיעים על שיעורי פשיעה במגזר היהודי ובמגזר הערבי, וכן לא מניח שוויון שוניות σ^2 בשני המגזרים. ברזולוציה מרחבית של תחנות משטרה, לא מתאפשרת אמידת מודל נפרד לכל מגזר מספר מועט של תחנות משטרה שבשטחן אחוז האוכלוסייה הערבית עולה על 50%.

במטרה לבחון קיום הנחות מודל OLS (5), לכל מודל נאמד נערכו בדיקות לאי-תלות ונורמאליות של שאריות והומוסקדסטיות. כמו כן, נבדק קיום מולטיקוליניאריות במודל, תוך ניתוח מטריצת שוניות משותפות של מקדמי הרגרסיה הנאמדים ומדד Conditional Number (אם ערך המדד אינו עולה על 10,

נהוג לומר כי לא נמצאה מולטיקוליניאריות במודל הנאמד). המודלים בהם Conditional Number גדול מ-10 נדחו. בעבודה זו השתמשנו במודלים בהן לא נמצאה עדות סטטיסטית מובהקת (ברמת 5%, בכל המבחנים הסטטיסטיים הרלוונטיים) לסטיות מהנחות מודל רגרסיה OLS, הן ברזולוציה של ישובים/מועצות, הן ברזולוציה של תחנות משטרה.

3.3 מודלים מרחביים

בעבודה הנוכחית נבחנו מודלים מרחביים פרמטריים המאפשרים כימות ההשפעות של גורמים שונים על שיעורי הפגיעה תוך ניצול השונות בין יחידות גיאוגרפיות בשתי רזולוציות שנבחרו: ישובים/מועצות ותחנות משטרה. במקרה בו קיימת תלות מרחבית בין התצפיות של המשתנה המוסבר y , הנחת OLS בנוגע לאי-תלות שאריות אינה מתקיימת. בדרך כלל הסיבה לתלות מרחבית נעוצה בקיום מטרופולינים ושטחי פעילות של עבריינים אשר לא מוגבלים לרמות של ישוב או תחנת משטרה.

בבדיקה לקיום התלות המרחבית באמצעות המדד Moran's I נמצאה תלות מרחבית מובהקת בנתוני עבירות רכוש. לכן, מטריצת השונויות המשותפות של שיעורי פגיעה תלויה במיקום של יחידה גיאוגרפית במרחב (תלות בשונות). בדומה למודל OLS, במחקר הנוכחי אנו מניחים כי לא קיימת תלות מרחבית במשתנים מסבירים הכלולים במטריצה Z .

נסמן מטריצת השונויות המשותפות ב- Ω . תחת הנחת נורמאליות במודל מרחבי, ההתפלגות של y נתונה על ידי: $y \sim N(Z\beta, \Omega)$. מודלים רגרסיה מרחביים שונים מציעים שיטות שונות לבניית Ω כפונקציה של מיקום התצפיות במרחב. מבין מספר מודלים מרחביים, במחקר הנוכחי נבחרו שני מודלים: Spatial Error Model (SER) (Anselin, 1988), ומודל (SGLS) Spatial Generalized Least Square (Anselin, 1988; Dormann et al., 2007).

3.3.1 מודלים SER

מודל SER, בכתוב מטריוצות, ניתן על ידי:

$$\begin{aligned} y &= Z\beta + e \\ e &= u + \varepsilon \end{aligned} \quad (6)$$

כשאר ε ו- u בלתי תלויים, ו- $\varepsilon \sim N(0, \sigma^2 I)$, $u \sim N(0, \rho\Sigma)$, ρ מסמן מקדם מרחבי בשונות, ו- Σ מטריצה המבטאת מבנה התלות המרחבית. ההתפלגות של y במודל (6) תהיה: $y \sim N(Z\beta, \sigma^2 I + \rho\Sigma)$, ומרחב הפרמטרים כולל ווקטור β של מקדמי רגרסיה ושני פרמטרים של שונות: ρ, σ^2 . אמידת מודלים SER נעשית באמצעות השיטה (Restricted Maximum Likelihood) REML.

נשאר להראות, כי במודל SER, שימוש במשתנה מסביר תלוי מרחב "שיעורי פגיעה מסוג עבירות רכוש בפיגור" לא מכניס להטיה במקדמי הרגרסיה ובשונותם. נסמן משתנה זה ב- X_1 , ואוסף משתנים מסבירים ללא תלות מרחבית ב- X_2 . תחת מודל חיבורי לתלות מרחבית במשתנה "שיעורי הפגיעה מסוג עבירות רכוש", ניתן לרשום מודל (6):

$$y = X_1\beta_1 + X_2\beta_2 + \tilde{e} = (M + \bar{e})\beta_1 + X\beta_2 + \tilde{e} = Z\beta + e^* \quad (א6)$$

כאשר β_1, β_2 מסמנים את הפרמטרים של המודל, \bar{e}, \tilde{e} - שאריות התלויות בהימצאות במרחב עבור משתנה מוסבר y ומשתנה מוסבר בפיגור X_1 , בהתאמה. כמו כן: $e^* = \bar{e}\beta_1 + \tilde{e}$, $\beta = [\beta_1, \beta_2]^T$.

$\bar{e} = \bar{u} + \bar{\varepsilon}, \tilde{e} = \tilde{u} + \tilde{\varepsilon}$: לטעות (6), המודל $Z = [M, X]$ לפי $M = EX_1$

זהה נובעת מכך שמדובר באותה תופעה בדיוק, אשר נמדדת בנקודות זמן שונות. נזכיר כי לפי הנחה של מודל SER, שאריות $\bar{\varepsilon}, \tilde{\varepsilon}$ בלתי תלויות ב- \bar{u}, \tilde{u} , בהתאמה. בנוסף מתקיים:

$e^* = u^* + \varepsilon^* = (\tilde{u}\beta_1 + \bar{u}) + (\tilde{\varepsilon}\beta_1 + \bar{\varepsilon})$ נסמן את מטריצת השונות המשותפת של שאריות $\bar{\varepsilon}, \tilde{\varepsilon}$ ב- \mathcal{A} :

$E\bar{\varepsilon}_i\tilde{\varepsilon}_j = \gamma$ ו- $E\bar{\varepsilon}_i\tilde{\varepsilon}_j = 0, j \neq i$. נניח עוד כי השונות המשותפת של \bar{u}, \tilde{u} עבור 2 יחידות גיאוגרפיות

שהוגדרו כשכנות, פרופורציונאלית למידת השכנות שלהן כפי שהוגדרה במטריצה Σ מרחבית:

$E\tilde{u}\bar{u} = \delta\Sigma$. קל לראות כי תחת הנחות אלו התפלגות המשתנה המוסבר בנוסחה (6א) ניתנת על ידי:

$y \sim N(Z\beta, (\sigma_1^2\beta_1^2 + \sigma_2^2 + 2\beta_1\gamma)I + (\rho_1\beta_1^2 + \rho_2 + 2\beta_1\delta)\Sigma)$ ולכן על ידי סימון

$\rho = \rho_1\beta_1^2 + \rho_2 + 2\beta_1\delta$ ו- $\sigma^2 = \sigma_1^2\beta_1^2 + \sigma_2^2 + 2\beta_1\gamma$ אנו מקבלים את מודל (6) גם עבור המקרה בו

קיים משתנה מוסבר תלוי מרחב. טיעונים אלה נכונים גם עבור המודל SGLS אשר נדונה בפרק 3.3.2

להלן. היות ואומדני מודל (6) חסרי הטיה אסימפטוטית, הכנסת משתנה מסביר "שיעורי פשיעה מסוג

עבירות רכוש בפגור" לא תכניס הטיה לאומדני מקדמי הרגרסיה ולשונותם.

בספרות נדונו מספר שיטות למידול התלות המרחבית בין התצפיות (ראה למשל (2009) Beenstock et al.),

(2000) Anselin et al.). במחקר הנוכחי נבנו מספר מודלים למבנה תלות מרחבית בשונות, דהיינו למבנה

של מטריצה Σ .

1. ברזולוציה של ישוב/מועצה נגדיר מטריצה W הכוללת מרחקים בין יחידת גיאוגרפית i לבין 10

היחידות הקרובים ביותר אליה:

$$W_{ij} = \begin{cases} w_{ij}, j \in S_i \\ 0, j \notin S_i \end{cases} \quad (7)$$

בנוסף, באותה הרזולוציה המרחבית, נגדיר מטריצה V המבוססת על פונקציה לא ליניארית של המרחקים בין היחידה הגיאוגרפית למרכז העיר תל-אביב. עיר תל-אביב מהווה מרכז פעילות כלכלית-חברתית במדינת ישראל; נתוני הפשיעה תומכים בכך שהיא מהווה בנוסף מרכז פעילות עבריינית משני סוגי הפשיעה הנבדקים בעבודה זו. לכן, תל-אביב נבחרה להיות נקודת מרכז, וביחס אליה נקבע "פריפריאליות"

של שאר האזורים. נגדיר $v_i = 1 - \sqrt{\frac{l_i}{2(\max_i(l_i))}}$ כאשר l_i הינו מרחק בין יחידה גיאוגרפית i לעיר תל-

אביב, ולכן $\max_i(l_i)$ הינו המרחק בין תל-אביב לבין הישוב המרוחק ביותר ממנו. קל לראות, כי בתל-אביב

$v_i = 1$, כאשר הערך של v_i הולך וקטן עם התרחקות ממנה. פונקציה v_i שייכת למשפחה רחבה של

פונקציות לא ליניאריות של מרחק המוצעות עבור המודלים מרחביים לשונות (Sun, Tu, Yu (2005)).

פונקציה זו הינה מונוטונית יורדת בקצב מתון יחסית, כאשר קצב הירידה הולך וקטן עם התרחקות מעיר

תל-אביב. מטריצה V למבנה התלות המרחבית נתונה על ידי:

$$V_{ij} = \begin{cases} v_i, i = j \\ 0, i \neq j \end{cases} \quad (8)$$

לפי נוסחה (8), מניחים כי שונות שיעורי הפשיעה בעיר תל-אביב ובישובים הקרובים אליה גבוהה יותר מאשר באזורים פריפריאליים. הנחה זו עקבית עם הגישה לפיה השונות תלויה ברמת המשתנה הנחקר; במרכז הארץ, שיעורי הפשיעה לרוב גבוהים יותר, ולכן סביר לצפות לשונות גבוהות יותר. כמובן, ניתן להציע פונקציות נוספות, מה שיכול להוות נושא להמשך המחקר (ראה דוגמאות ב- Anselin et al. (2000), Beenstock et al. (2009)). בנתוני שיעורי הפשיעה שנבדקו במסגרת המחקר הנוכחי נמצא, כי מטריצה V המוגדרת ב- (7) תואמת מבנה הנתונים טוב יותר מאשר מספר פונקציות מתחרות (כמו למשל לקיחת מרחק מתל-אביב בצורה ישירה, פונקציה הופכית של מרחק מת-אביב וכו'). ברזולוציה מרחבית של ישוב/מועצה, נגדיר מודל SER1 בו $\Sigma = W$, ומודל SER2 בו $\Sigma = V$, לפי הגדרות של מטריצות אלו בנוסחאות (7) ו-(8).

2. ברזולוציה מרחבית של תחנות משטרה, נגדיר מטריצת שונות U :

$$U_{ij} = \begin{cases} u_{ij}, & j \in T_i \\ 0, & j \notin T_i \end{cases} \quad (9)$$

מודל SER ברזולוציה מרחבית של תחנת משטרה מתקבל על ידי הצבה: $\Sigma = U$. חשוב לציין כי אין בבדיקה לקיום התלות המרחבית באמצעות המדד Moran's I כדי להצביע על קיום התלות של תוחלת המשתנה הנחקר בהימצאות במרחב (תלות בתוחלת). לכן, בנוסף למודלים למבנה של מטריצת השונות, נבנו משתנים מסבירים מרחביים: מרחק מתל-אביב ושיעורי פשיעה ביחידות גיאוגרפיות שכנות. לצרכי העבודה הנוכחית, המשתנים הנ"ל מוגדרים כדלקמן:

1. TA_i - מרחק מיחידה גיאוגרפית i עד למרכז העיר תל-אביב - יפו (מרחק אווירי, בקילומטרים);

2. $Y_{i,k,t}^{SH}$ - שיעורי פשיעה מסוג k בתקופה t ביחידות הגיאוגרפיות שהוגדרו כיחידות שכנות של יחידה i :

(א) ברזולוציה של ישובים/מועצות, זהו ממוצע משוקלל של שיעורי פשיעה ב- 10 יחידות גיאוגרפיות הקרובות ביותר. לכל יחידה גיאוגרפית i , סוג פשיעה k ונקודת זמן t ,

$$Y_{i,k,t}^{SH} = \sum_{j \in S_i} w_j Y_{j,k,t}, \text{ כאשר } w_j = \frac{pop_{j,t}}{\sum_{j \in S_i} pop_{j,t}}, \text{ ו- } pop_{j,t} \text{ מסמן את גודל אוכלוסייה}$$

בישוב/מועצה j בשנה t .

(ב) ברזולוציה גיאוגרפית של תחנות המשטרה, משתנה "שיעור פשיעה אצל שכנים" מוגדר כממוצע משוקלל של שיעורי הפשיעה מאותו הסוג בתחנות הגובלות עם תחנה i :

$$Y_{i,k,t}^{SH} = \sum_{j \in T_i} w_j Y_{j,k,t}, \text{ במקרה זה מסמן את גודל אוכלוסייה } w_j = \frac{pop_{j,t}}{\sum_{j \in T_i} pop_{j,t}}, \text{ ו- } pop_{j,t}$$

בשטח השיפוט של תחנת משטרה j בשנה t .

באמצעות המשתנים האלה ניתן לאמוד את השפעת המיקום במרחב על רמת שיעורי הפשיעה משני הסוגים שנחקרו.

שימוש במודלים מרחביים מסוג (6) לא מאפשר הפעלת האלגוריתם האיטרטיבי stepwise selection לבחירת סט המשתנים המסבירים שהשפעתם מובהקת, וזהו אחד החסרונות הבולטים של שיטות מרחביות בפתרון בעיית אמידת תחזיות וסטיות התקן שלהן. לכן, מלכתחילה נבחר סט משתנים מסבירים בעלי יכולת החיזוי הגבוהה ביותר, במונחי דיוק תחזיות מתקבלות, סטיות תקן של תחזיות ומובהקות המקדמים לאורך תקופת המחקר. סט זה לא שונה משנה לשנה על מנת ליצור עקביות בתוצאות המתקבלות.

3.3.2 מודל SGLS

נגדיר מודל ריבועים פחותים מוכללים (GLS):

$$\begin{aligned} y &= Z\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 \Sigma) \end{aligned} \quad (10)$$

כאשר Σ מסמן מטריצה מרחבית של שונויות משותפות. התפלגות המשתנה מוסבר y ניתנת על ידי: $y \sim N(Z\beta, \sigma^2 \Sigma)$. שימוש במטריצה מרחבית הופכת מודל (10) למודל מרחבי של ריבועים פחותים מוכללים; ברזולוציה מרחבית של ישובים/מועצות, הצבת מטריצה מרחבית במקום Σ תביא למודל Spatial GLS (SGLS). SGLS מניח השפעה כפלית (במקום חיבורית ב-SER) של המרכיב המרחבי בשונויות: במודל SGLS מניחים $\Omega = \sigma^2 \Sigma$, לעומת $\Omega = \sigma^2 I + \rho \Sigma$ במודל SER.

בשלבם המקדימים של המחקר, נבדקה אפשרות שימוש בשיטת SGLS עם מטריצות השונויות המשותפות המבוססות על מטריצת W (נוסחה 7) ברזולוציה של ישובים/מועצות ועל מטריצת U (נוסחה 9) ברזולוציה של תחנות משטרה. בשל חוסר יציבות האומדים המתקבלים בשיטות הני"ל, הוחלט שלא להשתמש בהן במסגרת המחקר הנוכחי. לכן, השימוש במודל SGLS נעשה ברזולוציה של ישובים/מועצות בלבד, כאשר $\Sigma = V$. בדומה למודל SER, במודל SGLS נכלל משתנה מסביר מרחבי (מרחק מתל-אביב), על מנת לבדוק קיום ההשפעה המרחבית על התוחלת של המשתנה המוסבר.

בניגוד למודל OLS, מודלים מרחביים לא נאמדו בנפרד לפי מגזר, היות ומודל מרחבי נפרד לכל מגזר פירושו - כל אחד משני המגזרים מהווה מרחב נפרד ובלתי תלוי במשנהו, והנחה זו נראית כבלתי סבירה ונוגדת את נתוני המשטרה בדבר ההשפעות הבין-מגזריות, במיוחד בעבירות רכוש. לצרכי בחירת מודל מרחבי, נעשה שימוש במדד AIC לטיב התאמת המודל, ואנו נעדיף מודלים עם מדד AIC הקטן ביותר. בעבודה זו, נעשה שימוש במודלים בהם לא נמצאה עדות סטטיסטית לסטייה מהנחות המודלים המרחביים (6) ו-(10), הן ברזולוציה של ישובים/מועצות, הן ברזולוציה של תחנות משטרה.

3.4 מודלים להסבר שיעורי פשיעה - ניתוח גמישויות

במטרה לבחון את עוצמת ההשפעה של משתנה מסביר X על משתנה מוסבר Y , חושבו גמישויות לכל משתנה מסביר במודלים להסבר שיעורי פשיעה. הנוסחה למקדם הגמישויות של משתנה מסביר X ניתנת על ידי:

$$E_x = \frac{\partial Y}{\partial X} \times \frac{\bar{X}}{\bar{Y}} \quad (11)$$

כאשר \bar{X} מסמן ממוצע ערכי משתנה X , ו- \bar{Y} - ממוצע ערכי משתנה Y . משמעות הערך E_x : בכמה אחוזים ישתנה ערך של משתנה מוסבר Y (אצלנו, שיעורי הפשיעה מסוגים אלימות ועבירות רכוש עיקריות)

כתוצאה משינוי בערך של משתנה מסביר X באחוז אחד. חישוב גמישויות בנתוני פשיעה נעשה בעבודה של רגב (2008), במודל להסבר פשיעה מסוג "עבירות צווארון לבן". קיימת חשיבות מיוחדת לאמידת גמישות לתכנון פעילות משטרתית, שכן באמצעותם ניתן לאמוד השפעות שוליות של שיטור, פעילויות אכיפה, הרתעה ועוד. כתוצאה, ניתן לתכנן הקצאת משאבים משטרתיים בצורה יעילה יותר להשגת מטרה של הורדת שיעורי הפשיעה.

נוכיר, כי המשתנה המוסבר במחקר הנוכחי הינו לוג שיעורי פשיעה. ניתן להראות, באמצעות גזירה לפי הנוסחה (11), כי במקרה זה:

(1) עבור משתנה מסביר לאחר לוג-טרנספורמציה: $E_X = \hat{\beta}_{\log(X)}$, כאשר $\hat{\beta}_{\log(X)}$ - מקדם רגרסיה של משתנה $\log(X)$.

(2) עבור משתנה מסביר X (ללא לוג-טרנספורמציה): $E_X = \hat{\beta}_X \times \bar{X}$, כאשר $\hat{\beta}_X$ מסמן מקדם רגרסיה של משתנה X .

סימן של E_X מצביע על כיוון ההשפעה; יצוין, כי סימן זה יכול להיות שונה מסימן של $\hat{\beta}_X$ במקרה בו ממוצע ערכי X - שלילי.

3.5 אמידת תחזיות שיעורי הפשיעה וסטיות תקן לתחזיות

נוכיר, כי ניתן ליישם שיטת האינדקאטורים המובילים באמצעות מודל פרמטרי המספק אומד לווקטור מקדמי המשתנים המסבירים $\hat{\beta}$ (כמו מודלים (5), (6) ו-(10)), ובזכות הכללת משתנים בפיגור למודל. לכן, לכל המודלים לעיל: OLS, SER ו-SGLS, ניתן לבנות תחזיות ולאמוד סטיות התקן שלהן בצורה פשוטה יחסית.

נסמן את סט המשתנים בפיגור ב- $Z^{(1)}$, וסט המשתנים ללא פיגור ב- $Z^{(2)}$, $Z = [Z^{(1)}, Z^{(2)}]$. ווקטורים של מקדמי הרגרסיה יהיו $\hat{\beta}^{(1)}$ ו- $\hat{\beta}^{(2)}$, בהתאמה. אזי, כתוצאה מאמידת מודל פרמטרי (5), (6) או (10), עבור נקודת זמן t ועבור סוג פשיעה k , אנו מקבלים (בכתיב מטריצות):

$$\hat{y}_{k,t} = Z\hat{\beta} = Z_k^{(1)}\hat{\beta}^{(1)} + Z_k^{(2)}\hat{\beta}^{(2)} \quad (12)$$

כאשר t מסמן תקופת הפיגור, ו- $\hat{y}_{k,t+r}$ - לוג שיעורי פשיעה מסוג k בנקודת זמן $t+r$. כמו כן מתקבל אומד למטריצת השונויות של $\hat{\beta}$ - $Var(\hat{\beta})$, כאשר $\hat{\beta} = [\hat{\beta}^{(1)}, \hat{\beta}^{(2)}]^T$. התחזית מתקבלת על ידי:

$$\hat{y}_{k,t+r} = Z_k^{(1)}\hat{\beta}^{(1)} + Z_k^{(2)}\hat{\beta}^{(2)} \quad (13)$$

מטריצת השונויות של התחזיות ניתנת על ידי:

$$Var(\hat{y}_{k,t+r}) = Z_k^{(1)}Var(\hat{\beta}^{(1)})(Z_k^{(1)})^T + Z_k^{(2)}Var(\hat{\beta}^{(2)})(Z_k^{(2)})^T \quad (14)$$

נוכיר, כי $\hat{y}_{k,t}$ מפולג בקירוב נורמאלית, לכן המשתנה "שיעורי פשיעה מסוג k " מפולג בקירוב לוג-נורמאלית. מ- (13) ו-(14), מהנחת ההתפלגות נורמאלית של המשתנה המוסבר במודלים (5), (6) ו-(10) ומנורמאליות (אסימפטוטית) של אומדני $\hat{\beta}$ במודלים הנ"ל נובע, כי בקירוב מתקיים:

$$\hat{y}_{k,t+r} = \log(\hat{Y}_{k,t+r}) \sim N(\hat{y}_{k,t+r}, Var(\hat{y}_{k,t+r})) \quad (15)$$

לפי תכונות ההתפלגות הלוג-נורמאלית, מטריצת השונויות של תחזיות לשיעורי פשיעה מסוג k לשנת $t+r$ ניתנת על ידי:

$$\text{Var}(\hat{Y}_{k,t+r}) = (\exp(\text{Var}(\hat{y}_{k,t+r}) - 1) \times \exp(2\hat{y}_{k,t+r} + \text{Var}(\hat{y}_{k,t+r}))) \quad (16)$$

על מנת לבחון את טיב הקירוב (16) לשונויות התחזיות של שיעורי פשיעה $\hat{Y}_{i,k,t+r} = \exp(\hat{y}_{i,k,t+r})$, בוצעה אמידת השונויות באמצעות שיטת ה-bootstrap, ונמצא כי ההטיה של אומדני bootstrap לשונויות מקירוב (16) הינה זניחה בשני סוגי הפשיעה הנחקרים, לכל השנים שבניתוח. על סמך תוצאה זו הסקנו כי ניתן שימוש בקירוב הלוג-נורמאלי לשונויות התחזיות אינו גורם להטיה באומדנים המתקבלים.

3.6 שיטה בלתי ישירה - חישוב תחזיות ואמידת סטיות תקן

לפי שיטות לאמידת תחזיות שנדונו עד כה, התחזית מתקבלת על ידי התאמת מודל פרמטרי ברזולוציה מרחבית נדרשת ואמידת תחזיות לשיעורי פשיעה ישירות על סמך המודל באותה הרזולוציה. נקרא לשיטות אלו "שיטות ישירות", כאשר שיטות אלו מתאימות הן לרזולוציה של ישובים/מועצות, הן לרזולוציה של תחנות משטרה. עבור יחידות מרחביות מצרפיות (תחנת משטרה, מחוז, רמה כלל-ארצית), ניתן לקבל תחזית המבוססת על תחזיות של היחידות קטנות יותר הנכללות בתוכן (במקרה שלנו, על סמך תחזיות לישובים/מועצות). נקרא לשיטה זו "שיטה בלתי ישירה" לאמידת תחזיות ליחידה מצרפית.

נגדיר St_m - אוסף יחידות גיאוגראפיות הנמצאות בשטח השיפוט של תחנת משטרה m . היות ושיעורי הפשיעה מחושבות ל-1000 תושבים, אומד לשיעור פשיעה מסוג k בשטח התחנה m בשנה t מתקבל כמוצע משוקלל של שיעורי פשיעה ביחידות גיאוגראפיות השייכות ל- St_m :

$$\tilde{Y}_{m,k,t} = \sum_{i \in St_m} \hat{Y}_{i,k,t} w_{i,t} \quad (17)$$

כאשר $w_{i,t} = \frac{pop_{i,t}}{\sum_{i \in St_m} pop_{i,t}}$, ו- $pop_{i,t}$ מסמן גודל אוכלוסיה של יחידה גיאוגראפית i בשנה t .

לפי נוסחה (13), לכל יחידה גיאוגראפית i התחזית מתקבלת על ידי $\hat{y}_{i,k,t+r} = Z_{i,k,t}^{(1)} \hat{\beta}^{(1)} + Z_{i,k,t}^{(2)} \hat{\beta}^{(2)}$. בכל המודלים במחקר הנוכחי (SER, OLS ו-SGLS) מקדמי הרגרסיה בלתי תלויים אסימפטוטית, וכן במודלים אלה מניחים כי לא קיימת תלות מרחבית במשתנים מסבירים. בנוסף נציין, כי שימוש במודלים מרחביים נועד לנכות את התלות המרחבית מהתצפיות בתהליך האמידה, שכן התלות המרחבית מתבטאת דרך שאריות אשר לא נלקחות בחשבון בתהליך החיזוי. לכן, נסיק כי תחת הנחות אלו תחזיות שיעורי הפשיעה $\hat{y}_{i,k,t+r}$ הינם בלתי תלויים בקירוב, וניתן לקרב את שונויות התחזית לשיעורי פשיעה ברזולוציה של תחנת משטרה m בשיטה הבלתי ישירה על ידי:

$$\text{Var}(\tilde{Y}_{m,k,t}) = \text{Var}\left(\sum_{i \in St_m} w_{i,t} \hat{Y}_{i,k,t}\right) = \sum_{i \in St_m} w_{i,t}^2 \text{Var}(\hat{Y}_{i,k,t}) \quad (18)$$

יש לסייג, כי ייתכן ושימוש בקירוב (18) עבור משתנה מסביר "שיעורי פשיעה מסוג עבירות רכוש בפיגור" עלול לגרום להטיה מסוימת כלפי מטה בשונויות התחזיות במודל SER1 (שכן תחת מודל זה השונויות המשותפת בין התצפיות בישובים/מועצות הנמצאים בשטח תחנת משטרה שונה מאפס; לפי ממצאי הבדיקה לקיום התלות המרחבית היא חיובית). נשים לב כי הטיה זו תוכנס לאומדני שונויות התחזיות בשיטה הבלתי ישירה על סמך מודלים OLS, SER2 ו-SGLS המניחים כי מטריצת השונויות של המשתנה

המוסבר הינה אלכסונית ($\Sigma = I$ ב- OLS, $\Sigma = V$ ב- SER2 ו- SGLS), ומתעלמים מקיום השונויות המשותפות בין הישובים/מועצות. היות ומטרת העבודה הנוכחית - השוואה בין שיטות חיזוי שונות, קיום הטיה בגודל זהה באותו כיוון לא משפיע על תוצאות ההשוואה בבחירת המודל הטוב ביותר לחיזוי. הנחת אי תלות מרחבית בכל המשתנים המסבירים מפשטת תהליכי אמידה וחיזוי במידה משמעותית. על מנת לבדוק טיב הקירוב, במחקר אמפירי נשווה את סטיות התקן המתקבלות ממודלים שונים. במידה ושונות התחזיות בשיטה הבלתי ישירה לפי מודל SER1 דומה לזו המתקבלת משאר השיטות הבלתי ישירות, בהינתן סטיות התקן דומות של תחזיות ישירות ברזולוציה של ישובים/מועצות, נסיק כי גודל ההטיה הינו זניח ונוסחה (18) מהווה קירוב מספיק טוב לשונות התחזיות בכל המודלים.

עבור תקופות חיזוי עתידיות הערכים של $pop_{i,t}$ אינם ידועים, ולכן יש לאמוד אותם על מנת לקבל ערכי $w_{i,t}$ בנוסחאות (17) ו- (18). תחזית האוכלוסייה לנקודה $t+1$ ביחידה i מתקבלת על ידי אקסטרפולציה של סדרה שנתית של גודלי האוכלוסייה באותה היחידה הגיאוגרפית, תוך השלכה של שיעור גידול ממוצע של אוכלוסייה ביחידה זו על פני שלוש השנים האחרונות לשנה הבאה:

$$pop_{i,t+1} = b_{i,t} \times pop_{i,t}$$

$$b_{i,t} = \left(\frac{pop_{i,t-2}}{pop_{i,t-3}} + \frac{pop_{i,t-1}}{pop_{i,t-2}} + \frac{pop_{i,t}}{pop_{i,t-1}} \right) / 3 \quad (19)$$

בבדיקות שנערכו, שיטה זו לחיזוי אוכלוסייה ברמת ישוב נמצאה מספיק מדויקת: סטייה ממוצעת מהנתון בפועל שווה ל- 0.5% לתחזיות של שנה אחת, 1% לתחזית לשנתיים ו- 2% לתחזית לשלוש שנים). דיוק זה הושג בזכות העובדה ששיעורי גידול האוכלוסייה, לאחר מיצוע על פני 3 שנים אחרונות, מספיק יציבים גם ברמת ישוב בודד/תחנה בודדת. התחזית לשנת $t+2$ מתקבלת על ידי הכפלה נוספת במקדם גידול האוכלוסייה $b_{i,t}$ בנוסחה (18): $pop_{i,t+2} = b_{i,t} \times pop_{i,t+1} = b_{i,t}^2 \times pop_{i,t}$. באופן דומה, עבור שנה $t+3$, מתקבל:

$$pop_{i,t+3} = b_{i,t} \times pop_{i,t+2} = b_{i,t}^3 \times pop_{i,t}$$

קיים מספר יתרונות לשיטה בלתי ישירה על פני השיטה הישירה. ראשית, מספר התצפיות הזמינות בכל שנה ברזולוציה של ישוב/מועצה גדול בהרבה ממספר התצפיות הזמינות ברמת תחנת משטרה (237 לעומת 60 בהתאמה, בנתוני פשיעה מסוג אלימות, שנת 2008). לכן, ברזולוציה של ישוב/מועצה קיימת אפשרות לפקח על מספר רב יותר של גורמים המשפיעים על שיעורי הפשיעה. שנית, אמידת תחזיות בשיטה הישירה עלולה לגרום לאי-דיוקים היות אותם הגורמים בישובים שונים יכולים להשפיע בכיוונים מנוגדים, והשפעות אלו מתקזזות בנתון מצרפי. תופעה זו קיימת בתחנות הכוללות ישובים הטרוגניים מבחינת הרכב האוכלוסייה, תכונות דמוגרפיות, מיקום במרחב וכיוצא בזה. לכן, השיטה הבלתי ישירה עשויה להיות עמידה יותר לערכים חריגים, שכן מודל המבוסס על 60 תצפיות (או אף פחות מכך) יהיה רגיש יותר לקיום החריג מאשר מודל ברזולוציה של ישוב/מועצה המבוסס על כ- 230 תצפיות בכל אחת משנות הניתוח. כמו כן, פירוק נתוני התחנות לרמה של ישובים מאפשרת זיהוי חריגים ברמת ישוב ובניית התחזית תוך פיקוח על השפעה שהביאה לחריגות זו. כתוצאה, סטיות התקן של תחזיות לפי השיטה הבלתי ישירה עשויות לקטון ביחס לאלו המתקבלות לפי השיטה הישירה. יתרון זה של השיטה הבלתי ישירה עשוי לבוא לידי ביטוי בתחנות הכוללות גם ישובים יהודיים וגם ישובים ערביים.

בנוסף, תחזית ברמה ארצית או ברמת מחוזות ניתנת לבנייה אך ורק לפי השיטה הבלתי ישירה, שכן לא ניתן להתאים מודל רגרסיה פרמטרי לתצפית אחת או ל- 8 תצפיות, בהתאמה.

3.7 מדדי טיב החיזוי

בעקבות Cohen (2006), Gorr and, Olligschlaeger (2005) Cohen, ואחרים, אנו משתמשים במדד MAPE לטיב תחזיות הפשיעה:

$$MAPE = \sum_i abs(1 - \frac{\hat{Y}_{i,k,t}^F}{Y_{i,k,t}^O}) \times 100 \quad (20)$$

כאשר $\hat{Y}_{i,k,t}^F$ מסמן תחזית לשנה t של שיעורי פשיעה מסוג k ביחידה גיאוגראפית i , ו- $Y_{i,k,t}^O$ - שיעורי פשיעה מסוג k שנצפו בפועל באותה שנה ובאותה יחידה גיאוגראפית. משמעותו של מדד MAPE - בכמה אחוזים, במוצע, תסטה התחזית מהנתון בפועל עבור שנת התחזית. הערך המקובל בספרות המקצועית של מדד MAPE לתחזיות שבועיות וחודשיות הינו 20% במוצע (Cohen (2006)). בשל הצורך להשוות תחזית לערך שהתקבל בפועל, המדד חושב עבור השנים 2006 - 2009, והמודל הטוב ביותר בשנים אלו שימש להפקת תחזיות לשנים 2010 - 2012. בנוסף, אנו משתמשים בסטיות תקן לתחזיות כמדד ליציבותן ואמידותן לחריגים, תוך השוואת סטיות התקן המתקבלות בשנים שונות בתקופת המחקר.

3.8 אמידת תחזית רבעונית של שיעורי הפשיעה - ניתוח סדרה עתית

לצרכי היערכות של כוחות משטרה, נדרשות תחזיות בתדירות גבוהה יותר מאשר שנתית. בעבודה הנוכחית, בשל היעדר נתונים חברתיים-כלכליים ודמוגרפיים רבעוניים לאמידת מודל רגרסיה, מוצעות שיטות לאמידת תחזיות פשיעה תוך שימוש במתודולוגיה של ניתוח סדרות עתיות. בחירת הרזולוציה המרחבית של תחנת משטרה לצרכי ניתוח סדרות עתיות נבעה משיקולי יציבות הסדרות, שכן הסדרות ברזולוציה של ישוב/מועצה חשופות לתנודות חריגות בעוצמה רבה יותר, בעיקר בישובים קטנים ובינוניים. תנודתיות זו עלולה לפגוע בדיוק ויציבות התחזיות.

3.8.1 תהליך ניכוי עונתיות ואמידת השפעת חגים וימי פעילות

לצרכי אמידת השפעה עונתית, השפעות חגים וימי פעילות במשק וסדרה מנוכת עונתיות נשתמש בשיטה X-12-ARIMA, אשר פותחה בלשכה המפקדית של ארה"ב (US Census Bureau). לפי האלגוריתם האיטרטיבי של השיטה הזו, לתחילה מתבצעת אמידת גורמי חג וימי פעילות והתאמת הסדרה המקורית לגורמים אלה. לאחר מכן, לסדרה המותאמת נבחר ונאמד מודל ARIMA להרחבת הסדרה. לבסוף, גורמים עונתיים, סדרה מנוכת עונתיות ומגמת הסדרה נאמדים על ידי הפעלת ממוצעים נעים מאורך מתאים. שיטות לאמידת השפעת החגים העבריים וימי הפעילות הנהוגים בישראל מתוארות ב- Burk and Gubman (2003). אמידת מודל ARIMA לסדרה עתית, מודלים לפירוק עונתי וניכוי עונתיות נסקרו בהרחבה ב- Box and Jenkins (1976), Hannan and Quinn (1979), Findley et al (2000). בפרק זה, נתייחס לשלבי X-12-ARIMA הרלוונטיים לחיזוי מספר אירועי פשיעה בתדירות רבעונית.

ניתן לפרק את סדרה עתית y_1, y_2, \dots, y_T למרכיבים עיקריים השונים זה מזה מבחינה מושגית: S - מרכיב עונתיות, C - מרכיב מגמה-מחזור (השם מגמה-מחזור נובע מכך שמרכיב C כולל בתוכו השפעות המחזוריים הכלכליים על הסדרה) ו- I - מרכיב האי-סדירות (שאריות הלא מוסברות). המודל לסדרה y יהיה:

$$y = f(S, C, I) \quad (21)$$

כאשר f מסמן פונקציה לפירוק סדרה עתית. המודל הנפוץ לפירוק סדרה עתית, בו אנו משתמשים במחקר הנוכחי, הינו המודל הלוג-חיבורי, כאשר המשוואה (21), לכל נקודת זמן t , ניתנת על ידי:

$$y_t = \log(S_t \times C_t \times I_t) = \log(S_t) + \log(C_t) + \log(I_t) \quad (22)$$

לוח שנה משפיע על התנהגות של סדרה עתית; לכן, קיים צורך לאמוד את השפעת חגים וימי פעילות ולנכותם מהסדרה על מנת למנוע עיוותים באמידת גורמים עונתיים. היות ובישראל החגים יכולים לנוע בין רבע לרבע, השפעתם על הסדרה לא תנוכה בתהליך ניכוי עונתיות, שכן מרכיב עונתי S משקף תנודות רבעוניות החוזרות על עצמן מדי שנה בצורה פחות או יותר סדירה, מבחינת עיתוין ועוצמתן. היות והשפעות חגים וימי פעילות אינן שייכות כאמור למרכיבים S או C , הן מהוות חלק ממרכיב האי-סדירות I . לכן, ניתן לרשום:

$$I_t = TH_t \times E_t \quad (23)$$

כאשר TH מסמן מרכיב השפעת החגים וימי הפעילות, ו- E אי-סדירות נשארת. אז, המודל הלוג-חיבורי המורחב לפירוק סדרה עתית יהיה:

$$y_t = \log(S_t \times C_t \times TH_t \times E_t) = \log(S_t) + \log(C_t) + \log(TH_t) + \log(E_t) \quad (24)$$

לצורך אמידת המרכיב P , פותח בלמ"ס אוסף מתאים של משתנים מסבירים, הכולל משתני ימי הפעילות: הפרשים בין מספר ימי ראשון, שני, ..., שישי לבין מספר ימי שבת בכל רבע, תאריכי החגים ומספר ימי חול המועד באותו הרבע. אומדן ראשוני למרכיב I משמש כמשתנה מוסבר. בתהליך האמידה, בוצע זיהוי חריגים קיצוניים, כאשר השפעתם נוכתה באמצעות הכנסת משתני דמי מתאימים. ממקדמי מודל רגרסיה OLS, לאחר טרנספורמציה מתאימה, התקבלו גורמי חג וימי פעילות לכל רבע (מרכיב TH_t). היות ולוח השנה ידוע מראש, ניתן לקבל תחזיות לגורמים החגים וימי הפעילות לכל תקופת החיזוי.

לסדרה מנוכת השפות חגים וימי פעילות: $y_t^* = y_t - \log(TH_t)$, מותאם מודל ARIMA. מודל ARIMA עונתי מסדר $(P, D, Q)_s$ (p, d, q) ניתן על ידי:

$$\phi_p(B)\Phi_p(B^s)(1-B)^d(1-B)^D y_t^* = \theta_q(B)\Theta_q(B^s)e_t \quad (25)$$

כאשר ϕ, Φ מסמנים פולינומים אוטורגרסיביים (AR) מסדר p ו- P , בהתאמה, θ, Θ פולינומים ממוצעים נעים (MA) מסדר q ו- Q , בהתאמה. B מסמן אופרטור הזזה אחורנית: $By_t = y_{t-1}$, ו- s אורך התקופה העונתית. עבור סדרות רבעוניות $s = 4$.

מתוך כל המודלים האפשריים מסדר $p, q, P, Q \leq 2$ ו- $d, D \leq 1$ נבחר המודל הטוב ביותר עבור הסדרה הספציפית, במונחי מספר בדיקות ומבחנים סטטיסטיים, וביניהם מבחן חי-בריבוע לטיב ההתאמה, טעות תחזיות המודל לשנה, שנתיים ושלוש שנים קדימה, באחוזים. הגבלת סדר המודל דרוש על מנת למנוע חוסר יציבות העלולה להופיע במודלים מסדר גבוה. בעבודה זו, אמידת מודל ARIMA נעשית תוך זיהוי וניטרול ערכים חריגים בסדרה (ראה Findley et al (2000)).

3.8.2 בניית תחזיות על סמך מרכיבי הסדרה החזויים

כתוצאה מביצוע האלגוריתם X-12-ARIMA לניכוי עונתיות, מתקבלים אומדנים לכל אחד ממרכיבי סדרה עתית. בהתבסס על אומדנים אלה, אנו מציעים 2 שיטות לחיזוי סדרה בתדירות רבעונית. לפי שיטה

1 (TS1), תחזית לתקופה $t+r$, לאחר העברת מרכיבי הסדרה לסקאלה מקורית על ידי טרנספורמציה exp, ניתנת על ידי:

$$\hat{y}_{t+r}^{(1)} = \hat{S}_{t+r} \times \hat{TH}_{t+r} \times (\hat{CE})_{t+r} \quad (26)$$

כאשר \hat{CE} מסמן סדרה מנוכת עונתיות הכוללת ממגה-מחזור C ואי-סדירות נשאר E . תחזית לסדרה מנוכת עונתיות \hat{CE}_{t+r} מתקבל על ידי התאמה נוספת של מודל ARIMA לסדרה מנוכת העונתיות. תחזית למרכיב העונתי S מתקבל בתהליך ניכוי עונתיות.

לפי שיטה 2 (TS2), נשתמש בתחזיות לסדרה מקורית מנוכת השפעות החגים וימי הפעילות y_t^* . נסמן

תחזיות אלו ב- \hat{SCE}_{t+r} , היות והן כוללת מרכיבים: מגמה-מחזור C , עונתיות S ואי-סדירות נשאר E . לכן, תחזית לנקודה $t+r$, לאחר העברת מרכיבי הסדרה לסקאלה מקורית על ידי טרנספורמציה exp, ניתנת על ידי:

$$\hat{y}_{t+r}^{(2)} = \hat{TH}_{t+r} \times y_{t+r}^* = \hat{TH}_{t+r} \times (\hat{SCE})_{t+r} \quad (27)$$

תחזיות המתקבלות לפי נוסחאות (26) ו-(27) עדיפות על התחזית ממודל ARIMA לסדרה מקורית, שכן ניכוי השפעות החגים וימי הפעילות מקטין רעש בסדרה ומשפר את טיב התאמת המודל. שימוש בשיטות עמידות לחריגים בכל אחד משלבי האמידה (לרבות שלב אמידת גורמים עונתיים), עשוי להעלות את טיב החיזוי ודיוקו. שיטות אמידת התחזיות (1) ו-(2), לפי נוסחאות (26) ו-(27), בהתאמה, הושו תוך שימוש במדד MAPE (נוסחה 19) לעיל). כל ההרצות נעשו באמצעו התוכנית X-12-ARIMA גרסה 3.

אמידת סטיות התקן לסדרה מנוכת עונתיות מהווה בעיה סטטיסטית מורכבת שלא פתורה במלואה עד כה. הסיבה לכך נעוצה בעובדה, שהאלגוריתם X-12-ARIMA מהווה שילוב של שיטות פרמטריות (התאמת מודל ARIMA ואמידת השפעת החגים וימי הפעילות) ושיטות לא פרמטריות של ממוצעים נעים המופעלות

על הסדרה בצורה איטרטיבית. לכן, סטיות התקן לתחזיות $\hat{y}_{t+r}^{(1)}$ ו- $\hat{y}_{t+r}^{(2)}$ לא חושבו בעבודה זו.

ניתוח סדרה עתית כפי שמתואר לעיל בוצע על סדרות של מספר אירועי פשיעה מסוג k , בשטח תחנת משטרה m . על מנת לחשב שיעורי הפשיעה חזויים ל-1000 תושבים $\hat{Y}_{m,k,t+r}$, יש לאמוד תחזית לגודל אוכלוסייה באותה היחידה הגיאוגרפית, לכל רבע. חישוב זה נעשה על סמך הנוסחה (19). על מנת לאפשר השוואה של שיטות TS1 ו-TS2 עם שיטות הרגרסיה, מדדי טיב החיזוי חושבו ברמה שנתית בלבד.

4. נתונים וסטטיסטיקה תיאורית

4.1 בסיסי נתונים

הוכן בסיס נתונים ייחודי על מנת לאפשר את ביצוע הפרויקט, המבוסס על קישור רשומות מתוך בסיס נתוני המשטרה, שהתקבל מהמטה הארצי, מרשמי האוכלוסין לשנים הרלוונטיות לחקירה, קובצי חינוך ובסיס נתוני עובד/מעביד, ברזולוציה גיאוגרפית הנבחרת, כפי שיפורט להלן. לוח A1 בנספח A מכיל את הרשימה המלאה של המשתנים שנבנו ונחקרו במסגרת העבודה הנוכחית.

בסיס הנתונים המשטרתית. בסיס הנתונים המשטרתית מכיל 2 טבלאות עיקריות – טבלת התיקים וטבלת העבריינים. טבלת התיקים מכילה מידע על כל התיקים שנפתחו במשטרה. טבלת העבריינים כוללת מידע על כל העבריינים אשר התיק שנפתח במשטרה מיוחס אליהם. מבסיס הנתונים ניתן היה לבנות את

המשתנים הבאים ברזולוציה גיאוגרפית מתאימה: שיעור התיקים ל-1000 תושבים בשנה, שיעור העבריינים ל-1000 תושבים באותה שנה ואחוז הגילויים בתיקים שנפתחו באותה שנה. בנוסף התקבלה טבלה המפרטת את התקן ומצבת השוטרים בתחנות המשטרה, בפירוט המקצועות העיקריים (אג"מ, אח"ק ואמ"ן). מטבלה זו ניתן היה לחשב את מצבת או פרופרציית השוטרים על פי מקצועות ואת שיעור השוטרים ל-1000 תושבים באותה השנה.

מרשם האוכלוסין. מרשם האוכלוסין מכיל מידע דמוגרפי אודות תושבי מדינת ישראל, וכך ניתן היה לבנות מספר משתנים דמוגרפיים ברזולוציה הגיאוגרפית הנבחרת. כך, נבנו המשתנים הבאים: פרופרציית היהודים ואחרים באוכלוסייה, מבנה הגילים, יחס התלות, פרופרציית עולי בריה"מ לשעבר ופרופרציית העולים מאתיופיה. זיווג מרשם האוכלוסין עם קובץ רמת הדתיות שנבנה בלמ"ס על סמך אוסף קבצי חינוך איפשר חישוב פרופרציית האוכלוסייה החילונית, דתית (שאינה חרדית) וחרדית.

קובצי חינוך. בלשכה המרכזית לסטטיסטיקה מדי שנה מתקבלים קבצי נתונים ממועד החינוך. בין היתר, מתקבלים קבצים על אוכלוסיית התלמידים הלומדים בבתי ספר במסגרת משרד התמ"ת ומידע על בני נוער השייכים למסגרות כגון "קידום נוער". כך חושבה עבור האוכלוסייה פרופרציית בני נוער הלומדים במסגרת התמ"ת ופרופרציית בני הנוער השייכים למסגרת "קידום נוער", ומשתנים אלה יכולים להוות אינדיקציה על פוטנציאל של עבריינות בקרב בני נוער.

בסיס נתוני עובד/מעביד. מדי שנה מתקבל מרשות המסים מידע על כל העובדים, המעבידים, ההכנסה משכר, היקפי משרה ועוד. מתוך בסיס נתונים זה חושבו מספר משתנים אגרטיביים על האוכלוסייה בהסתמך על הרזולוציה הגיאוגרפית הנבחרת – פרופרציית השכירים בגילי העבודה, פרופרציית העצמאיים בגילי העבודה, פרופרציית הקטינים המועסקים, ההכנסה השנתית החציונית, פרופרציית המשתכרים הנמצאים בחמישון הכנסה ארצי תחתון ועליון, פערי ההכנסה באזור הגיאוגרפי בין ממוצע שכר השכירים באזור הנמצאים בחמישון ההכנסה העליון וממוצע שכר השכירים באזור הנמצאים בחמישון ההכנסה התחתון.

בסיס נתוני עובד/מעביד זווג עם מרשם המשפחות המנהליות בישראל שנבנה בלמ"ס על בסיס מרשמי האוכלוסין, וכך ניתן היה לחשב גם משתנים כגון חציון הכנסה משפחתית באזור, חציון הכנסה שנתית לנפש סטנדרטית, פרופרציית חמישוני הכנסה משפחתית באזור ופרופרציית המשפחות המלאות.

שימוש במערכות המידע הגיאוגרפיות (GIS) בלמ"ס. בסיס הנתונים המשטרתי עוגן במערכות המידע הגיאוגרפיות (להלן: ממ"ג) בלמ"ס לאזורים סטטיסטיים של מפקד 2008, וכך התקבל איתור של אירועי הפשיעה (בטבלת התיקים) ואיתור מגורי העבריינים (בטבלת העבריינים). יש לציין כי מדובר באיתור חלקי – 45% מהעבריינים ו-40% מהתיקים לא אותרו בשל חוסרים בנתוני הכתובת או מכיוון שמדובר בישוים ללא כתובות, דוגמת הישוים הערבים. לכן, לא נעשה שימוש ברזולוציה גיאוגרפית זו. כמו כן, חושבו בממ"ג מרחקים אוויריים בין יחידות גיאוגרפיות שונות, ונקבעו יחסי שכנות בין ישובים ובין שטחי תחנות משטרה בהסתמך על קיום גבולות משותפים. בנוסף, חושבו אורכי הגבולות הני"ל עבור רזולוציה של תחנות המשטרה.

בניית משתנה מוסבר: שיעור תיקי עבירות משני סוגים עיקריים ל-1000 תושבים.

הסיווג הסטטיסטי של העבירות הינו מסועף מאוד ורחב היקף. במטה הארצי של המשטרה הוחלט לבצע ניתוח תיקים מסוגים אלימות ורכוש עיקריים, בהתאם להגדרת יעד מפכ"ל. מפרט ההגדרות של משתנים אלה מוצג בנספח A2. המשתנה המוסבר הינו שיעור התיקים הנפתחים ברזולוציה גיאוגרפית ובנקודות זמן רלוונטיות, ל-1000 תושבים. שיעור יחסית לגודל אוכלוסייה נבחר בגלל שספירה של תיקים לא תשקף

את רמת הפשיעה במקום מסוים, שכן באזורים גדולים ומרובי אוכלוסייה מספר התיקים יהיה רב יותר מאשר במקום קטן יותר.

לצורך מתן תחזיות הפשיעה נבחרו שתי רזולוציות גיאוגרפיות: רשויות מקומיות בישראל – קרי, עיריות, מועצות מקומיות ומועצות אזוריות, ושטח תחנות משטרה, כאשר תחנות המשטרה כוללות בדרך כלל כמה רשויות מקומיות. כדי ליצור מצב שבו רזולוציית הרשויות המקומיות תהיה מוכלת ברזולוציית תחנות המשטרה הוחלט לאחד את 6 תחנות המשטרה בירושלים לתחנה אחת, וכך גם לגבי 4 תחנות המשטרה של ת"א-יפו. כמו כן, היה צורך קבוע מהי תחנת המשטרה אליה שייכות המועצות האזוריות אשר בשטחן ישנם יותר מתחנת משטרה אחת. לצורך כך, נקבע כי תחנת המשטרה שבשטחה מתגוררת מרבית אוכלוסיית המועצה האזורית היא תחנת המשטרה הקובעת.

4.2 סטטיסטיקה תיאורית של נתוני פשיעה בישראל

במהלך השנים 2003-2009 חלה ירידה הדרגתית של סך תיקי העבירות (ראו לוח 2). הירידה ניכרת בעיקר בארבע השנים האחרונות של התקופה, ובסך הכל מספר התיקים שנפתחו בשנת 2009 נמוך בכ- 19% ממספר התיקים שנפתחו בשנת 2003. לגבי תיקי הרכוש העיקריים - בתחילת התקופה (2003-2005) חלה עלייה די משמעותית במספר התיקים, ובהמשך (2006-2009) חלה ירידה דרסטית במספרם, כך שמספר תיקי הרכוש העיקריים שנפתחו בשנת 2009 נמוך בכ- 25% ממספר התיקים שנפתחו בשנת 2003. מספר תיקי האלימות נשאר יציב עם תנודות קטנות מאוד לאורך התקופה. בסך הכל תיקי עבירות האלימות ותיקי הרכוש העיקריים שבסל המשטרתי מהוות כ- 30% מכלל התיקים הנפתחים.

לוח 2: סך תיקי עבירות בשנים 2003-2009, לפי סוגים (כלל ארצי)

שנה	סך התיקים לפי סוג עבירה				אחוזים מסה"כ		
	רכוש עיקריים	אלימות	שאר התיקים	כלל התיקים	רכוש עיקריים	אלימות	שאר התיקים
2003	83,444	65,983	355,984	505,119	16.52	13.06	70.48
2004	90,194	67,978	361,238	519,048	17.38	13.10	69.60
2005	95,972	65,572	334,076	495,269	19.38	13.24	67.45
2006	92,562	66,621	319,254	478,050	19.36	13.94	66.78
2007	82,692	64,075	299,471	445,869	18.55	14.37	67.17
2008	70,288	63,476	288,035	421,477	16.68	15.06	68.34
2009	62,919	61,732	283,886	408,537	15.40	15.11	69.49

בלוח 2, מספר התיקים אינם מסתכמים בדיוק לכלל התיקים, וכך גם האחוזים אינם מסתכמים בדיוק ל-100.0% מכיוון שישנה ספירה כפולה של כ-300-400 תיקים מדי שנה שבהם יש הן סעיפי עבירות אלימות והן סעיפי עבירות רכוש. שאר התיקים הם תיקים שבהם אין בכלל סעיפי עבירות אלימות ורכוש עיקריים. קיימים הבדלים משמעותיים בין הישובים היהודים (והמעורבים) כאשר האוכלוסייה היהודית מהווה לפחות 50% מאוכלוסיית הישוב, להלן "ישובים יהודיים") לבין הישובים הערבים בנתוני הפשיעה. מנתוני שנת 2008 המוצגים בלוח 2 ניתן ללמוד על כמה הבדלים בולטים:

1. בישובים יהודיים שיעורי תיקי רכוש עיקריים גבוהים באופן משמעותי מאשר בישובים ערבים, ולעומת זאת שיעורי תיקי האלימות שבהם נמוכים בהשוואה לנתונים המקבילים בישובים הערביים.
2. בישובים יהודיים שיעורי תיקי רכוש עיקריים גבוהים באופן משמעותי משיעורי תיקי אלימות, ואילו בישובים הערביים התמונה הפוכה: הפשיעה מסוג אלימות שכוחה הרבה יותר מאשר הפשיעה מסוג רכוש.

3. אחוז הגילויים בתיקים (הן תיקי אלימות והן תיקי רכוש עיקריים) בישובים יהודים נמוך יותר מאשר בישובים ערביים. כך, אחוז הגילויים של תיקי רכוש בישובים ערביים גדול כמעט פי שניים בהשוואה לישובים יהודים (26.3% לעומת 14.3%).

4. שיעור העבריינים בישובים יהודים נמוך משיעור זה בישובים ערביים. לאור הבדלים אלו ומתוך מחשבה שסט המשתנים המתואמים עם שיעורי התיקים בין המגזרים שונה זה מזה, חושבו המתאמים בנפרד עבור כל אחד מהמגזרים. יצוין, כי המתאמים להלן הינם מתאמים גולמיים אשר אינם לוקחים בחשבון את מגוון גורמים נוספים המשפיעים על שיעורי הפשיעה. רמת המובהקות עבור מקדמי המתאם נקבעה ברמה של 5%.

לוח 3: סיכום הנתונים המשטרתיים (לשנת 2008), לפי מגזר מוצגים הממוצעים, בסוגריים מוצגים סטיית התקן

ישובים ערביים N=73		ישובים יהודים ומעורבים N=165		
תיקי אלימות	תיקי רכוש	תיקי אלימות	תיקי רכוש	
6.2 (3.3)	1.6 (1.1)	7.4 (4.5)	11.1 (8.8)	שיעור תיקים ל-1000 תושבים
7.1 (3.2)	1.2 (1.4)	4.7 (2.8)	0.7 (0.7)	שיעור העבריינים ל-1000 תושבים
67.3 (14.9)	26.3 (23.7)	60.4 (12.9)	14.3 (11.3)	אחוז הגילויים

4.2.1 ישובים יהודים ומעורבים

מבדיקת מתאמים בין המשתנים הדמוגרפים, החברתיים, הכלכליים והמשטרתיים לבין שיעור תיקי האלימות/רכוש בישובים יהודים ניתן לראות כי קיימים הבדלים במתאמים בין המשתנים המסבירים שנבנו לבין שיעור תיקי האלימות או שיעור תיקי הרכוש.

תיקי אלימות:

בישובים שבהם אחוז האוכלוסייה הצעירה גדול יותר שיעור התיקים נמוך יותר. בישובים שבהם פרופורציית העולים מבריה"מ לשעבר ו/או אתיופיה גבוהה יותר שיעור התיקים גבוה יותר אף הוא. בישובים בעלי מדד חברתי-כלכלי נמוך יותר שיעור התיקים גבוה יותר. כמו כן, בישובים שבהם פרופורציית בני הנוער הלומדים במסגרת תמ"ת או שייכים למסגרת קידום נוער שיעור התיקים גבוה, בישובים שבהם ההכנסה החציונית משכר גבוהה יותר שיעור התיקים נמוך יותר. ככל ששיעור עברייני אלימות המתגוררים בישוב גבוה יותר, שיעורי תיקי האלימות בישוב גדל, כאשר מתאם זה נמצא מובהק סטטיסטית. לא נמצאה השפעה מובהקת של גורם ההרתעה (אחוז גילויים בתיקי אלימות בשנה שעברה) על שיעורי הפשיעה מסוג זה.

תיקי רכוש עיקריים:

שיעור תיקי הרכוש העיקריים נמוך יותר בישובים שבהם אחוז האוכלוסייה הצעירה גדול יותר; כמו כן שיעור התיקים גדול יותר כאשר אחוז האוכלוסייה המבוגרת גדול יותר. בישובים שבהם ישנם יותר עולים מבריה"מ לשעבר שיעור תיקי הרכוש נמוך יותר (השפעה בכיוון הפוך מזו שראינו בנתוני אלימות). שיעורי פשיעה מסוג עבירות רכוש עיקריות קטנים יותר בישובים בעלי אחוז גבוה יותר של אוכלוסייה חילונית. שיעור תיקי הרכוש גבוה יותר בישובים חזקים יותר מבחינה חברתית כלכלית: בישובים שבהם ההכנסה החציונית משכר גבוהה יותר שיעור התיקים גבוה יותר. ככל ששיעור העבריינים המתגוררים בישוב גדול יותר יורד שיעור תיקי הרכוש. כמו כן, נמצא כי קיימת השפעה של גורם ההרתעה: ככל שאחוז גילויי תיקי הרכוש גבוה יותר (בשנה הקודמת) כך קטן שיעור תיקי הרכוש.

4.2.2 ישובים ערביים

סט המשתנים בעלי מתאם יציב ומובהק עם שיעורי הפשיעה בקרב האוכלוסייה הערבית שונה מהסט המקביל במגזר היהודי. נמצא כי המתאם בין שיעור התיקים לבין שיעור העבריינים המתגוררים בישוב הינו חזק ומובהק הן בנתוני אלימות והן בנתוני רכוש.

מבנה הגילים המאפיין ישובים ערביים עם שיעורים גבוהים של תיקי אלימות ורכוש: יש יותר צעירים ופחות מבוגרים. משתנה בעל מתאם מובהק עם שיעורי תיקי האלימות והרכוש הוא שיעור המשפחות המנהליות המלאות בישוב. ככל שפרופורציית המשפחות המנהליות המלאות בישוב גבוהה כך יורדים שיעורי תיקי האלימות ותיקי הרכוש. משתנה נוסף בעל המתאם עם שיעורי הפשיעה הוא המרחק מעיר תל אביב-יפו. ככל שמתרחקים מהמרכז, שיעורי תיקי הרכוש העיקריים והאלימות קטנים יותר.

ההבדלים העיקריים שנמצאו בין המתאמים של שיעורי תיקי האלימות ושיעורי תיקי הרכוש מצויים דווקא במשתנים הכלכליים: בישובים חלשים יותר מבחינה כלכלית, שיעורי תיקי אלימות גבוהים יותר. בישובים שבהם פרופורציית הקטינים המועסקים גדולה יותר ופרופורציית השכירים הנמצאים בחמישון העליון נמוכה יותר שיעור תיקי האלימות גבוה יותר.

שיעור תיקי הרכוש גבוה יותר בישובים גדולים יותר, בניגוד לממצא מקביל עבור שיעור תיקי האלימות. כמו כן, נמצא כי יש השפעה של הגורם הרתעה, וככל שאחוז הגילויים בתיקי האלימות בישוב גבוה יותר (בשנה הקודמת) כך קטן שיעור תיקי האלימות.

4.2.3 תחנות משטרה

רזולוציה גיאוגרפית של תחנות משטרה נמוכה יותר מזו של רשויות מקומיות (בד"כ תחנת משטרה מכילה בתוכה כמה רשויות מקומיות), וכן מספר תחנות המשטרה קטן יותר. בנוסף, ברזולוציה של תחנות המשטרה אין הפרדה בין לפי מגזר (יהודי וערבי). ברמה של תחנות משטרה, נמצאו מתאמים הבאים בין המשתנים המסבירים המתוארים לעיל לבין שיעורי הפשיעה:

1. יש שיעור גבוה יותר של תיקי רכוש באזורים שבהם האוכלוסייה מבוגרת יותר.
2. משתנים המתוארים רמה כלכלית של אוכלוסייה מתואמים בעיקר עם שיעור תיקי רכוש: באזורים החזקים יותר מבחינה כלכלית שיעורי תיקי הרכוש גבוהים יותר. לא נמצאו מתאמים מובהקים בין שיעור תיקי האלימות לבין המשתנים הכלכליים.
3. נמצא כי מספר השוטרים בתחנה מתואם שלילית עם שיעור תיקי האלימות, ולא נמצא מתאם מובהק עם שיעור תיקי הרכוש. כמו כן, שיעור תיקי הרכוש מתואם חיובית עם אחוז שוטרי האח"ק בתחנה. נמצא כי שיעור העבריינים מתואם חיובית עם שיעור תיקי האלימות. ניתן למצוא את אפקט ההרתעה בנתוני עבירות רכוש בלבד, כאשר ככל שאחוז הגילויים בתיקי הרכוש בתחנת המשטרה היה גדול יותר בשנה הקודמת, כך בשנה הזאת יקטן שיעור תיקי הרכוש העיקריים.

5 תוצאות אמידה

5.1 מודלים להסבר שיעורי פשיעה

מניתוח מקדמי המתאם בין המשתנים המסבירים המוצגים בלוח A1 עולה, כי לא ניתן להשתמש בכל אוסף המשתנים שייצרנו במודלים של רגרסיה. לכן, למודלים שונים נבחרו סטים שונים של משתנים מסבירים בהתאם לקריטריונים לגבי מקדמי מתאם, כושר הסבר ויכולת חיזוי. הקריטריונים לגבי מקדמי המתאם במודלים שונים מפורטים בפרק 3. כושר הסבר של שיעורי פשיעה נבחן על סמך רמת המובהקות של מקדם המשתנה המסביר במודל הנאמד ומדד $Partial R^2$, על פני תקופת המחקר. יכולת החיזוי

נבחנה באמצעות הבדיקה: האם כתוצאה מהכנסת המשתנה המסביר שנבדק, מדד ה-MAPE יורד, ואם כן, האם תוקף סטטיסטי של מודל לא נפגע (קיום מולטיקוליניאריות, תלות בשאריות וכו').
כאמור בסעיף 1.2, בעבודה הנוכחית נעשתה הבחנה בין מודלים להסבר תופעת הפשיעה ולבין מודלים לחיזוי של שיעורי הפשיעה מסוגים שונים. ההבחנה נובעת מכך שהקריטריונים לבחירת המודל הטוב ביותר משתנים לפי המטרה לשמה נאמד המודל.

נאמדו 4 מודלים להסבר פשיעה – לפי 2 סוגי פשיעה (עבירות רכוש עיקריות ואלימות), ב-2 רזולוציות מרחביות (ישוב/מועצה ותחנת משטרה). נמצא, כי למטרות ההסבר כדאי להשתמש במודלים OLS2, ברזולוציה של ישובים/מועצות, ו- OLS1 ברזולוציה של תחנות משטרה. מודלים אלה נבחרו בזכות יכולתם לנתח מספר רב של משתנים "מועמדים" להיות משתני המודל, ולבחור סט המובהק ביותר, ברמת מובהקות הנקבעת מראש, לכל שנה. כמו כן, אי-הכללת משתנים שמקדמיהם אינם מובהקים מונעת ניפוח מדדי טיב התאמת המודל כתוצאה מריבוי הפרמטרים הנאמדים. סיבת ההעדפה של מודלים OLS על פני המודלים המרחביים למטרות הסבר תופעות הפשיעה מסוג עבירות רכוש עיקריות תידון בסעיף 4.2.1 להלן.

במחקר הנוכחי, רף המובהקות להכנסת המשתנה למודל נקבע ברמת 10%. בכל המודלים הסופיים להסבר תופעת הפשיעה לא זוהו סטיות מהנחות מודל OLS. הלוחות בנספח B מציגים מקדם גרסיה, רמת המובהקות, ממוצע של ערכי המשתנה ומקדם גמישות, עבור כל משתני המודל.
מוציגים מודלים להסבר הפשיעה שנאמדו עבור שנת 2008, תוך שימוש במשתנים מסבירים עם פיגור בזמן של שנה אחת. יצוין, כי אין הבדל משמעותי בין התוצאות המוצגות עבור 2008 לבין התוצאות שהתקבלו עבור שנים אחרות בתקופת הניתוח (החל מ-2003). ניתוח גורמים המשפיעים על שיעורי פשיעה נעשה על סמך גמישויות המשתנים הרלוונטיים.

5.1.1 מודלים להסבר שיעורי פשיעה ברמת ישוב/מועצה

לוח B1 מראה השפעתם של גורמים שונים על שיעורי פשיעה מסוג אלימות ברזולוציה של ישובים/מועצות. היות ומדובר במודל OLS2, נאמדו 2 מודלים: עבור המגזר היהודי ועבור המגזר הערבי.
(1) המגזר היהודי: ממודל עבור המגזר היהודי ניתן לראות, כי הגורם בעל ההשפעה החזקה ביותר על שיעורי הפשיעה בשנה t הוא שיעורי הפשיעה מאותו הסוג בשנה $t-1$. כך, עליה ב-1% בשיעורי הפשיעה בשנה כלשהי גורמת בממוצע לעליה בשיעורי הפשיעה בשנה הבאה 0.7658%. כמו כן, שיעורי הפשיעה מסוג אלימות מושפעים חיובית משיעורי הפשיעה מסוג עבירות רכוש עיקריות שבוצעו באותו ישוב: עליית שיעורי הפשיעה מסוג עבירות רכוש עיקריות ב-1% גורמת בממוצע לעלייה של הפשיעה מסוג אלימות ב-0.0928%. קיימת השפעה שלילית מובהקת של שינוי בפרופורציית הגילויים בתיקים שנפתחו על שיעורי הפשיעה מסוג אלימות (גורם ההרתעה).

בנוסף, נמצאה השפעה חיובית של המשתנים הבאים על שיעורי הפשיעה מסוג אלימות: פרופורציית עולי בריה"מ לשעבר (משנת 1990), פרופורציית בני נוער במסגרות התמי"ת ושינוי בפערי הכנסות משכר. נמצאה השפעה שלילית של המשתנה: שינוי באחוז השייכים לחמישון התחתון של הכנסות.
בנתוני אלימות נמצאה השפעה מרחבית חיובית על רמת שיעורי פשיעה. עם התרחקות הישוב מעיר תל-אביב ב-1%, שיעורי הפשיעה מסוג אלימות גדלים ב-0.05% בממוצע.

(2) המגזר הערבי: בדומה למגזר היהודי, במגזר הערבי נמצאה השפעה חזקה של שיעורי פשיעה מסוג אלימות בשנה שעברה על שיעורי הפשיעה מאותו סוג בשנה הנוכחית. בנוסף, זוהו מספר גורמים המגבירים את שיעורי הפשיעה מסוג אלימות במגזר הערבי: פערי הכנסות, גידול בפרופורציית השייכים לחמישון

התחתון של הכנסות וגידול של פרופורציית השייכים לחמישון העליון של הכנסות. בניגוד למגזר היהודי, במגזר הערבי שיעורי הפגיעה מסוג אלימות הולכים וקטנים עם התרחקות מתל-אביב, כאשר עוצמת ההשפעה גבוהה בהרבה מזו שנצפית במגזר היהודי: עם הגדלת המרחק מתל-אביב ב-1%, שיעורי הפגיעה קטנים ב-0.5188% במוצע.

בלוח B2 מוצגים משתנים שנכללו במודל OLS2 עבור שיעורי פגיעה מסוג עבירות רכוש עיקריות ברזולוציה של ישובים/מועצות. בדומה ללוח B1, לוח B2 מציג 2 מודלים - עבור המגזר היהודי ועבור המגזר הערבי.

1) המגזר היהודי: נמצא, כי גידול ב-1% בשיעורי הפגיעה מסוג עבירות רכוש עיקריות בשנה שעברה גורם לגידול שיעורי הפגיעה מאותו הסוג בשנה זו ב-0.7763%. בנוסף, גידול בשיעורי הפגיעה מסוג אלימות בשנה שעברה ב-1% גורם לגידול שיעורי הפגיעה מסוג בעבירות רכוש עיקריות בשנה זו ב-0.1193%. עלייה ב-1% בפרופורציית גילויים בתיקים שנפתחו מורידה את שיעורי הפגיעה מסוג עבירות רכוש עיקריות ב-0.1373% במוצע. יש להדגיש, כי עוצמת ההשפעה של עלייה בפרופורציית גילוי בתיקים על שיעורי הפגיעה חזקה יותר בנתוני עבריינות מסוג עבירות רכוש עיקריות מאשר בנתוני עבריינות מסוג אלימות.

בנוסף, נמצאה השפעה חיובית של המשתנים הבאים על שיעורי הפגיעה מסוג עבירות רכוש עיקריות במגזר היהודי: פרופורציה של אוכלוסייה דתית (לא חרדית), מדד פערי הכנסות (עוצמת השפעה זו היא הגדולה ביותר מבין הגורמים החברתיים-כלכליים שהשפעתם נמצאה מובהקת) ועלייה ביחס תלות. במגזר היהודי, נמצאה השפעה שלילית של המשתנה המרחבי "מרחק מעיר תל-אביב" על שיעורי הפגיעה מסוג עבירות רכוש עיקריות.

2) המגזר הערבי: ניתוח השפעות של משתנים משטרתיים מראה, כי גידול שיעורי הפגיעה בשנת $t-1$ ב-1% יגרום לגידול שיעורי הפגיעה בשנת t ב-0.588%. בדומה, קיימת השפעה חיובית של שיעורי הפגיעה מסוג אלימות על שיעורי הפגיעה מסוג עבירות רכוש עיקריות – גידול של שיעור עבירות אלימות ב-1% בשנה שעברה יגרום לגידול בשיעור עבירות מסוג עבירות רכוש עיקריות ב-0.5279% בשנה זו.

נמצאה השפעה שלילית של המשתנים הבאים על שיעורי הפגיעה מסוג עבירות רכוש עיקריות במגזר הערבי: פרופורציית עצמאים, כאשר עלייה של פרופורציית עצמאים ב-1% תגרום לירידה בשיעורי פגיעה מסוג עבירות רכוש עיקריות ב-0.5531%, שינוי בפערי הכנסות משכר ושינוי ביחס תלות.

עבור שיעורי עבירות רכוש עיקריות, התוצאות ממודל OLS2 לעיל הושוו לתוצאות שהתקבלו ממודל מרחבי SER1 (התוצאות המלאות זמינות אצל החוקרים). בחירת מודל מרחבי SER1 להשוואה מול מודל OLS2 נובעת מקלות יחסית של פירוש המקדם המרחבי במודל לשונו. במודל SER1, המקדם המרחבי בשונות הינו חיובי, מה שתומך בהנחת המחקר שככל שמרחק בין שני הישובים גדל, כך התלות המרחבית הולכת וקטנה (ההנחה בנוסחה (4) למשקלות מרחביים עבור רזולוציה של ישובים/מועצות).

ההבדלים בין מקדמי המשתנים המסבירים ובין הגמישויות שהתקבלו ממודל OLS2 וממודל SER1 אינם משמעותיים. היות ולא זוהו סטיות מובהקות מהנחות מודל OLS בניתוח שאריות של OLS2, ניתן להעדיף מודל זה על פני מודל מרחבי למטרות הסבר פגיעה. ממצא זה עולה בכנה אחד עם הממצאים שדווחו בספרות (ראה למשל רגב, 2008).

5.1.2 מודלים להסבר שיעורי פגיעה ברמת תחנות משטרה

אחד המטרות המרכזיות של אמידת מודלים להסבר שיעורי פגיעה ברמת תחנות משטרה – אמידת ההשפעה של גורם השיטור על שיעורי פגיעה, שכן משתני שיטור מוגדרים ברמת תחנות משטרה ולא

מוגדרים היטב ברמת ישוב/מועצה. האמידה בוצעה תוך פיקוח על משתנים חברתיים-כלכליים ודמוגרפיים המאפיינים אוכלוסייה המתגוררת בשטח שיפוט התחנה. בכל המודלים, לא נמצאה השפעה מובהקת של המשתנה "סך השוטרים בתחנה" על שיעורי הפשיעה, גם לא בפיגור של שנה אחת, שנתיים או שלוש. לכן, נבדקה השפעה פרטנית של כל אחד מ-3 מקצועות השיטור: אג"מ, אמ"ן ואח"ק. נמצא מתאם שלילי בין שיעור השוטרים בכל אחד מהמקצועות הנ"ל לבין שיעורי הפשיעה, הן מסוג אלימות והן מסוג עבירות רכוש עיקריות. עם זאת, מקדם המתאם נמצא מובהק סטטיסטית בנתוני אג"מ בלבד (ברמת מובהקות 10%, לכל השנים עבורן חלוקת השוטרים למקצועות זמינה), ולכן משתנה "אחוז שוטרי אג"מ מכלל השוטרים בתחנה" נכלל במודלים סופיים.

נזכיר כי בנתוני אלימות ברזולוציה מרחבית של תחנת משטרה נאמד מודל OLS1 בלבד, ותוצאות האמידה מוצגות בלוח B3. ניתן לראות, כי גידול בשיעורי הפשיעה מסוג אלימות בשנה $t-1$ ב-1% גורם לגידול בשיעורי הפשיעה מאותו הסוג בשנה t ב-0.8973%. הגדלת אחוז שוטרי אג"מ בתחנה ב-1% תקטין את שיעורי הפשיעה מסוג אלימות ב-0.2574%.

בין השפעות הגורמים החברתיים-כלכליים והדמוגרפיים אשר נמצאו מובהקות, ניתן למנות: השפעה חיובית של פרופורציית השכירים ופרופורציית המשפחות המלאות; השפעה שלילית של פרופורציית העצמאים.

לוח B4 מציג תוצאות האמידה של מודל OLS1 להסבר שיעורי הפשיעה מסוג עבירות רכוש עיקריות. יצוין, כי בדומה לרזולוציה של ישובים/מועצות, תוצאות המודל המרחבי SER ברזולוציה של תחנת משטרה דומות מאוד לתוצאות OLS1, כאשר המקדם המרחבי אינו מובהק. עם זאת יש לציין, כי מודל OLS1 לתחנות משטרה אינו מתעלם מתלות מרחבית, כאשר המשתנה המרחבי "שיעור פשיעה בתחנות הגובלות" נכלל במודל.

לוח B4 מראה, כי הגדלת אחוז שוטרי אג"מ בתחנה ב-1% תגרום להורדת שיעורי הפשיעה מסוג עבירות רכוש עיקריות ב-0.4868% בממוצע. יודגש, כי הגדלת אחוז שוטרי אג"מ בתחנה תגרום להורדת הפשיעה מסוג עבירות רכוש עיקריות בעוצמה גבוה כמעט פי שתיים מאשר להורדת הפשיעה מסוג אלימות. נמצאה השפעה חיובית חזקה של שיעורי הפשיעה מאותו הסוג בתחנות השכנות (בעלות גבול משותף) – גידול פשיעה בתחנות השכנות ב-1% מגדיל שיעורי הפשיעה מסוג עבירות רכוש עיקריות בתחנה הנחקרת ב-1.6994%. בנוסף נמצא, כי גידול ב-1% בפערי הכנסות יגדיל שיעורי פשיעת עבירות רכוש עיקריות ביותר מ-2%, וגידול ב-1% בפרופורציית המשפחות המלאות יקטין שיעורי פשיעת עבירות רכוש עיקריות ב-1.3887%.

5.2 מודלים לחיזוי - בחירת המודל הטוב ביותר והתחזיות המתקבלות

לצורך בניית תחזיות, בשתי רזולוציות מרחביות ועבור שני סוגי פשיעה, הושאו מודלים לחיזוי שיעורי פשיעה שנתיים כפי שתוארו בפרק 3 לעיל. לצורך בחירת השיטה הטובה ביותר, נאמדו תחזיות וסטיות תקן שלהן לשנה אחת, לשנתיים ולשלוש שנים קדימה, והושאו עם הנתונים בפועל עבור השנים בהן נתוני הפשיעה הינם זמינים. נזכיר, כי מזעור מדד MAPE יגרום להעלאת דיוק התחזיות, ומזעור סטיות התקן (מסומנות ב-Std בלוחות) – להעלאת יציבותן. המדדים הטובים ביותר (הקטנים ביותר) בכל שנה מודגשים בלוחות של נספח C, כמו גם המדדים המסכמים עבור המודל הטוב ביותר לפי הקריטריונים הנ"ל.

מלוח C1 ניתן לראות, כי בנתוני שיעורי פשיעה מסוג אלימות ברזולוציית ישובים/מועצות, המודל OLS2 מביא למזעור מדד MAPE על פני תקופת המחקר, אך מודל OLS1 – למזעור סטיות התקן. עם זאת ניתן

לראות, כי סטיות התקן המתקבלות משתי המודלים קרובות מאוד. כמו כן, מדדי MAPE יציבים יותר על פני תקופת המחקר במודל OLS2. לכן, נעדיף מודל OLS2 לצרכי חיזוי של שיעורי פשיעה מסוג אלימות ברזולוציה מרחבית של ישובים/מועצות. הסטייה הממוצעת על פני תקופת המחקר של התחזית מהנתון בפועל תחת המודל שנבחר: 24%, עם סטיית תקן ממוצעת 0.33.

לוח C2 משווה תוצאות המתקבלות מ-5 המודלים לחיזוי שיעורי פשיעה מסוג עבירות רכוש עיקריות ברזולוציה של ישובים/מועצות. בכל שנה, צוינו שני מודלים הטובים ביותר. הן מבחינת דיוק התחזיות והן מבחינת מזעור סטיות תקן, קיים יתרון ברור למודלים מרחביים, בהשוואה למודלים המבוססים על שיטת ריבועים פחותים (הפרש של כ-2 נקודות אחוז במונחי דיוק וכן הפרש של כ-0.1 בסטיות תקן). תוצאה זו ממחישה את הצורך בשימוש במודלים מרחביים כאשר קיימת תלות מרחבית מובהקת. התעלמות מקיום תלות זו מורידה את טיב החיזוי.

מדדי MAPE עבור המודלים המרחביים קרובים אחד לשני עם יתרון קל למודל SER1. מבחינת סטיות התקן, קיים יתרון קל למודל SGLS, אך סטיות תקן המתקבלות ממודלים SER1 ו-SER2 קרובים מאוד לאלו המתקבלים מ-SGLS. לכן, בהתחשב בדיוק ויציבות התחזיות המתקבלות, נבחר במודל SER1 לחיזוי שיעורי פשיעה מסוג עבירות רכוש עיקריות ברזולוציה מרחבית של ישובים/מועצות. בשיטה זו, הסטייה הממוצעת של תחזית מנתונים שהתקבלו בפועל הינה 36%, וסטיית תקן ממוצעת של תחזיות: 0.46.

ברזולוציה מרחבית של תחנות משטרה, אנו משווים שיטות ישירות, המבוססות על התאמת מודל לנתוני פשיעה ברזולוציה של תחנות המשטרה באופן ישיר (בדומה לרזולוציה של ישובים/מועצות) ושיטות בלתי ישירות המבוססות על חישוב תחזית לתחנה על סמך תחזיות ליחידות גיאוגרפיות המרכיבות אותה, כאשר האחרונות מתקבלות ממודלים לנתוני ישוב/מועצה שנסקרו לעיל.

מלוח C3 ניתן להסיק, כי עבור שיעורי פשיעה מסוג אלימות ברזולוציה מרחבית של תחנות משטרה, קיים יתרון למודל ישיר של OLS מבחינת הדיוק, ויתרון לשיטה בלתי ישירה על סמך תחזיות לישובים ממודל OLS1 מבחינת סטיות התקן. נציין, כי ההבדל בין סטיות התקן המתקבלות לפי שלושת השיטות שנבדקו בלוח C3 אינו משמעותי. נשים לב, כי בשיטה הישירה, טיב החיזוי רגיש מאוד לשינויים בנתונים ואינו יציב: מדד MAPE הקטן ביותר הינו 6.65%, עבור תחזית לשנתיים בשנת 2007, ומדד MAPE הגבוהה ביותר מושג עבור שנת 2009, בתחזית לשלוש שנים (34.1%). לעומת זאת, אף מדד MAPE בשיטה הבלתי ישירה ממודל OLS1 לא עבר גבול עליון של 19%, ולפי מודל OLS2 – לא עבר גבול של 18%. ניתן לראות, כי השיטה הבלתי ישירה המבוססת על OLS2 היא היציבה ביותר. בזכות עובדה זו, השיטה הבלתי ישירה המבוססת על מודל OLS2 נבחרה לחיזוי שיעורי פשיעה מסוג אלימות ברזולוציה של תחנות משטרה. שיקול נוסף לבחירת שיטה זו – עקביות עם השיטה שנבחרה ברמת ישובים/מועצות. תחת שיטה זו, הסטייה הממוצעת של התחזית מהנתונים בפועל הינה 15.8%, וסטיית התקן של תחזיות שווה בממוצע ל-0.3.

לוח C4 משווה שיטות חיזוי עבור שיעורי פשיעה מסוג עבירות רכוש עיקריות ברזולוציה של תחנות משטרה. ניתן לראות, כי קיים יתרון משמעותי לשיטות בלתי ישירות המבוססות על מודלים מרחביים על פני כל השיטות המתחרות. ממצא זה ממחיש יתרון של שיטת חיזוי בלתי ישירה בנתוני פשיעה מסוג עבירות רכוש עיקריות, תוך התחשבות בתלות המרחבית הקיימת בנתונים אלה. בנוסף ניתן לראות, כי המודל המרחבי הישיר (SER) לתחנות המשטרה אינו יציב וסטיות התקן המתקבלות הינן גבוהות. כשולן

המודל המרחבי הישיר בחיזוי שיעורי פשיעה ברזולוציה של תחנות משטרה יכול להצביע על הצורך בשיפור המודל למבנה התלות המרחבית ברזולוציה זו (8), מה שיהווה נושא להמשך המחקר. לפי מדדי MAPE, קיים יתרון לשיטה הבלתי ישירה המבוססת על מודל SER1. מבחינת סטיות התקן, למרות שקיים יתרון קל לשיטה המבוססת SGLS, לא ניתן להעדיף מודל כלשהו מבין שלושת השיטות המרחביות הבלתי ישירות. לכן, נעדיף שיטה בלתי ישירה המבוססת על מודל מרחבי SER1. שיקול נוסף לבחירת שיטה זו - עקביות אומדני תחזיות ברזולוציות מרחביות שונות: ישוב/מועצה ותחנת משטרה, שכן, כמו בנתוני אלימות, בשתי הרזולוציות האמידה מתבצעת על סמך אותו המודל. תחת המודל שנבחר, הסטייה הממוצעת של התחזית מהנתונים שהתקבלו בפועל שווה ל- 23.9%, וסטית התקן הממוצעת של תחזיות שווה ל- 0.29.

יצוין, כי המודל שנבחר לחיזוי שיעורי פשיעה מסוג עבירות רכוש עיקריות אינו זהה למודל להסבר הפשיעה מאותו סוג. שיטות החיזוי ברזולוציה של תחנות משטרה שונות אף הן ממודלים להסבר שיעורי הפשיעה ברזולוציה מרחבית זו, הן עבור פשיעה מסוג אלימות, הן עבור עבירות רכוש עיקריות. המודלים שנבחרו לחיזוי שיעורי הפשיעה נאמדו עם משתנים מסבירים בפיגור הולך וגדל בזמן - שנה, שנתיים ושלוש שנים. לפי שיטת האינדקסאטורים המובילים, התקבלו תחזיות לשיעורי הפשיעה לשנה, שנתיים ושלוש שנים קדימה, בהתאמה. באיור C1 מוצגים שיעורי פשיעה ברמה ארצית, תחזיות לעומת שיעורי פשיעה בפועל, עבור התקופה 2006 - 2009. בנוסף, מוצגים תחזיות לשנים 2010, 2011 ו- 2012. ניתן לראות, כי במידה והגורמים המשפיעים על שיעורי הפשיעה לא ישתנו בתקופת החיזוי ביחס לתקופה הנוכחית, ועוצמת השפעתם לא תשתנה משמעותית, לשלוש השנים הקרובות צפויה ירידה בשיעורי הפשיעה מסוג עבירות רכוש עיקריות (בכ- 28 אחוזים, משנת 2009 עד סוף תקופת החיזוי). בשיעורי פשיעה מסוג אלימות צפויה ירידה מתונה יותר (בכ- 12 אחוזים, משנת 2009 ועד סוף תקופת החיזוי). איור C2 מראה מדדי MAPE עבור שיעורי פשיעה ברזולוציה כלל-ארצית. ניתן לראות, כי הסטייה המרבית בין תחזית לבין נתון בפועל נצפתה בשנת 2006 בנתוני עבירות רכוש עיקריות (16%). מאז, הטעות היחסית הממוצעת של תחזיות פשיעה מסוג זה יורדת באופן עקבי, עד לרמת דיוק של 2% בשנת 2009. לגבי תחזיות פשיעה מסוג אלימות, נצפית תנודתיות בדיוק תחזיות, אך הסטייה המרבית (עבור שנת 2007) שווה ל- 12% בלבד. סטיית התחזית מהנתון בפועל בשנת 2009 שווה ל- 8% בלבד, בנתוני הפשיעה מסוג אלימות. הסטיות המתקבלות ברזולוציה זו נמוכות משמעותית מאלו המדווחים בספרות המקצועית.

5.3 תחזיות ממודל לסדרה עתית

נזכיר, כי ניתוח הסדרות כלל 2 שלבים - אמידת מרכיבי הסדרה באמצעות השיטה X-12-ARIMA ואמידת תחזיות, כאשר תקופת העיבוד נקבעה החל מרבע I שנת 2001, עבור 6 תחנות משטרה נבחרות. נמצא, כי בכל הסדרות קיימת השפעה מובהקת של החגים העבריים וימי הפעילות, והשפעות אלו נוכו מכל הסדרות שבניתוח (פעולת "התאמה מראש"). לסדרות המותאמות, ונבחר ונאמד מודל ARIMA, תוך איתור ואמידה של החריגים הקיצוניים.

ניתוח מדדי טיב ניכוי עונתיות - מבחן F לקיום עונתיות, מדדי תרומות יחסיות של מרכיבי הסדרה לשינויים בין שני נתונים עוקבים בסדרה ומדד מסכם M לטיב ניכוי עונתיות מראה, כי השפעה עונתית קיימת בכל הסדרות שנבדקו. תרומת מרכיב האי-סדירות ולא עולה על 15% וקטנה מתרומתו של מרכיב העונתיות, בכל הסדרות. לכן ניתן לסכם, כי כל 12 הסדרות שנתחו - עבור 6 תחנות משטרה ושני סוגי פשיעה - מושפעות ממיקום החגים העבריים בשנה, ממספר ימי פעילות בכל רבע וכן מתנודות עונתיות.

לוח D1 משווה תחזיות רבעוניות המתקבלות לפי הנוסחה (31) (שיטה TS1) לעומת תחזיות המתקבלות לפי הנוסחה (32) (שיטה TS2). היות והנתונים בתדירות רבעונית עבור 2009 לא היו זמינים בתקופת המחקר, נעשתה השוואה ברמה שנתית מול התחזית שנבחרה בסעיף 4.2 לעיל עבור התחנות (לפי שיטה בלתי ישירה המבוססת על המודל OLS2 עבור נתוני אלימות, ולפי שיטה בלתי ישירה המבוססת על המודל SER1 עבור נתוני עבירות רכוש עיקריות). ברמה שנתית ניתן לחשב שיעורי פשיעה ל- 1000 תושבים (היות וגודלי האוכלוסייה זמינים בתדירות שנתית), מה שמקל על ההשוואה בין שיטות שונות. לצרכי ההשוואה חושבו מדדי MAPE לשלושת השיטות עבור 6 התחנות שבניתוח ושני סוגי הפשיעה.

מלוח D1 ניתן לראות, כי קיים יתרון, במונחי דיוק תחזיות, לשיטות שנתיות בלתי ישירות שנבחרו לעיל, על פני שיטות המבוססות על ניתוח סדרה עתית, הן בנתוני אלימות, הן בנתוני עבירות רכוש עיקריות. יתרון זה בולט במיוחד בנתוני עבירות רכוש עיקריות. אחת הסיבות האפשריות לכך - קיום נקודת המפנה בשנים 2005 – 2006 בסדרות עבירות רכוש עיקריות, מה שמקשה על חיזוי של סדרות אלו. בהשוואה בין שתי השיטות המבוססות על חיזוי סדרה עתית, קיים יתרון לשיטה TS2, הן בנתוני אלימות, הן בנתוני עבירות רכוש עיקריות. לפי שיטה TS2, עבור התחנות הנבחרות, מדד MAPE באירועי אלימות שווה ל- 6.6% (מדד זה נמוך בהרבה מהמדדים המקבילים המדווחים בספרות), ובאירועי עבירות רכוש עיקריות - 24.2%.

על אף העובדה ששיטה TS2, המבוססת על חיזוי סדרה עתית באמצעות הנוסחה (32), פחות מדויקת במונחי MAPE בתדירות שנתית מאשר שיטה בלתי ישירה מהסעיף הקודם, תחזיות המתקבלות בתדירות רבעונית טומנות בתוכן אינפורמאציה נוספת אודות הצפי למספר אירועי הפשיעה מסוגים שונים לתקופה של רבעון אחד עד ארבע רבעונים הבאים. מידע זה יכול להוות בסיס לתכנון טקטי של פעילות משטרתית, לרבות פריסת כוחות וניוד משאבים בטווח קצר. נראה, כי לא כדאי לאמוד תחזיות על סמך סדרה עתית לתקופה העולה על שנה, שכן דיוק התחזיות ומהימנותן יורדים במהירות עם התרחקות מהנתון העדכני האחרון, במיוחד בנתוני עבירות רכוש עיקריות. יודגש כי לפי השיטות שנדונו במסגרת העבודה הנוכחית, סכום התחזיות הרבעוניות עבור שנה מסוימת אינו שווה לתחזית שנתית. על ידי שימוש בשיטות benchmarking (ראה גובמן, 2008) ניתן להתאים תחזיות רבעוניות לתחזית שנתית, אך התאמה זאת מצריכה בדיקה נוספת בנוגע למודל לחלוקה בינרבעונית בתוך השנה, מה שחורג ממסגרת העבודה הנוכחית ומהווה נושא למחקר עתידי.

6. דיון ומסקנות

בעבודה הנוכחית נסקרו והשוו שיטות שונות לחיזוי הפשיעה מסוגים שונים וברזולוציות גיאוגרפיות שונות. פותחו מודלים לחיזוי שיעורי הפשיעה בתדירות שנתית, ומספר אירועי הפשיעה בתדירות רבעונית. בנתונים שנתיים, לכל סוג פשיעה ולכל רזולוציה מרחבית נבחרה השיטה הטובה ביותר להסבר תופעת הפשיעה ולחיזוי הפשיעה, כאשר המודלים להסבר ולחיזוי עשויים להיות שונים. שימוש במודלים מרחביים - אוטורגרסיביים העלה את טיב התחזיות וסיפק מידע נוסף בנוגע להתפתחות הפשיעה במרחב ולאורך זמן. במסגרת המחקר, התייחסנו לתלות מרחבית הן דרך התוחלת, תוך שימוש במשתנים מסבירים מרחביים, הן דרך השונות, תוך בניית מטריצות מרחביות לשונות בנתוני עבירות רכוש עיקריות. בחיזוי פשיעה, הושג דיוק גבוה ביחס למדדים המדווחים בספרות. הסטייה הממוצעת של התחזיות מהנתון המתקבל בפועל עבור אותה השנה נע בין 15% ל- 25%, למעט נתוני פשיעה מסוג עבירות רכוש עיקריות ברזולוציה של ישוב/תחנה, שם רמת הדיוק הממוצעת נעה סביב 35%. בהסתמך על רמות דיוק אלו, יחד

עם רמה נמוכה של סטיות התקן לתחזיות כפי שדווחה בפרק 5, ניתן להסיק על טיב גבוה של תחזיות שנאמדו, מה שמאפשר שימוש בהם לטווח קצר, בינוני ארוך על ידי מקבלי ההחלטות במשטרת ישראל. בניית תוחות נתונים שנתיים, ניתן לסכם את הממצאים העיקריים:

(1) זוהתה ונאמדה התלות המרחבית בנתוני פשיעה מסוג עבירות רכוש עיקריות, ברזולוציה מרחבית של ישובים/מועצות. בנתוני פשיעה מוג אלימות לא נמצאה תלות מרחבית מובהקת. ברזולוציה מרחבית של תחנות משטרה, זוהתה תלות מרחבית בנתוני עבירות רכוש עיקריות, אך תלות זו נמצאה מובהקת רק בשנים 2007 - 2009.

(2) נבנו מודלים תלויי זמן ומרחב עבור שני סוגי הפשיעה, כאשר הנחנו קיום תלות מרחבית הן בתוחלת והן בשונות. נמצא, כי מיקום היחידה במרחב (ביחס לעיר תל-אביב המהווה מרכז פעילות חברתית-כלכלית ותרבותית של מדינת ישראל) משפיע על רמה של משתנה "שיעורי הפשיעה" משני הסוגים. מטריצת השונות המשותפות עבור נתוני עבירות רכוש עיקריות ברזולוציה של ישובים/מועצות אזוריות תלויה במרחקים בין היחידה הגיאוגרפית הנחקרת לבין היחידות השכנות.

(3) זהו מספר גורמים המשפיעים על שיעורי הפשיעה בישראל. ניתן לחלק את הגורמים האלה לשני סוגים.

א. השפעות הפעילות המשטרתית ושיפוטית - העלאת אחוז הגילויים בתיקים שנפתחו, העלאת אחוז ההרשעות בתיקים, הגברת הנוכחות המשטרתית בשטח (עליה באחוז שוטרי אג"מ) - פועלים בכיוון הורדת שיעורי הפשיעה, בעוצמות שונות.

ב. השפעות מאפיינים חברתיים-כלכליים ודמוגרפיים - זהו מספר גורמים המשפיעים על שיעורי הפשיעה, כגון השתייכות מגזרית (המגזר היהודי והמגזר הערבי), אחוז עולים, מידת הדתיות של האוכלוסייה היהודית, אחוז המשפחות המלאות, פערי הכנסות ועוד. כיווני השפעה של הגורמים האלה לא אחידים על פני השנים ועל פני סוגי פשיעה שונים. ניתוח כיווני ההשפעות אלה ועוצמתן הוצג בהרחבה בסעיף 4.2.

(4) בנתוני הפשיעה, קיימת תלות אוטורגרסיבית חזקה, יציבה ומובהקת (תלות בזמן). תלות זו קיימת גם במקרה בו שיעורי הפשיעה בעבר נלקחים עם פיגור של עד שלוש שנים. לכן, ניתוח שיעורי פשיעה אמור להתבסס על מתודולוגיה לניתוח סדרות עתיות. בנוסף, קיימת תלות הדדית בין שיעורי הפשיעה מסוגים שונים: נמצאה השפעה חזקה ומובהקת של שיעורי הפשיעה מסוג אלימות על שיעורי הפשיעה מסוג עבירות רכוש עיקריות, וההיפך.

(5) בבניית תחזיות לשיעורי פשיעה, לא ניתן להתעלם מהתלות המרחבית הקיימת בנתונים, והתעלמות ממנה גורמת לפגיעה משמעותית בדיוק התחזיות וביציבותן. המודל המרחבי Spatial Error Model, עם מטריצת השונות התלויה בפונקציה הופכית של המרחקים בין היחידות הגיאוגרפיות, נבחר לחיזוי נתוני עבירות רכוש עיקריות. עם זאת, למטרות הסבר תופעת הפשיעה, יש להעדיף מודל OLS המבוסס על משוואת אמידה (5), כאשר השוני בין מקדמי הרגרסיה המתקבלים ממודל מרחבי וממודל OLS אינו משמעותי. אפשרות לבחור את המשתנים שהשפעתם מובהקת ולהכליל רק אותם במודל הסופי (stepwise selection) מהווה יתרון חשוב של שיטת OLS. יצוין כי אלגוריתם זה אינו זמין עבור מודלים מורכבים יותר, ובפרט עבור המודלים המרחביים.

(6) השוני בין המגזר היהודי לבין המגזר הערבי, הן ברמות שיעורי הפשיעה והן בשונותם, מצריך התחשבות בהשתייכות מגזרית; לכן, בנתוני פשיעה מסוג אלימות, נאמדו שני מודלים OLS שונים לשני המגזרים. בנתוני עבירות רכוש עיקריים הדבר אינו אפשרי בגלל התלות המרחבית שזוהתה.

7) השיטה הבלתי ישירה נמצאה עדיפה על השיטות הישירות בבניית התחזית עבור יחידה גיאוגרפית מצרפית כמו שטח שיפוט של תחנת משטרה. עדיפות זו נובעת ממספר סיבות:

א. יכולת לפקח על מספר רב יותר של גורמים המשפיעים על שיעורי הפשיעה ברזולוציה גיאוגרפית של ישובים/מועצות (בגלל מספר רב יותר של תצפיות).

ב. בישובים עם מאפיינים חברתיים-כלכליים או/ו דמוגרפיים שונים הנכללים בשטח של אותה תחנת משטרה, גורמים המשפיעים על שיעורי הפשיעה עתידיים עשויים להיות שונים (למשל, ישובים יהודיים וערביים). ברזולוציה מרחבית של תחנות משטרה, התאמת מודל נפרד לכל אחד מהמגזרים אינה מתאפשרת בגלל מספר קטן של תצפיות. כתוצאה, חיזוי על סמך מודל רגרסיה המותאם לנתונים ברזולוציה של תחנת משטרה עלול לגרום לחוסר דיוק בתחזיות.

ג. השפעות הפעולות בכיוונים שונים אינן "נבלעות" בתוך הנתון המצרפי, מה שמעלה את רמת הדיוק של תחזיות המתקבלות לפי השיטה הבלתי ישירה.

8) לכל התחזיות מכל המודלים נאמדו סטיות תקן אשר שימשו כמדד ליציבות תחזיות בביצוע השוואות בין מודלים שונים. תחת שיטות החיזוי שנבחרו, סטיות התקן לשיעורי הפשיעה קטנים מהנתון החזוי לפחות בסדר גודל.

במסגרת המחקר הנוכחי, בוצע ניתוח סדרות עתיות רבעוניות של אירועי פשיעה. בכל הסדרות שנבדקו במסגרת המחקר הנוכחי, נמצא השפעה מובהקת של החגים העבריים וימי הפעילות הנהוגים בישראל וכן השפעה עונתית. לכן, כדאי לנכות השפעות אלו על מנת לאמוד את המגמה הבסיסית בסדרות של אירועי פשיעה מסוגים שונים. מומלץ להרחיב ניתוח זה לשאר תחנות המשטרה בארץ תוך אמידת הגורמים העונתיים וגורמי החג וימי הפעילות עבור כל אחת מתחנות המשטרה.

נבחנו שיטות חיזוי של סדרות עתיות תוך פירוקן למרכיבים; נמצא, כי שיטה המבוססת על:

א. חיזוי של סדרה מותאמת השפעות החגים וימי הפעילות,

ב. חיזוי של גורמי החג וימי הפעילות במשק (שיטה TS2),

עדיפה, במונחי דיוק תחזיות, על פני השיטה המתחרה המבוססת על חיזוי של כל אחד ממרכיבי הסדרה בנפרד.

תחזיות בתדירות רבעונית יכולות לשמש את המשטרה בתכנון פעילותה לטווח קצר. כמו כן, תחזיות לגורמים עונתיים ולגורמי חג וימי פעילות יכולות לספק למקבלי החלטות במשטרה מידע חשוב המסייע לתכנן מראש תגובת המשטרה להתגברות הפשע בתקופות "חמות". במובן מסוים, ניתוח סדרות עתיות מאפשר לזהות "תקופות חמות", במקום "נקודות חמות" בניתוח רב-משתני של Neural Data Mining ו- Neural Networks.

7. רשימה ביבליוגרפית

- Anselin, L., (1988) *Spatial Econometrics: Methods and Models*, *Kluwer Academic*, Dordrecht
- Anselin, L., Cohen, J., Cook, D., Gorr, W., and Tita, G., (2000) *Spatial analysis of crime*, *Measurement and Analysis of Crime and Justice*, vol. 4, pp 213-261
- Beenstock, M., Ben Zeev, N., and Felsenstein, D., (2009) *Nonparametric estimation of the spatial connectivity matrix by the method of moments using spatial panel data*, Working paper, Hebrew University of Jerusalem
- Box, G.E.P and Jenkins, G.M., (1976) *Time series analysis: forecasting and control* (2nd ed.), San-Francisco, CA, USA: Holden-Day
- Brown, D.E., and Smith, M., (2005) *Terrorist threat prediction in urban environments*, Knowledge fusion, Research workshop, Nov.-Dec. 2005, Annapolis, Maryland, USA
- Cohen, J., Gorr, W.L., Olligschlaeger, A.M., (2005) *Leading indicators and spatial interactions: a crime forecasting model for proactive police deployment*, *Geographical Analysis*
- Cohen, J., (2006) *Development of crime forecasting and mapping systems for use by police*, Final report, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, USA
- Corcoran, J.J., Wilson, I.D., and Ware, J.A., *Predicting the geo-temporal variations of crime and disorder*, *International Journal of Forecasting*, Vol. 19, pp 623-634
- Dorman, C.F., McPherson, J.M., Araujo, M.B., Bivand, R., Bolliger, J., Carl, G., Davies, R.G., Hirzel, A., Jetz, W., Kissking, D., Kuhn, I., Ohlemuller, R., Peres-Neto, P.R., Reineking, B., Schroder, B., Schurr, F.M., Wilson, R., (2007) *Methods to account for spatial autocorrelation in the analysis of species distribution data: a review*, *Ecography*, Vol. 30, pp 609-628
- Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C., (2000) *New capabilities and methods of the X-12-ARIMA seasonal adjustment problem*, U.S. Census Bureau, Washington DC, USA
- Gorr, W., Olligschlaeger, A., and Thompson, Y., (2000) *Assessment of crime forecasting accuracy for deployment of police*, Working paper, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, USA
- Gorr, W., and Harries, R., (2003) *Introduction to crime forecasting*, *International Journal of Forecasting*, Vol. 19, pp 551-555
- Hannan, E.G., and Quinn, B.G., (1979) *The determination of the order of autoregression*, *Journal of the Royal Statistic Society, Series: B*, 41, pp 190-195

- Liu, H., and Brown, D.E., (2003) *Criminal incident prediction using a point-pattern based density model*, International Journal of Forecasting, Vol. 19, pp 603-622
- Mitchell, M.B., Brown, D.E., and Conklin, J.H., (2007) *A crime forecasting tool for the web-based crime analysis toolkit*, System and Information Engineering Design Symposium, Apr. 2007
- Morenoff, J.D., Sampson, R.J., (1997) *Violent crime and the spatial dynamics of neighborhood transition: Chicago, 1970-1990*, Social Forces, Vol. 76, pp 31-64
- Oatley, G.C., Ewart, B.W., and Zeleznikow, J., (2006) *Decision support systems for police: lessons from the application of data mining techniques to 'soft' forensic evidence*, Artificial Intelligence and Law, Vol. 14(1-2) pp 35-100
- Olligschlaeger, A., (1997) *Spatial analysis of crime using GIS - based data: weighted spatial adaptive filtering and chaotic cellular forecasting with applications to street level drug markets*, PhD dissertation, H. John Heinz III School of Public Policy and Management, Carnegie Mellon University, USA
- Schubert, J.L., (2004) *Identifying crime hot spots in space and time: an analysis of Dallas police department reports of criminal activity*, GIS Workshop, University of Texas at Dallas, USA
- Sun, H., Tu, Y., and Yu, S.M., (2005) *A spatio-temporal autoregressive model for multi-unit residential market analysis*, The Journal of Real Estate Finance and Economics, Vol. 31-2, pp 155-187
- X-12-ARIMA Reference Manual*, Version 03, U.S. Census Bureau, Washington DC, USA, 2007

אפיון רשויות מקומיות וסיווג לפי רמה חברתית-כלכלית של האוכלוסייה – 2003, הלמ"ס 2006 (פ"מ 1281)

גורמים עונתיים וגורמי התאמה מראש לשנת 2010, מגמות ל- 2006 - 2010, פרסום הלשכה המרכזית לסטטיסטיקה, ירושלים

בורק, לואיזה, וגובמן, יורי (2007), ניכוי עונתיים באמצעות X-12-ARIMA, נייר עבודה מס' 30, הלשכה המרכזית לסטטיסטיקה, ירושלים

גובמן, יורי, (2009) שימוש בשיטות סטוכסטיות לביון וחיוץ של סדרות עתיות, נייר עבודה מס' 44, הלשכה המרכזית לסטטיסטיקה, ירושלים

ציבל, נטליה (2009), מדד פריפריאליות של רשויות מקומיות בישראל: שילוב של מדד נגישות פוטנציאלית עם קרבה למחוז תל אביב, נייר עבודה מס' 45, הלשכה המרכזית לסטטיסטיקה, ירושלים

פורטנוי, חיים (2007) אפיון רמת הדתיות באוכלוסייה היהודית על פי זיקה למוסדות חינוך, נייר טכני מס' 19, הלשכה המרכזית לסטטיסטיקה, ירושלים

רגב, בשורה (2008) היבטים כלכליים וכמותיים למדיניות אכיפת החוק של עבירות צווארון לבן בישראל, עבודה לקבלת תואר דוקטור, הפקולטה למשפטים, האוניברסיטה העברית בירושלים

רשימת נספחים

נספח A – תיאור משתנים

נספח B – מודלים להסבר שיעורי פשיעה

נספח C – מודלים לחיזוי שיעורי פשיעה

נספח D – טיב תחזיות על סמך ניתוח סדרות עתיות רבעוניות

נספח A

לוח A1: רשימת משתנים ברזולוציות של ישוב/מועצה ותחנת משטרה

זמינות	תיאור	משתנה	סוג
2001-2009, רבעוני ושנתי	שיעור תיקי האלימות/עבירות רכוש עיקריות ל-1,000 תושבים הגרים באזור שבו התרחשו אירועי האלימות/עבירות רכוש עיקריות. הגדרת התיק נעשית על פי משתנה משטרי "סמל מקוצר" בנספח A2	משתנה מוסבר: שיעורי הפשיעה	משתנים משטריים
2001 - 2008, שנתי	פרופרציית התיקים המקושרים לנתוני עבריינים, קרי התיקים אשר לפחות עבריון אחד קושר לאירוע. ישנם תיקים שבהם אין "מובילים" (חשודים) – ואלו למעשה תיקים ללא "גילוי"	פרופרציית גילויים בתיקים שנפתחו	
	שיעור העבריינים שקושרו לתיקים ונידונו בעבירות אלימות/עבירות רכוש עיקריות ל-1,000 תושבים הגרים במקום המגורים של העבריון	שיעור עבריינים	
	שיעור השוטרים לפי התקן בתחנת המשטרה (כפי שנקבע מדי שנה) בכל המקצועות ל-1,000 תושבים	שיעור שוטרים	
2006 - 2008, שנתי	פרופרציה של שוטרי אג"מ (מבצעים) שבתקן תחנת המשטרה מבין כלל השוטרים בתחנה	שוטרי אג"מ	
	פרופרציה של שוטרי אח"ק (חקירות) שבתקן תחנת המשטרה מבין כלל השוטרים בתחנה	שוטרי אח"ק	
	פרופרציה של שוטרי אמ"ן (מודיעין) שבתקן תחנת המשטרה מבין כלל השוטרים בתחנה	שוטרי אמ"ן	
2003 - 2008, שנתי	מקור: מרשם האוכלוסין (לפי סיווג קבוצת אוכלוסייה: 1. יהודים ואחרים 2. ערבים, ראה הסבר בפרק 3). בכל ישוב חושבה פרופרציית היהודים ואחרים באוכלוסיית הישוב	קבוצת אוכלוסייה: יהודים ואחרים	משתנים דמוגרפיים
2003 - 2008, שנתי	מקור: מרשם האוכלוסין. כל פרט שויד לקבוצת גיל לפי תאריך לידה: 0-24, 25-54, 55+. חושבה פרופרציה של השייכים לכל קבוצת גיל ביחס לכלל אוכלוסיית היחידה הגיאוגרפית הרלוונטית	מבנה גילים	
2003 - 2008, שנתי	מקור: מרשם האוכלוסין. מספר התושבים המוגדרים כ"תלויים", שאינם בגילי העבודה (צעירים בגילים 0-17 ומבוגרים בגילים 65+), לחלק למספר התושבים הנמצאים בגילי העבודה (18-64)	יחס התלות	
2003 - 2008, שנתי	מקור: קובץ רמת דתיות שפותח בלמ"ס (פורטנוי, 2007). חושבה פרופרציית האוכלוסייה החרדית, הדתית והלא דתית.	רמת דתיות עבור אוכלוסייה יהודית	
2003 - 2008, שנתי	מקור: מרשם האוכלוסין. חושב על סמך המידע על ארץ הלידה ושנת העלייה. עולי בריה"מ לשעבר – משנת 1990; עולי אתיופיה נלקחו ללא התייחסות לשנת העלייה. בשני המקרים, בחישוב נכללים ילדי העולים.	פרופרציית עולי בריה"מ לשעבר ועולי אתיופיה	
2003 - 2008, שנתי	מקור: מרשם האוכלוסין. על סמך מספר משתני מרשם האוכלוסין, זוהו משפחות מנהליות גרעיניות. משפחה מלאה מוגדרת כזוג הורים עם ילדיהם עד גיל 18.	פרופרציה של משפחות מנהליות מלאות	
2003 - 2008, שנתי	מקור: קובצי חינוך - תלמידים בדרג חינוך על-יסודי הלומדים בבתי ספר שבפיקוח משרד התמ"ת ולא בבתי הספר שבפיקוח משרד החינוך. חושבה פרופרציית תלמידי תמ"ת בקרב האוכלוסייה בגילי 15-20.	פרופרציה של בני הנוער הלומדים במסגרת תמ"ת	משתני חינוך
2003 - 2008, שנתי	מקור: קובצי חינוך – "קידום נוער". כל בני הנוער שפונים ל"קידום נוער" ומקבלים מענה כלשהו במסגרת זו. חושבה פרופרציית בני נוער השייכים לקידום בקרב האוכלוסייה בגילי 15-21.	פרופרציה של בני הנוער במסגרת "קידום נוער"	
2003 - 2008, שנתי	המדד חברתי-כלכלי משקלל מספר רב של מאפיינים חברתיים כלכליים ודמוגרפיים של האוכלוסייה בכל אחת מרשויות מקומיות בארץ. רמה חברתית-כלכלית נעה בין 1 (הרשויות ה"חלשות" ביותר) לבין 10 (הרשויות ה"חזקות" ביותר). לפירוט ראה פרסום למ"ס משנת 2006.	מדד חברתי-כלכלי של רשות מקומית, 2003	משתנים כלכליים
2003 - 2008, שנתי	מקור: רשות המסים. בסיס נתונים עובד/מעבידה המכיל את נתוני הכנסת שכירים ועצמאים בישראל. מתוך כלל התושבים בגילי 25-54, חושבה פרופרציית שכירים ופרופרציית עצמאים.	פרופרציית השכירים/עצמאיים	
2003 - 2008, שנתי	מקור: בסיס נתונים עובד/מעביד. מתוך כלל הצעירים בגילי 13-17, חושבה פרופרציית הנמצאים בבסיס הנתונים עובד/מעביד והיה להם לפחות שלושה חודשי עבודה בשנה.	פרופרציית קטינים המועסקים	
2003 - 2008, שנתי	מקור: בסיס נתונים עובד/מעביד. בהסתמך על כלל בעלי הכנסות משכר חושב חציון הכנסה שנתית משכר (ברוטו).	הכנסה שנתית חציונית משכר	
2003 - 2008, שנתי	מקור: בסיס נתונים עובד/מעביד. כל בעלי הכנסה משכר חולקו לחמישונים, כך שבחמישון 1 נמצאים 20% מהפרטים ששכרם הוא הנמוך ביותר ובחמישון 5 נמצאים 20% מהפרטים ששכרם הוא הגבוה ביותר.	פרופרציית שכירים בחמישוני הכנסה 1 ו-5	
2003 - 2008, שנתי	חושב השכר הממוצע של בעלי הכנסה משכר הנמצאים בחמישון הארצי הגבוה (5), והוא חולק בשכר הממוצע של המשתכרים הנמצאים בחמישון הארצי הנמוך (1), עבור אותה יחידה גיאוגרפית. ככל שפערים גדולים יותר כך המדד יהיה גבוה יותר, וככל שפערים קטנים יותר כך המדד יהיה נמוך יותר.	מדד לפערי הכנסה משכר בין חמישונים 5 ו-1	
2001 - 2009, שנתי	מרחק אוירי בין מרכז הישוב/מועצה לבין מרכז עיר תל-אביב; מוסבר בפרק 3.	מרחק מתל-אביב	משתנים מרחביים
2003 - 2009, שנתי	שיעור הפשיעה ביחידות גיאוגרפיות שכנות (ראה הגדרה בפרק 3), משוקלל בגודל אוכלוסייה של היחידות הנ"ל.	שיעורי פשיעה אצל השכנים	

נספח A2 – הרכב "סל עבירות אלימות" ו"סל עבירות רכוש עיקריים

תיקי אלימות ותיקי עבירות רכוש עיקריות סווגו על פי הסיווגים הסטטיסטיים, בהתאם להגדרות של משטרת ישראל.

תיקי אלימות – תיקים שבהם יש לפחות סעיף עבירה אחד השייך לסיווגים הסטטיסטיים הבאים:

202, 203 – קטטות והפרעות במוסדות וברחובות (כולל 17 סעיפי עבירות, לדוגמה: קטטה, העלבת אדם

במקום ציבורי, התנהגות פרועה תוך שכרות, קבוץ נדבות ועוד)

217 – תקיפת זוטר בנסיבות חמורות (כולל 3 סעיפי עבירות)

401 – תקיפה וחבלה גופנית חמורה (כולל 23 סעיפי עבירות, לדוגמה: הנחת חומר נפץ, השקאת רעל, חבלה

ע"י שניים או יותר, ניסיון השחתת גוף ועוד)

403 – תקיפת עובד ציבור (כולל 3 סעיפי עבירות)

404 – תקיפה למעט עובד ציבור (כולל 42 סעיפי עבירות, לדוגמה: התעללות פיזית או נפשית בקטין, שידול

קטין להמיר את דתו, פציעת בן זוג ועוד)

701 – שוד בנסיבות חמורות (כולל 4 סעיפי עבירות, לדוגמה: שוד מזוין, ניסיון לשוד עם נשק ועוד)

702 – שוד (כולל 4 סעיפי עבירות, כולל: שוד, תקיפה לשם שוד)

704 – תקיפה לשם גניבה (כולל 2 סעיפי עבירות)

722 – חטיפת תיקים (כולל 4 סעיפי עבירות).

תיקי רכוש עיקריים - תיקים שבהם יש לפחות סעיף עבירה אחד השייך לסיווגים הסטטיסטיים הבאים:

706 – התפרצות לבתי עסק ולמוסדות (כולל 6 סעיפי עבירות, לדוגמה: כניסה והתפרצות בנסיבות

מחמירות, פריצה לבתי עסק ומוסדות וגניבה ועוד)

707 – התפרצות לבתי דירה (כולל 5 סעיפי עבירות, לדוגמה: כניסה לדירה בכוונה לגנוב, התפרצות

למגורים/ תפילה לבצע עבירה ועוד)

711 – גניבת רכב (כולל 8 סעיפי עבירות, לדוגמה: שימוש ברכב בלי רשות, גונב, מזייף או משתמש במרמה

במסמכי רכב, גניבת כלי שיט או כלי טיס ועוד)

נספח B

מודלים להסבר שיעורי פשיעה

לוח B1 – מודל להסבר שיעורי פשיעה מסוג אלימות – רזולוציה של ישוב/מועצה

מודל רגרסיה לישובים/מועצות - שיעורי פשיעה מסוג אלימות, מגזר יהודי					
משתנה		מקדם	ממוצע	גמישות	רמת המובהקות
משתני רמה בנשה שעברה (t-1)	שיעור פשיעה מסוג אלימות (לוג טבעי)	0.7658	1.8604	0.7658	<0.0001
	שיעור פשיעה מסוג פע"ר (לוג טבעי)	0.0928	2.3223	0.0928	<0.0001
	פרופורציית עולי בריה"מ לשעבר, משנת 1990	0.4398	0.0945	0.0416	0.0386
	פרופורציית בני נוער הלומדים במסגרות תמ"ת	3.8080	0.0136	0.0518	0.0187
שינוי שחל בערכי המשתנה משנה (t-2) עד לשנה (t-1)	שינוי בפרופורציית גילוי	-0.0628	0.0091	-0.0006	0.0133
	שינוי בפערי הכנסות משכר	-0.0307	-2.3476	0.0720	0.0177
	שינוי בפרופורציית השייכים לחמישון התחתון של הכנסות	2.7056	-0.0030	-0.0081	0.0378
משתנה מרחבי	מרחק מתל-אביב	0.0008	62.8883	0.0503	0.0334
מדד טיב המודל		MSE	0.0496		

מודל רגרסיה לישובים/מועצות - שיעורי פשיעה מסוג אלימות, מגזר ערבי					
משתנה		מקדם	ממוצע	גמישות	רמת המובהקות
משתני רמה בנשה שעברה (t-1)	שיעור פשיעה מסוג אלימות (לוג טבעי)	0.8820	1.7916	0.8820	<0.0001
	פערי הכנסות משכר	0.0411	14.5003	0.5960	0.0036
שינוי שחל בערכי המשתנה משנה (t-2) עד לשנה (t-1)	שינוי בפרופורציית השייכים לחמישון התחתון של הכנסות	5.9685	0.0114	0.0680	0.0079
	שינוי בפרופורציית השייכים לחמישון העליון של הכנסות	6.9836	-0.0036	-0.0251	0.0710
משתנה מרחבי	מרחק מתל-אביב	-0.0058	89.4406	-0.5188	0.0012
מדד טיב המודל		MSE	0.0924		

לוח B2 – מודל OLS2 להסבר שיעורי פשיעה מסוג עבירות רכוש עיקריות – רזולוציה של ישוב/מועצה

מודל רגרסיה לישובים/מועצות - שיעורי פשיעה מסוג עבירות רכוש עיקריות, מגזר יהודי					
משתנה		מקדם	ממוצע	גמישות	רמת המובהקות
משתני רמה בנשה שעברה (t-1)	שיעור פשיעה מסוג פע"ר (לוג טבעי)	0.7763	2.3223	0.7763	<.0001
	שיעור פשיעה מסוג אלימות (לוג טבעי)	0.1193	1.8604	0.1193	0.0100
	פרופורציית אוכלוסיה דתית (לא חרדית)	0.2421	0.1740	0.0421	0.0573
	פערי נכסות משכר	0.0184	17.3578	0.3201	0.0008
	פרופורציית גילויים בתיקים שנפתחו	-1.0601	0.1295	-0.1373	0.0001
שינוי שחל בערכי המשתנה משנה (t-2) עד לשנה (t-1)	שינוי ביחס תלות	2.5399	-0.0022	-0.0056	0.0946
משתנה מרחבי	מרחק מתל-אביב	-0.0013	62.8883	-0.0792	0.0205
מדד טיב המודל		MSE	0.0919		

מודל רגרסיה לישובים/מועצות - שיעורי פשיעה מסוג עבירות רכוש עיקריות, מגזר ערבי					
משתנה		מקדם	ממוצע	גמישות	רמת המובהקות
משתני רמה בנשה שעברה (t-1)	שיעור פשיעה מסוג פע"ר (לוג טבעי)	0.5880	0.3425	0.5880	<.0001
	שיעור פשיעה מסוג אלימות (לוג טבעי)	0.5279	1.7916	0.5279	<.0001
	פרופורציית עצמאים	-8.9935	0.0615	-0.5531	0.0012
שינוי שחל בערכי המשתנה משנה (t-2) עד לשנה (t-1)	שינוי בפערי הכנסות משכר	0.1234	-1.5784	-0.1948	0.0369
	שינוי ביחס תלות	12.6281	-0.0157	-0.1983	0.0300
מדד טיב המודל		MSE	0.2704		

לוח B3 – מודל להסבר שיעורי פשיעה מסוג אלימות – רזולוציה של תחנת משטרה

מודל רגרסיה לתחנות משטרה - שיעורי פשיעה מסוג אלימות					
משתנה		מקדם	ממוצע	גמישות	רמת המובהקות
משתני רמה בנשה שעברה (t-1)	שיעור פשיעה מסוג אלימות (לוג טבעי)	0.8973	2.1228	0.8973	<0.0001
	אחוז שוטרי אג"מ מתוך כלל שוטרי התחנה (כפול 100)	-0.0039	66.8626	-0.2574	0.0505
	פרופורציית שכירים	0.4722	0.6527	0.3082	0.0029
	פרופורציית עצמאים	-2.7356	0.0583	-0.1595	0.0145
	פרופורציית משפחות מלאות (שני הורים נשואים)	0.6041	0.5182	0.3130	0.0014
מדד טיב המודל	MSE	0.0084			

לוח B4 – מודל להסבר שיעורי פשיעה מסוג עבירות רכוש עיקריות – רזולוציה של תחנת משטרה

מודל רגרסיה לתחנות משטרה - שיעורי פשיעה מסוג עבירות רכוש עיקריות					
משתנה		מקדם	ממוצע	גמישות	רמת המובהקות
משתני רמה בנשה שעברה (t-1)	אחוז שוטרי אג"מ מתוך כלל שוטרי התחנה	-0.0073	66.8626	-0.4868	0.0137
	פערי נכסות משכר	0.1258	16.0664	2.0208	<0.0001
	פרופורציית משפחות מלאות (שני הורים נשואים)	-2.6801	0.5182	-1.3888	0.0003
משתנה מרחבי	שיעור משוקלל של עבירות רכוש (פע"ר) בתחנות הסמוכות	0.8641	1.9666	0.8641	<0.0001
מדד טיב המודל	MSE	0.2134			

נספח C בחירת מודל לחיזוי ותחזיות פשיעה

לוח C1 – בחירת מודל לחיזוי שיעורי פשיעה מסוג אלימות, ברמת ישוב/מועצה

מדדי טיב תחזיות - אלימות, רמת ישוב/מועצה					
תקופת תחזית	שנת תחזית	MAPE OLS1	MAPE OLS2	Std OLS1	Std OLS2
לשנה אחת	2006	25.8819	24.8128	0.3018	0.3947
	2007	20.9006	21.0715	0.3407	0.3616
	2008	25.6092	22.2764	0.4599	0.4989
	2009	22.2301	21.1166	0.3227	0.4705
לשנתיים	2007	29.2265	27.6607	0.2991	0.4598
	2008	23.3696	24.8594	0.2603	0.4260
	2009	27.0952	25.5463	0.5012	0.6685
	2010	.	.	0.4587	0.6064
לשלוש שנים	2008	26.3773	26.2018	0.3103	0.3736
	2009	22.6782	23.7045	0.2884	0.4413
	2010	.	.	0.2605	0.3038
	2011	.	.	0.2523	0.2712
	2012	.	.	0.2339	0.2565
Average		24.8187	24.1389	0.3300	0.4256

לוח C2 – בחירת מודל לחיזוי שיעורי פשיעה מסוג עבירות רכוש עיקריות, ברמת ישוב/מועצה

מדדי טיב תחזיות - עבירות רכוש עיקריות, רמת ישובים/מועצות						
תקופת תחזית	שנת תחזית	MAPE OLS1	MAPE OLS2	MAPE SER1	MAPE SER2	MAPE SGLS
שנה אחת	2006	33.2474	33.6118	33.6714	33.6637	34.0150
	2007	42.0902	39.3969	35.5973	35.4978	35.3221
	2008	35.4117	39.1768	36.3222	36.3726	36.6109
	2009	38.5451	37.7955	32.5480	32.7471	33.0626
שנתיים	2007	37.8836	38.1678	37.0830	36.7939	37.0837
	2008	45.7151	38.6605	38.1702	38.0422	38.1486
	2009	36.9191	38.5494	37.7068	37.8535	37.8783
	2010
שלוש שנים	2008	41.4026	41.1648	40.2597	40.3471	40.7030
	2009	34.6180	35.6918	32.8493	32.9673	33.0862
	2010
	2011
	2012
Average MAPE		38.4259	38.0239	36.0231	36.0317	36.2123

תקופת תחזית	שנת תחזית	Std OLS1	Std OLS2	Std SER1	Std SER2	Std SGLS
שנה אחת	2006	0.9344	0.6332	0.7067	0.7085	0.6817
	2007	0.5402	0.6691	0.6167	0.6038	0.6001
	2008	0.3792	0.6084	0.3909	0.4040	0.3906
	2009	0.5538	0.5041	0.4085	0.3956	0.3848
שנתיים	2007	1.0168	0.8462	0.6415	0.6853	0.6773
	2008	0.5032	0.5051	0.5405	0.5247	0.5096
	2009	0.3669	0.6940	0.3663	0.3383	0.3301
	2010	0.3092	0.3678	0.3918	0.3796	0.3755
שלוש שנים	2008	0.6450	0.5736	0.6236	0.6125	0.5934
	2009	0.6416	0.5576	0.4593	0.4457	0.4374
	2010	0.2946	0.3645	0.3259	0.3257	0.3194
	2011	0.2618	0.3068	0.2907	0.2905	0.2852
	2012	0.2293	0.2828	0.2539	0.2537	0.2501
Average Std		0.5135	0.5318	0.4628	0.4591	0.4489

לוח C3 – בחירת מודל לחיזוי שיעורי פשיעה מסוג אלימות, ברמת תחנת משטרה

מדדי טיב תחזית - אלימות, תחנת משטרה							
תקופת תחזית	שנת תחזית	MAPE OLS1 ישיר	MAPE OLS1 בלתי ישיר	MAPE OLS2 בלתי ישיר	Std OLS1 ישיר	Std OLS1 בלתי ישיר	Std OLS2 בלתי ישיר
שנה אחת	2006	21.0508	18.4716	17.6763	0.5106	0.2214	0.2683
	2007	9.3178	14.8663	15.5081	0.1211	0.2460	0.2568
	2008	10.0925	16.0222	16.1541	0.1925	0.3562	0.4153
	2009	12.7576	12.1118	14.4610	0.2299	0.2572	0.3420
שנתיים	2007	6.6507	16.9209	16.7036	0.1998	0.1983	0.3111
	2008	8.4804	15.6544	15.6292	0.1538	0.1985	0.2992
	2009	13.4165	16.5509	14.9237	0.2197	0.3831	0.4769
	2010				0.3949	0.3509	0.4207
שלוש שנים	2008	10.9449	17.6828	16.9268	0.2147	0.2213	0.2586
	2009	34.0682	12.8389	14.2460	0.2089	0.2313	0.3428
	2010				0.3176	0.2205	0.2298
	2011				0.3082	0.2186	0.2220
	2012				0.3016	0.2054	0.2076
Average		14.0866	15.6800	15.8032	0.2595	0.2545	0.3116

לוח C4 – בחירת מודל לחיזוי שיעורי פשיעה מסוג עבירות רכוש עיקריות, ברמת תחנת משטרה

מדדי טיב תחזית - עבירות רכוש עיקריות, רמת תחנת משטרה								
תקופת תחזית	שנת תחזית	MAPE OLS1 ישיר	MAPE SER ישיר	MAPE OLS1 בלתי ישיר	MAPE OLS2 בלתי ישיר	MAPE SER1 בלתי ישיר	MAPE SER2 בלתי ישיר	MAPE SGLS בלתי ישיר
שנה אחת	2006	15.5367	15.1231	25.2545	27.2708	25.8590	25.8453	26.1140
	2007	15.5843	15.7100	22.4689	24.3004	21.4126	21.3645	21.5653
	2008	35.6922	35.5543	20.1704	26.9008	21.4144	21.3597	21.2693
	2009	30.8815	32.8378	26.4040	25.1144	21.8950	21.9339	22.0613
שנתיים	2007	19.2411	18.8929	30.3219	30.3090	29.1368	28.5847	29.2089
	2008	20.1210	20.7825	26.9005	24.9547	22.6640	22.6638	22.7510
	2009	31.6000	34.5191	22.1240	24.4966	21.9228	22.2071	22.0022
	2010							
שלוש שנים	2008	24.4350	24.3612	29.6515	29.9185	29.7891	30.0331	30.2333
	2009	43.1120	44.6413	22.4884	25.4207	20.6541	20.8247	20.6776
	2010							
	2011							
	2012							
Average MAPE		26.2449	26.9358	25.0871	26.5207	23.8609	23.8685	23.9870

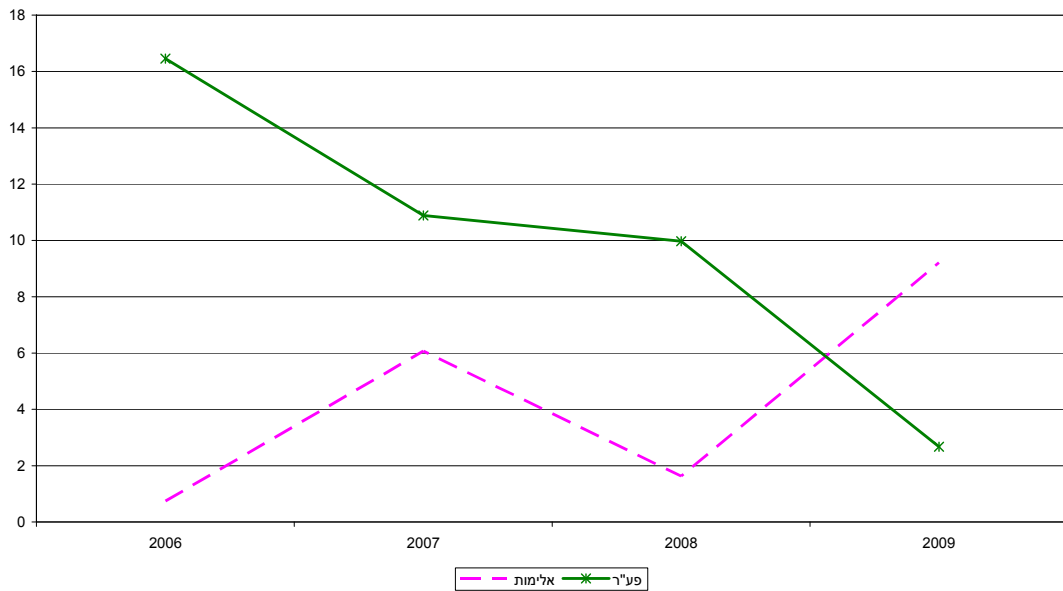
תקופת תחזית	שנת תחזית	Std OLS1 ישיר	Std SER ישיר	Std OLS1 בלתי ישיר	Std OLS2 בלתי ישיר	Std SER1 בלתי ישיר	Std SER2 בלתי ישיר	Std SGLS בלתי ישיר
שנה אחת	2006	0.4394	12.0624	0.4869	0.3763	0.4152	0.4163	0.4019
	2007	0.3852	10.4723	0.2489	0.3481	0.3584	0.3508	0.3509
	2008	0.7820	5.6077	0.2102	0.3666	0.2468	0.2549	0.2485
	2009	0.6989	7.2041	0.3577	0.3707	0.2639	0.2559	0.2497
שנתיים	2007	0.6726	10.9644	0.5389	0.4507	0.3852	0.4150	0.4114
	2008	0.3484	9.1248	0.2621	0.2715	0.3126	0.3037	0.2968
	2009	0.4491	4.9304	0.2262	0.4126	0.2403	0.2208	0.2166
	2010	0.5917	6.2487	0.1905	0.1952	0.2681	0.2603	0.2593
שלוש שנים	2008	0.6352	9.5445	0.4027	0.3678	0.3839	0.3760	0.3657
	2009	0.2808	8.0151	0.3280	0.3182	0.2706	0.2618	0.2565
	2010	0.4659	4.7237	0.1989	0.2601	0.2205	0.2203	0.2176
	2011	0.4579	4.8581	0.1730	0.2012	0.1953	0.1952	0.1932
	2012	0.4376	4.6765	0.1563	0.1844	0.1763	0.1762	0.1748
Average Std		0.5111	7.5717	0.2908	0.3172	0.2875	0.2852	0.2802

איורים C1 ו-C2 – נתונים בפועל, תחזיות נאמדות ומדד MAPE לטיב תחזיות, ברמה ארצית

שיעורי פשיעה ברמה ארצית: תחזיות (לפי השיטה הבלתי ישירה) ושיעורי פשיעה בפועל



טעויות יחסיות של התחזיות ברמה ארצית, באחוזים



נספח D – אמידת תחזיות על סמך ניתוח סדרות עתיות, לתחנות נבחרות

לוח D1 – מדדי טיב תחזיות רבעוניות

מדדי טיב תחזיות רבעוניות			
סוג פשיעה	מדדי MAPE, ממוצע על פני 6 תחנות נבחרות		
	שיטה 1	שיטה 2	שיטה בלתי ישירה ממודל רגרסיה
אלימות	7.67	6.60	6.34
עבירות רכוש	25.42	24.22	11.10

Abstract

Spatial allocation of limited police force requires extensive information and planning to optimize its effectiveness in terms of crime rates reduction. In this paper we compare a number of parametric temporal–spatial models for crime rates forecasting, where the obtained forecasts should meet the police requirements in terms of accuracy, spatial resolution and timeliness according to a planning period.

The crime data should be analyzed within a spatial dependence model. Spatial dependence may affect estimation procedure, but its significance and influence on crime forecasts accuracy for sufficiently large geographic areas is unclear a-priori. Also, crime activity is time dependent, and one cannot ignore seasonal fluctuations and holiday and trading day influence in the forecasting procedure. The forecasts should internalize existing police data, economic, social and demographic variables.

Additional challenge in the analysis of crime data is endogenous relation between crime and police activity: more police resources are allocated to high crime areas thereby creating a positive correlation between the levels of crime and police resources across geographical areas.

In the current study, we analyze two major types of crime activities: violence and offences against property. Forecasts are constructed for two geographical levels: (1) municipalities, local and regional councils (high resolution); (2) police stations (low resolution, which includes one or more localities or/and regional councils, excluding Tel Aviv and Jerusalem). We use unique linked database which contains police data at the individual crime incident level, for 2003-2009, socio-economic and demographic variables and GIS data on spatial features of geographic units.

Leading indicators method is used for medium and long term forecasting (up to three years). For geographical resolution of police station, indirect approach is applied, where the forecast for aggregate unit (police station) is obtained as a weighted sum of forecasts over localities included in the station jurisdiction. The proposed forecasting methods are compared, in terms of accuracy, stability during the research period and robustness, and the appropriate method is chosen for each geographical resolution in for each type of crime activity.

For quarterly crime forecasting we propose time series based approach, which includes seasonal adjustment procedure and estimation of Jewish holidays and trading day effects.

Key words: crime forecasting, spatial models, leading indicators, seasonal adjustment

**Published by the Central Bureau of Statistics, 66 Kanfe Nesharim St.,
Corner Bachi St., P.O.B 34525, Jerusalem 91342, Israel
Tel. 972-2-6592666; Fax: 972-2-6521340
Internet Site: www.cbs.gov.il
E-Mail: info@cbs.gov.il**

The Central Bureau of Statistics (CBS) encourages research based on CBS data. Publications of this research are not official publications of the CBS, and they have not undergone the review accorded official CBS publications. The opinions and conclusions expressed in these publications, including this one, are those of the authors and do not necessarily represent those of the CBS. Permission for republication in whole or part must be obtained from the authors.

WORKING PAPER SERIES

NO.55

Estimation of temporal - spatial forecasts for crime rates in Israel

Yury Gubman* Nir Fogel* Dmitri Romanov*
Besora Regev** Shai Amram**

כסלו תשע"א, נובמבר 2010 November