

מדינת ישראל
משרד ראש הממשלה



מודל רגרסיה הדונית לחישוב שינויים במחירי דירות בעיר חיפה דו"ח מתודולוגי

עבודת מחקר במסגרת הפרויקט "מדד מחירי דירות בבעלות פרטית"
של גף מחירים לצרכן, אגף מיקרו כלכלה

אלון שפירא
תחום ניתוח סטטיסטי
אגף מתודולוגיה סטטיסטית

לואיזה בורק
תחום ניתוח סטטיסטי
אגף מתודולוגיה סטטיסטית

ירושלים, יולי 2005

הקדמה

דו"ח מתודולוגי זה מתאר את המחקר שנעשה במסגרת הפרויקט "מדד מחירי דירות בבעלות פרטית" של אגף מיקרו כלכלה. מטרת המחקר היא בניית מודל לחישוב שינויים במחירי דירות עבור העיר חיפה. שלבי המחקר כללו: בניית תוכנית העבודה, טיפול בקובץ נתונים, עריכת בדיקות מקדימות, פיתוח המודל וניתוח ממצאים ראשוניים. נציין, כי עבודה זו נעשתה בתאום, סיוע ושיתוף פעולה עם מר דורון סייג מגף מחירים לצרכן.

תוכן עניינים

4	מבוא	1
4	בדיקת אמינות ושלמות הנתונים	2
4	ניפוי רשומות בקובץ	2.1
5	הגדרת משתנים	2.2
8	זקיפת ערכי נתונים חסרים	2.3
8	ערכים חריגים	2.4
11	בניית מודל רגרסיה הדונית ללוג-מחיר הדירה	3
11	המודל ההדוני לחישוב השינויים במחירי דירות – המודל התקופתי	3.1
12	חישוב שינויים במחירי דירות בבעלות פרטית	3.2
14	בחירת משתנים מסבירים במודל	3.3
15	בחירת משתנים מסבירים למודל	3.4
17	ממצאים	4
17	המשתנים המסבירים במודל רגרסיה הדונית	4.1
19	בדיקת הנחות רגרסיה ליניארית	4.2
22	חישוב מדד דיור	4.3
23	סיכום	5
24	נספח מידע סטטיסטי ותרשימים	6

1 מבווא

מטרת עבודה זו הייתה לבחון אפשרות התאמת מודל לחישוב שינויים במחירי הדירות על סמך נתוני עסקאות בגין מכירת נכסי מגורים (שע"מ). עצם העובדה כי בכל חודש נמכרים נכסי מגורים בעלי מאפיינים שונים ובאיכות שונה, כגון דירות בעלי מספר חדרים שונה, דירות הנמצאות באזורים בהם נקבעה רמת ארנונה שונה, ותכונות אחרות כמו מספר קומה, סוג מגורים כמו בית פרטי או דירה בבניין, מחייב בניית מודל המנכה את שינויי איכות נכס המגורים מחודש לחודש. השיטה הידועה והמקובלת לביצוע התאמה מסוג זה נקראת שיטה הדונית, וכאן אנו דנים בגרסה הדונית, כלומר, גרסה של משתנה מחיר הדירה על משתנים המודדים את איכות הדירה. כשלב מקדים לבניית התוכנית לפיתוח מודל כלל ארצי לחישוב מדד דירות בבעלות פרטית בישראל נבחרה העיר חיפה לניסוי התאמת מודל גרסה הדונית למחירי דירות. באמצעות המודל שנבחן נתקבלה סדרה של אחוזי שינוי חודשיים במחירי דירות בחיפה המתייחסים לתקופה של ינואר 1999 ועד יוני 2004. בשלב מקדים זה נבחנו רלוונטיות ומובקות המשתנים המסבירים, יציבות המודל על פני זמן וטיב ההתאמה לאורך התקופה הנחקרת. דו"ח זה מכיל תיאור תהליך בדיקת אמינות ושלמות נתונים ופעולות הכנה נוספות של הרשומות לקראת התאמת המודל, וכן שלבי התאמת המודל וניתוח ראשוני של ממצאים.

2 בדיקת אמינות ושלמות הנתונים

2.1 ניפוי רשומות בקובץ

הקובץ המעובד של שע"מ המכיל מידע על עסקאות שבוצעו בין חודש ינואר 1999 לחודש יוני 2004 בגין מכירת דירות כלל 19,942 רשומות ושימש כקובץ העבודה לצורכי מחקר וניתוח. על מנת להימנע מהכללת נתונים חוזרים בניתוח אותרו ונפחו רשומות שהופיעו יותר מפעם אחת בקובץ. באופן כזה נגרעו מהקובץ 4,777 רשומות שעבור כל אחת מהן נמצאה רשומה זהה בכל ערכי המאפיינים המופיעים בקובץ העבודה.

2.1.1 גריעת רשומות שאינן נכללות באוכלוסיית היעד (רשומות לא רלוונטיות למחקר)

לצורך המחקר נכללו רשומות המתייחסות לעסקאות הבאות בלבד:

- א. דירות ובתים לצורך מגורים בלבד (קוד 10 בטבלת תפקוד כללי, ראה טבלה 7 בנספח). שאר הרשומות הכוללות עסקאות בגין מכירת מבנים, דירות ובתים שלא לצורכי מגורים נגרעו מהקובץ.
- ב. עסקאות על מכירת דירות ובתים מאדם פרטי לאדם פרטי ומחברה פרטית לאדם פרטי בלבד ע"פ קוד סוג קונה-מוכר 11 ו-21.

2.1.2 גריעת רשומות לגביהן אזור ארנונה מופיע כחסר

כל הרשומות בהן משתנה EZOR123 (ראה טבלה 1) המציין את שייכות הדירה לאזור ארנונה הופיע חסר נגרעו מהקובץ. הסיבה לגריעתן היא חוסר יכולת לשייך את הדירה לאזור הספציפי ע"פ רישום בעירייה ובהתאם לכך חוסר אפשרות לכלול רשומות אלה בניתוח הכולל. סה"כ נגרעו 1,043 רשומות עבורן אזור הארנונה הופיע כחסר. בתום טיפול בכפל רשומות ולאחר גריעת אלה שאינן רלוונטיות למחקר ובעלות מידע חיוני חסר נותרו לצורך מחקר 13,588 רשומות בקובץ חיפה.

2.2 הגדרת משתנים

בטבלה מס' 1 מופיעה רשימת המשתנים המקוריים כפי שהופיעו בקובץ שע"מ. המשתנים הללו שימשו בסיס לעבודה המתוארת כאן, כלומר, לבדיקות שלמות ואמינות הנתונים, לניפוי נתונים אי-רלוונטיים, לבדיקות לוגיות על הרשומות ולהגדרת משתנים נוספים לצורך בניית מודל לחישוב שינויים במחירי דירות בעיר חיפה.

2.2.1 הגדרת מחיר הדירה

קובץ העבודה מכיל שני משתנים שונים המתייחסים למחיר הדירה¹: מחיר מוערך ומחיר מוצהר. המחיר המוערך מתייחס לעלות הדירה המוערכת ע"י שמאי, והמחיר המוצהר מדווח ע"י הצדדים בעסקה לצורכי מיסוי ורישום. בהשוואה בין ערכי המשתנים הללו נמצאו בחלק מהרשומות הבדלים משמעותיים בין שני המחירים הנובעים ע"פ הערכה עקב רצונם של הצדדים בעסקה לדווח על מחיר נמוך יותר ע"מ לשאת בתשלומי מס ורישום קטנים יותר. כדי לאמת האם המחיר המוערך אכן משקף טוב יותר את העלות האמיתית של הנכס נעשו בדיקות לוגיות עבור המקרים בהם נמצא הבדל משמעותי בין שני המחירים באמצעות מאפיינים נוספים של הנכס כגון: מספר חדרים, אזור ארנונה והכתובת הרשומה של הנכס. בתום הבדיקות נבחר משתנה "מחיר המוערך" של הנכס כמחיר הדירה בניתוח ובניית המודל.

הערה: כל הערכים של משתנה MEHIR_MUARAH של הדירה השווים ל-0 הוחלפו בערך חסר על מנת לא להכלילם בניתוח הסטטיסטי.

¹ מונח הדירה מובא כאן גם בהקשר של בית פרטי, קוטג', או כל נכס אחר המיועד לצורכי מגורים בלבד

2.2.2 סיווג לקבוצות ע"פ מספר חדרים

משתנה קטגורי (KVUDIR) המסווג את נכסי מגורים לקבוצות שונות בהתאם למספר חדרים בדירה/בית הוגדר על בסיס משתנה המציין את מספר החדרים בדירה/בית (MISPAR_HADARIM) ומקבל ערכים בהתאם לחלוקה המופיעה בטבלה 2.

טבלה 1: רשימת משתנים של קובץ דירות שע"מ

שם משתנה	תאור המשתנה
EZOR123	משתנה קטגורי המציין את איזור הארנונה ע"פ חלוקה של עיריית חיפה (3 אזורים)
GUSH	מס' גוש בו נמצאים המגורים ע"פ רישום העיריה
HILCA	מס' חלקה בה נמצאים המגורים ע"פ הרישום בעירייה
KOMA	מס' קומה שבה נמצאת דירת מגורים
MAHUT_SHUT	משתנה המציין את הזכויות על הנכס (בעלות, חכירה, הרשאה, אופציה וכד')
MEHIR_LM2	מחיר למ"ר של המגורים
MEHIR_LM2_AVRG	מחיר ממוצע למ"ר (מחושב לכל אזור ארנונה בנפרד)
MEHIR_MUARAH	מחיר מוערך ע"י שמאי
MEHIR_MUZHAR	מחיר מוצהר בעת דווח ורישום עסקה
MISPAR_BAIT	מספר הבית ע"פ כתובת מגורים
MISPAR_HADARIM	מספר חדרים במקום מגורים
MISPAR_KOMOT	מספר הקומות בבניין/בית
MISPAR_MAALIOT	מספר המעליות בבניין
SEMEL_REHOV	סמל רחוב
SEMEL_SHUNA	סמל שכונה
SHEM_REHOV	שם רחוב
SHEM_SHUNA	שם שכונה
SHETAH_BRUTTO	שטח ברוטו של המגורים (כולל כל התוספות ע"פ חישוב מקובל)
SHETAH_MADAD	משתנה עזר לחישוב שטח לצורכי מגורים ע"פ נוסחה נתונה
SHETAH_NETTO	שטח נטו (שטח לצורכי מגורים בלבד ע"פ חישוב מקובל)
SHNAT_BNIA	שנת בניה של המגורים
SUG_KONE_MOHER	משתנה המציין את הצדדים המעורבים בעסקת מכירה/רכישה של המגורים (סמל עסקה)
TAARIH_DIVUAH	תאריך שבו דווח לרשם על העסקה בגין מכירה/רכישה של המגורים
TAARIH_ISCA	תאריך שבו בוצעה העסקה בפועל
TIFKUD_MEGURIM	ייעוד תפקודי של הנכס (מגורים, תעשייה, מסחר, משרדים וכד')

טבלה 2: סיווג קבוצות נכסי מגורים ע"פ מספר חדרים

מספר חדרים	ערך משתנה KVUDIR
1.5-2	2
2.5-3	3
3.5-4	4
4.5-5	5
5.5 +	6

2.2.3 סיווג ע"פ תפקוד מגורים

משתנה סוג מגורים (CODMEG) הוגדר לצורך סיווג נכס המיועד למגורים בלבד בהתבסס על המשתנה תפקוד מגורים באופן הבא:

1 - בתים פרטיים ודירות פאר כגון: קוטג'ים, וילות, בתים צמודי קרקע,

(קוד 1,2,3,4,5,8,9,37,40,50,55,60 ברשימת תפקוד מגורים).

2 - דירות בבנייני מגורים (קוד 10,11,12,13,70 ברשימת תפקוד מגורים)

3 - דירות דופלקס ודירות גג (קוד 14,15,16,30,35)

לרשימה מלאה של קודים לעסקאות וסוגים של נכסי מגורים עיין בטבלה 7 בנספח לדו"ח.

הערה: המשתנה הקטגורי (CODMEG) שימש לתהליך בדיקת שלמות הנתונים, בהמשך הוגדר משתנה דיכוטומי (MEGDUM) שהתייחס לנכסי מגורים בבנייני דירות ואחרים. ראה בהמשך.

2.2.4 משתנה זיהוי בניין רב קומות (RAVKOM)

משתנה זה הוגדר על מנת להפריד בין הדירות הנמצאות בבניינים רבי קומות לבין הדירות הממוקמות בבתיים של 4 קומות ומטה. המשתנה הוגדר כדיכוטומי וקיבל:

ערך 1, אם הבניין הוא בעל יותר מ-4 קומות (ערכו של משתנה MISPAR_KOMOT גדול מ-4), אחרת קיבל ערך 0.

2.2.5 משתנה גיל הדירה (AGE)

משתנה זה הוגדר כהפרש בין שנת העסקה לשנת הבניה.

$$AGE = SHNAT_ISCA - SHNAT_BNIYA$$

במקרים בהם קיבל משתנה AGE ערכים השווים ל-0 הוחלפו ערכים אלו ב-0.1 על מנת להימנע מפעולה בלתי חוקית של חישוב ערכי לוגריתם של המשתנה. בנוסף, בעת חישוב ערכי משתנה AGE נתקבלו ערכים שליליים של גיל הדירה עבור המקרים בהם הדירות שעסקת מכירתן התבצעה בטרם הושלמה בנייתה. גם כאן, לצורכי החישובים הוחלפו הערכים השליליים של AGE בערך 0.1.

2.2.6 משתנה דירה חדשה/ישנה (HADASH)

משתנה המסווג את נכסי המגורים לחדשים/ישנים (HADASH) הוגדר כך שהוא מקבל ערך 1 אם גיל הדירה ע"פ ההגדרה בד' קטן או שווה ל-1, כלומר הדירה היא חדשה, ו-0 אחרת.

2.3 זקיפת ערכי נתונים חסרים

2.3.1 זקיפת שטח הדירה (SHETAH_MADAD)

בקובץ הנתונים המקורי הופיעו שלושה משתנים המתייחסים לשטח הדירה: שטח ברוטו (SHETAH_BRUTTO), שטח נטו (SHETAH_NETTO) ושטח מדד (SHETAH_MADAD) בבדיקה התברר כי שני המשתנים האחרונים מציינים את שטח הדירה בפועל (נטו), אך בשניהם נמצאו ערכים חסרים. המשתנה (SHETAH_MADAD) נבחר לשמש כמשתנה רשמי לציון שטח הדירה נטו בפועל, כאשר ערכיו החסרים נזקפו באמצעות ערכי משתנה SHETAH_NETTO. במקומות שלא היה ניתן לזקוף ערכים באמצעות ערכי SHETAH_NETTO חושב השטח לצורכי מגורים באופן הבא: אם שנת הבניה של הדירה קטנה מ-1970, הוכפל SHETAH_BRUTTO ב-0.85, אחרת הוכפל SHETAH_BRUTTO ב-0.8.

2.3.2 תיקון ערכי משתנה קומה (KOMA)

בקובץ חיפה נתגלה מספר רב של רשומות (3195 מתוך 14631 רשומות, כ-22%) בהן משתנה מספר קומה (KOMA) קיבל ערך 0. מתוך ידיעה כללית כי בבניינים רבי קומות בדרך כלל אין קומת קרקע נעשתה בדיקת אמינות נתון זה ובוצע תהליך זקיפת קומה באופן הבא:

א. אם ערך משתנה KOMA נתגלה כחסר והדירה השתייכה לקטגוריית בתים פרטים/וילות (CODMEG=1), אזי ערך משתנה KOMA נקבע ל-0.

ב. אם ערך משתנה KOMA שווה ל-0 והדירה השתייכה לקטגוריית דירות בבנייני קומות (CODMEG=2) ומספר הכולל של הקומות בניין עולה על 4, (MISPAR_KOMOT>4) נקבע ערכו החדש של משתנה KOMA באותה רשומה כערך השווה לממוצע משתנה הקומה באותה הקטגוריה.

ג. אם ערך משתנה KOMA שווה ל-0 ברשומה כלשהי, והדירה השתייכה לקטגוריית דירות גג (CODMEG=3), הנמצאות בקומה האחרונה ע"פ ההגדרה, אזי ערך משתנה KOMA באותה רשומה נקבע כערכו של משתנה MISPAR_KOMOT.

2.4 ערכים חריגים

2.4.1 שטח מדד

בכל קטגוריה של דירות ע"פ מספר חדרים נופו רשומות בהן ערך של משתנה SHETAH_MADAD של הדירה חרג מהגבולות המפורטים בטבלה 3. סה"כ נופו 46 רשומות כאלה.

טבלה 3: קריטריונים לניפוי רשומות עם ערכים חריגים במשתנה שטח הדירה

מספר חדרים	גבול תחתון (מ"ר)	גבול עליון (מ"ר)
1.5-2	10	77
2.5-3	37	100
3.5-4	55	132
4.5-5	73	165

2.4.2 מחיר הדירה (MEHIR_MUARAH)

איתור ערכים חריגים של מחירי הדירות נעשה באמצעות בחינת התפלגות משתנה "מחיר הדירה למ"ר" (MEHIR_LM2_TI) בכל תת-קטגוריה של סוג מגורים (בתים פרטיים/דירות בבניינים), אזור ארנונה (1,2,3) ושנה וחודש עסקה (מינואר 1999 עד יוני 2004). עקב השוני במדדי מרכז ופיזור של התפלגויות מחיר למ"ר באזורי ארנונה שונים נקבעו קריטריונים שונים של איתור וגריעת ערכים חריגים בהתאם לאזורים השונים כדלקמן:

נגדיר $IR^{(i)} = Q_{(75)}^i - Q_{(25)}^i$ כתחום בין-רבעוני של התפלגות משתנה MEHIR_LM2_TI בתת קטגוריה מסוימת של סוג מגורים, שנה וחודש עסקה ואזור ארנונה $i = 1, 2, 3$; כאשר $Q_{(\alpha)}^i$ מציין את האחוזון ה- α של התפלגות MEHIR_LM2_TI בתת קטגוריה זו, כלומר, התחום הבין-רבעוני מחושב ע"י הפרש בין האחוזון ה-75 וה-25 של התפלגות MEHIR_LM2_TI בתת קטגוריה מסוימת. (א) אם הדירה נמצאת באזור ארנונה 3 (הפחות יוקרתי משלושתם) אזי ערך חריג נקבע במידה ומחיר למ"ר של הדירה נמצא מעבר למרחק של תחום בין רבעוני מהאחוזון ה-25 או מעבר למרחק של 1.5 תחומים בין רבעוניים מהאחוזון ה-75. כלומר רשומה תגרע מהקובץ כבעלת ערך חריג במשתנה MEHIR_LM2_TI אם לא יתקיים:

$$Q_{(25)}^{(3)} - 1.0IR^{(3)} < MEHIR_LM2_TI^{(3)} < Q_{(75)}^{(3)} + 1.5IR^{(3)}$$

הסיבה לכך כי התחום לערכים בלתי חריגים נקבע כלא סימטרי היא צורת ההתפלגות של מחיר הדירה למ"ר באזור ארנונה 3 (התפלגות אסימטרית עם זנב בצד של התצפיות מעל הממוצע, ראה תרשים 1). מספר תחומים בין רבעוניים נקבע בהתאם למידת הפיזור של התצפיות. ניתן לראות בתרשים 1 כי התצפיות מרוכזות סביב הממוצע ולכן הערכים החריגים מאותרים באופן אסימטרי ובהתאם למידת הפיזור של ההתפלגות.

(ב) אם הדירה נמצאת באזור ארנונה 2 אזי ערך חריג נקבע במידה ומחיר למ"ר של הדירה נמצא מעבר למרחק של 1.5 תחומים בין רבעוניים מהאחוזון ה-25 או מעבר למרחק של 2 תחומים בין רבעוניים

מהאחזון ה-75 (ראה תרשים 2 להמחשת האסימטריות ומידת הפיזור של ההתפלגות), כלומר אם לא מתקיים:

$$Q_{(75)}^{(2)} - 1.5IR^{(2)} < MEHIR_LM2_TI^{(2)} < Q_{(75)}^{(2)} + 2.0IR^{(2)}$$

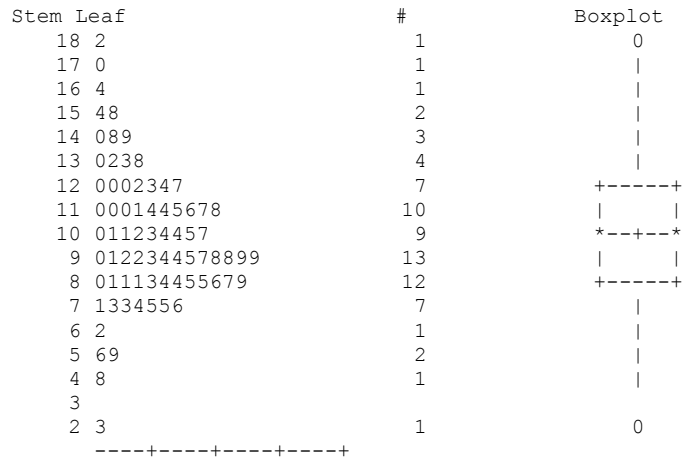
תרשים 1 : התפלגות שטח הדירה למ"ר – אזור 3 – שנת עסקה - 1999, חודש עסקה - מרס

Stem Leaf	#	Boxplot
13 3	1	0
12		
12		
11		
11		
10		
10		
9 69	2	
9 01	2	
8 9	1	
8 0123344	7	
7 68889	5	+-----+
7 01233344	8	
6 55667899	8	*---+---*
6 01334	5	
5 5677777888	11	+-----+
5 01122333	8	
4 668	3	
4 1	1	
3		
3		
2		
2 1	1	0
1		
1		
0 6	1	0

תרשים 2 : התפלגות שטח הדירה למ"ר – אזור 2 – שנת עסקה - 1999, חודש עסקה - מרס

Stem Leaf	#	Boxplot
15 0	1	0
14		
14		
13 7	1	0
13 0	1	
12		
12 2	1	
11 579	3	
11 234	3	
10 56777	5	
10 000122244444	12	
9 5555666677778999	17	+-----+
9 00001111233344444	18	
8 55666778888889	14	*---+---*
8 1111112233344	13	+
7 55555666688	11	
7 0001111122233444	17	+-----+
6 56668899	8	
6 022334	6	
5 55777788	8	
5 0000122	7	
4 789	3	
4 34	2	
3		
3		
2 6	1	0

תרשים 3: התפלגות שטח הדירה למ"ר – אזור 1 – שנת עסקה - 1999, חודש עסקה - מאי



ג) אם הדירה נמצאת באזור ארנונה 1 (היוקרתי ביותר) אזי ערך חריג נקבע במידה ומחיר למ"ר של הדירה נמצא מעבר למרחק של 1.75 תחומים בין רבעוניים מהאחוזון ה-25 או מעבר למרחק של 2 תחומים בין רבעוניים מהאחוזון ה-75, (ראה תרשים 3 להמחשת ההתפלגות), כלומר אם לא מתקיים:

$$Q_{(25)}^{(1)} - 1.75IR^{(1)} < MEHIR_LM2_TI^{(1)} < Q_{(75)}^{(1)} + 2IR^{(1)}$$

בתום תהליך זה, מתוך 13,542 רשומות שעמדו למבחן ערכים חריגים של מחיר למ"ר זוהו ונכחו מהקובץ 331 רשומות בהן אותרו ערכי מחיר למ"ר כחריגים (2.44%, ראה התפלגות ערכים חריגים בנספח). כך נותרו לצורך בניית מודל סטטיסטי 13,211 רשומות מתוך 19,942 בקובץ העבודה.

3 בניית מודל רגרסיה הדונית ללוג-מחיר הדירה

3.1 המודל ההדוני לחישוב השינויים במחירי דירות – המודל התקופתי

נניח כי העסקאות בגין מכירת דירות פרושות על פני T חודשים. נגדיר סדרה של מודלים ליניאריים נעים בזמן, כאשר כל מודל בסדרה מבוסס על הנתונים של τ חודשים עוקבים, והמודל הראשון מותאם לנתונים המתייחסים לתקופת החודשים $[1, K, \tau]$; המודל השני מתייחס לתקופה המוזנת בחודש אחד, כלומר לחודשים $[2, K, \tau + 1]$, והמודל האחרון בסדרה זו מותאם לנתוני תקופה החודשים $[T - \tau + 1, K, T]$. באופן כללי, המודל ה- k תכלול בתוכה החודשים $[k, K, \tau + k - 1]$. סדרת המודלים הליניאריים מונה $T - \tau + 1$ מודלים מותאמים לתקופה של τ חודשים הזזים כל פעם בחודש אחד. המודל לתקופה $k, T - \tau + 1, K, k = 1$ מותאם באופן הבא:

$$\log Y_{jtk} = \beta_k X_{jtk} + \delta_k I_{jk}(t) + \varepsilon_{jk} \quad \varepsilon_{jk} \sim N(0, \sigma_k) \quad (3.1)$$

כאשר Y_{jtk} הוא מחיר הדירה j בחודש t בתקופה k ; β_k הוא וקטור ממימד $1 \otimes P$ של מאפייני דירה הנכללת במודל רגרסיה ליניארית בתקופה k ; $X_{jtk} = [\log(x_{jtk}^1), K, \log(x_{jtk}^N), x_{jtk}^{N+1}, K, x_{jtk}^P]^T$; וקטור ממימד $P \otimes 1$ של מאפייני דירה j , בחודש t , במודל בתקופה k . וקטור המאפיינים מורכב מ- N מאפיינים שהם משתנים רציפים ונכללים במודל לאחר הפעלת טרנספורמציה לוגריתמית, ו- $P-N$ מאפיינים הם משתנים קטגוריים (דיכוטומיים). $I_{jk}(t) = [I_j\{t = k+1\}, K, I_j\{t = \tau+k-1\}]^T$ הוא וקטור ממימד $(\tau-1) \otimes 1$ מציינות של $\tau-1$ חודשים אחרונים בתקופה k ו- $[\delta_k^2, K, \delta_k^\tau]$ הוא וקטור של $\tau-1$ מקדמי מציינות במודל רגרסיה בתקופה k .

3.2 חישוב שינויים במחירי דירות בבעלות פרטית

אמידת הפרמטרים של המודל (3.1) מתבססת על שיטת הריבועים הפחותים. מודל זה מאפשר חישוב שינויים במחירי הדירות בתקופה k ללא צורך בחישוב "דירה ממוצעת". עבור חודש t , $t = 2, K, T$, נחשב את השינוי החודשי במחיר הדירה באופן הבא:

$$\hat{R}_{tk} = \frac{\hat{Y}_{.tk}}{\hat{Y}_{.(t-1),k}} = \frac{\exp(\beta_k X_{.tk} + \delta_k I_{.k}(t))}{\exp(\beta_k X_{.t-1,k} + \delta_k I_{.k}(t-1))} = \exp(\delta_k^{t-(k-1)} - \delta_k^{t-k}) \quad (3.2)$$

כפי שניתן לראות ב-(3.2), עבור דירה מסוימת, וקטור המאפיינים בחודש t זהה לווקטור המאפיינים בחודש $t-1$, כלומר $X_{.tk} = X_{.t-1,k}$, ולכן השינוי במחיר דירה מסוימת מתבטא כפונקציה של δ_k בלבד. השימוש ב-(3.2) מאפשר לחשב עד ל- $\tau-1$ אומדנים לשינוי במחירי דירות חודש t . בהנחה כי שונויות האומדנים ע"פ המודלים השונים זהים ושווים ל- σ^2 , הרי כי באמצעות חישוב הערך הממוצע של אומדנים אלו על פני כל המודלים האומדים את השינוי במחירי הדירות בחודש t , נקבל אומדן לשינוי במחירי הדירות בעל שונות נמוכה יותר $\sigma^2/(\tau-m-1)$, כאשר m מציין את מספר המודלים החסרים עד ל- $(\tau-1)$, $m = 0, K, \tau-2$. למשל, במודל השנתי, כלומר $\tau = 12$, עבור מקרי קצה, השינוי במחירי דירות במרס 1999 יכול להיות מחושב רק באמצעות שני מודלים בלבד, האחד מותאם לחודשים ינואר 1999 עד לדצמבר 1999, והשני לחודשים פברואר 1999 עד ינואר 2000 (נשים לב, כי במקרה השני $\hat{R}_{32} = \exp(\delta_3^2)$, היות ו- δ_3^1 אינו מוגדר ע"פ (3.1), ולכן לצורכי חישוב (3.2) $\delta_3^1 = 0$!). כמו כן, חשוב לשים לב, כי לא ניתן לחשב את השינוי במחירי הדירות במרס 1999 במודל המותאם עבור החודשים מרס 1999 עד פברואר 2000, היות ומקדם המציינות של חודש מרס 1999 אינו מוגדר במודל זה. כלומר, במקרה שהוסבר זה עתה, $m = (\tau-1) - 2 = 9$, ראה הסבר מפורט בהמשך).

כפי שתארנו קודם, ניתן לחשב ממוצע שינויי מחירים של דירות על פני האומדנים על פי המודלים הנעים בזמן באופן הבא:

$$\widehat{R}_t = (\tau - m - 1)^{-1} \left[\sum_{k=t-\tau+1+(t-\tau)I\{t<\tau\}}^{t-1-(t-(T-\tau+2))I\{t>T-\tau+2\}} \widehat{R}_{tk} \right] \quad (3.3)$$

$$. m = -((t - (T - \tau + 2))I\{t > T - \tau + 2\} + (\tau - t)I\{t < \tau\})$$

הערה: נשים לב כי רק אחת מבין המציינות ב-3.3 מקבלות ערך של 1 במקרי קצה, ובמקרים רגילים שתיהן מתאפסות.

כעת נתייחס לנתוני קובץ הדירות ששימש אותנו לצורך התאמת מודלים. נתוני הקובץ כאמור כוללים עסקאות מחודש ינואר 1999 ועד יולי 2004, כלומר סה"כ 67 חודשים. נניח כי כל מודל של רגרסיה ליניארית הותאם ל-12 חודשי נתונים עוקבים, כלומר, $(\tau = 12)$, כפי שהוסבר בסעיף 3.1. באופן כזה המודל הראשון יתבסס על נתוני ינואר 1999 ועד דצמבר 1999, המודל השני יותאם עבור נתוני פברואר 1999 ועד ינואר 2000, וכך הלאה, עד שהמודל האחרון ייבנה על סמך נתוני אוגוסט 2003 ועד יולי 2004. עקב החפיפה בחודשים בין המודלים, אומדן השינוי במחירי דירות מחושב מספר פעמים, וכאמור במקסימום $\tau - 1$ פעמים, או במקרה הספציפי המתואר כאן, 11 פעמים, ובמינימום פעם אחת בלבד. לדוגמה: ברצוננו לאמוד שינוי במחירי הדירות בין החודשים פברואר 2000 וינואר 2000. 11 מודלים שונים של רגרסיה ליניארית מותאמים באופן נע ומסתמכים על נתוני דירות של חודש פברואר 2000: הראשון ביניהם מותאם לנתוני מרס 1999 ועד פברואר 2000. בתקופה זו חודש פברואר 2000 הוא האחרון בסדרת החודשים. האומדן האחרון לשינוי במחירי דירות לחודש פברואר 2000 מתקבל מהמודל המותאם עבור התקופה מינואר 2000 עד דצמבר 2000, כאשר חודש פברואר 2000 הוא השני בסדרת החודשים. נציין, כי במודל שבו פברואר 2000 מופיע כחודש ראשון, לא ניתן לחשב שינוי במחירי דירות לחודש זה, כאמור ב-3.1. נרשום את סדרת אומדני המדד ואופן חישובם: חודש פברואר 2000 הוא החודש ה-14, $(t = 14)$ מבין ה-67, $(T = 67)$ חודשים מינואר 1999 ועד יולי 2004. עבור חודש זה $m = 0$, כלומר זהו אינו מקרה קצה. המודל הראשון שבו מופיע חודש זה של נתונים הוא המודל השלישי $(k = 3)$ בסדרה של $56 = 67 - 12 + 1 = T - \tau + 1$ מודלים. המודל האחרון שבו מופיע פברואר 2000 הוא המודל השלושה עשר $(k = 13)$. השינוי במחירי דירות לחודש פברואר 2000 יחושב, אם כן, באמצעות 11 אומדנים שונים באופן הבא:

$$\hat{R}_{14,3} = \frac{Y_{14,3}}{Y_{13,3}} = (\delta_3^{12} - \delta_3^{11}), k = 3$$

$$\hat{R}_{14,4} = \frac{Y_{14,4}}{Y_{13,4}} = (\delta_4^{11} - \delta_4^{10}), k = 4 \quad (3.4)$$

ℕ

ℕ

$$\hat{R}_{14,13} = \frac{Y_{14,13}}{Y_{13,13}} = (\delta_{13}^2 - \delta_{13}^1), k = 13$$

נשים לב כי במקרה האחרון $\delta_{13}^1 = 0$ היות ובמודל האחרון בו מופיע חודש פברואר 2000 מקדם המציינת של שייכות הנתונים לחודש ינואר 2000 שווה ל-0. האומדן הסופי יתקבל כממוצע של 11 אומדנים אלו באופן הבא:

$$\bar{\hat{R}}_{14} = \frac{1}{11} \sum_{k=3}^{13} \hat{R}_{14,k}$$

3.3 בחירת משתנים מסבירים במודל

כפי שנאמר בפרק הקודם מענייננו להתאים מודל אחיד המותאם לתקופות נעות של חודש אחד, כלומר, כל אחד מהמודלים המותאמים לתקופות השונות יכלול אותם משתנים מסבירים מובהקים סטטיסטית. לכן, עלינו לקבוע, מי מהמשתנים הפוטנציאליים יכלולו במודלים, כך שמקדמיהם יהיו מובהקים לאורך הסדרה והשינויים בערכי מקדמיהם במודלים העוקבים יהיו ככל האפשר מינימאליים, כלומר עדיין הנחת שוויון מקדמי המודלים בתקופות מספיק קצרות תהיה סבירה.

בעבודה זו בחרנו לבחון שתי סדרות של מודלים, האחת באורך שנה, $\tau = 12$, והשנייה באורך חודשיים $\tau = 2$, בדומה למודל המקובל כיום לחישוב שינויים במחירי הדירות. במטרה לקבוע אילו מאפיינים של

הדירה ישמשו כמשתנים מסבירים בסדרת המודלים של רגרסיה הדונית, נקבעו מספר קריטריונים: א. רלוונטיות למחקר – נקבע כי על המשתנה המסביר להיות רלוונטי למחקר ובעל פרשנות כלכלית, כלומר עליו לאפיין את הדירה במפורש, לשמש כמדד איכות למחיר הדירה ולהיות מתואם עם מחיר הדירה במתאם שונה מ-0 באופן מובהק. בהקשר זה יצוין כי המאפיינים שסוננו מהמחקר היו קשורים בעיקר לפרטי הדירה כגון הכתובת, סמל שכונה ושם הרחוב. מידת הפרשנות של המשתנים המסבירים נידונה בעיקר בפרק העוסק בהגדרת משתני אינטראקציה (ראה פרק 3.4).

ב. אמינות ערכי המשתנים – על המשתנה לכלול מספר מזערי ככל האפשר של ערכים חסרים וערכים לגביהם קיים ספק בדבר אמינותם. (ראה פרקים 2.3 ו-2.4 העוסקים בזקיפת ערכים חסרים וטיפול בחריגים).

ג. יציבות המשתנה לאורך הסדרה – קריטריון זה מתייחס לתדירות הופעתו של המשתנה המסביר כמובהק לאורך המודלים ואחידות בכיוון השפעתו על מחיר הדירה לאורך הסדרה (יציבות סימן המקדם).

ד. סדרה "חלקה" של המקדמים בזמן – על מנת להקנות למודל יציבות נוספת נדרש כי השינויים לאורך זמן של אומדני המקדמים יהיו הדרגתיים ללא קפיצות ושינויי סימן פתאומיים לאורך הסדרה. בטבלה 5 רוכזו כל המשתנים שנבחרו לצורך שילובם במודל או נלקחו בהגדרתם המקורית מקובץ נתוני דירות. המשתנים הללו עברו מבחן של רלוונטיות, אמינות ועמדו בהמשך למבחני יציבות וסטציונריות לפני שילובם במודל הסופי.

3.4 בחירת משתנים מסבירים למודל

3.4.1 בניית סדרה של מודלים ליניאריים באורך שנה

בשלב זה הותאמו 56 מודלים ($T = 67, \tau = 12$) של רגרסיה ליניארית בצעדים (stepwise) על המשתנה LOGMEHIR, כאשר בצעד הפתיחה שולבו כל המשתנים המופיעים בטבלה 5. כל מודל הותאם עבור 12 חודשי נתונים עוקבים בשיטה שתוארה ב- 3.1. בכל צעד הרצה הטיפול במשתנים מסבירים פוטנציאליים נעשה באופן הבא:

א. הוסף משתנה למודל, אם רמת המובהקות של מקדמו ברגרסיה הינה לכל היותר 0.05.

ב. גרע משתנה מהמודל, אם רמת מובהקות של מקדמו היא 0.10 לכל הפחות.

ג. עצור, אם אין עוד משתנים לגריעה והוספה.

תהליך קבלת החלטה אילו משתנים יכללו בסדרת המודלים נעשתה על פי השלבים הבאים:

(1) ריכוז תוצאות הרצה של שלבים א' עד ג' בטבלה וביצוע מעקב לאורך חודשי התקופה אחר משתנים שנגרעו מהמודלים בעקבות ביצוע שלבי א' עד ג'.

(2) סינון משתנים שלאורך חודשי התקופה הנחקרת תדירות גריעתם מהמודלים היא הגבוהה ביותר ובפרט אם היא גבוהה ממחצית המקרים (למעט אינדיקטורים המציינים את שייכות הנתונים לחודש מסוים).

(3) הרצה נוספת של סדרת המודלים (בשיטת enter) עם כל המשתנים המסבירים שנבחרו להישאר במודל ע"פ 2). מעקב אחר מובהקות המשתנים לאורך הסדרה וגריעת משתנים בלתי יציבים שתדירות מובהקותם לאורך הסדרה היא נמוכה.

טבלה 4: רשימת משתנים פוטנציאליים במודלים של רגרסיה ליניארית עבור המשתנה LOGMEHIR

KOMA	מס' קומה בה נמצאת הדירה (משתנה מקורי של קובץ ש"מ)
LOGSHMD	הלוגריתם הטבעי של משתנה שטח הדירה SHETAH_MADAD
RAVKOM	משתנה דיכוטומי המקבל ערך 1 אם הדירה נמצאת בבניין רב קומות ו-0 אחרת
LOGAGE	הלוגריתם הטבעי של משתנה גיל הדירה AGE ע"פ הגדרה 2.3.4
LSHMDAGE	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה לוג-גיל הדירה ולוג-שטח הדירה SHETAH_MADAD* AGE
HADASH	משתנה שייכות הדירה לדירות חדשות ע"פ ההגדרה ב- 2.3.5
RAVHAD	משתנה דיכוטומי שהוא אינטראקציה כפולית של משתנה RAVKOM ו-HADASH
MEGDUM	משתנה דיכוטומי המקבל ערך 1 אם הדירה היא בבניין מגורים ו-0 אחרת
KVUDUM_2	משתנה דיכוטומי המקבל ערך 1 אם הדירה שייכת לקבוצה 2 (היא בת 1.5-2 חדרים) ו-0 אחרת
KVUDUM_3	משתנה דיכוטומי המקבל ערך 1 אם הדירה שייכת לקבוצה 3 (היא בת 2.5-3 חדרים) ו-0 אחרת
KVUDUM_4	משתנה דיכוטומי המקבל ערך 1 אם הדירה שייכת לקבוצה 4 (היא בת 3.5-4 חדרים) ו-0 אחרת
EZORDUM_1	משתנה דיכוטומי המקבל ערך 1 אם הדירה היא באזור ארנונה 1 ו-0 אחרת
EZORDUM_2	משתנה דיכוטומי המקבל ערך 1 אם הדירה היא באזור ארנונה 2 ו-0 אחרת
KVUEZDUM_21	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ואזור ארנונה 1
KVUEZDUM_31	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ואזור ארנונה 1
KVUEZDUM_41	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ואזור ארנונה 1
KVUEZDUM_22	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ואזור ארנונה 2
KVUEZDUM_32	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ואזור ארנונה 2
KVUEZDUM_42	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ואזור ארנונה 2
KVURAV_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ומשתנה RAVKOM
KVURAV_3	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ומשתנה RAVKOM
KVURAV_4	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ומשתנה RAVKOM
KVUAGE_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים וגיל הדירה
KVUAGE_3	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים וגיל הדירה
KVUAGE_4	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים וגיל הדירה
KVUHAD_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ומשתנה HADASH
KVUHAD_3	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ומשתנה HADASH
KVUHAD_4	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ומשתנה HADASH
NUMMON2	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-2 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON3	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-3 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON4	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-4 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON5	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-5 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON6	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-6 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON7	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-7 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON8	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-8 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON9	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-9 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON10	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-10 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON11	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-11 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON12	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-12 מתוך 12 החודשים הנכללים במודל רגרסיה שנתית
NUMMON	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-2 מתוך 12 החודשים הנכללים במודל רגרסיה זו חודשית
EZORAVDUM_1	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות הדירה לבניין רב קומות לבין שייכות הדירה לאזור ארנונה 1
EZORAVDUM_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות הדירה לבניין רב קומות לבין שייכות הדירה לאזור ארנונה 2
KVUMEG_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ושייכות לסוג מגורים דירה בבניין
KVUMEG_3	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ושייכות לסוג מגורים דירה בבניין
KVUMEG_4	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ושייכות לסוג מגורים דירה בבניין
MEGHAD	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה שייכות לסוג מגורים (MEGDUM) ושייכות לדירות חדשות (HADASH)
MEGRAV	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה שייכות לסוג מגורים (MEGDUM) ושייכות לדירות בבניין רב קומות (RAVKOM)
HADEZDUM1	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה שייכות לדירות חדשות (HADASH) ושייכות הדירה לאזור ארנונה 1
HADEZDUM2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה שייכות לדירות חדשות (HADASH) ושייכות הדירה לאזור ארנונה 2
EZMEGDUM1	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה שייכות ושייכות הדירה לאזור ארנונה 1 ושייכות לסוג מגורים דירה בבניין
EZMEGDUM2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה שייכות ושייכות הדירה לאזור ארנונה 2 ושייכות לסוג מגורים דירה בבניין
KVUSHMD_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ולוג-שטח הדירה
KVUSHMD_3	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ולוג-שטח הדירה
KVUSHMD_4	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ולוג-שטח הדירה

3.4.2 בניית סדרה של מודלים ליניאריים זו חודשיים

באופן דומה לזה שתואר ב-3.3.1 נעשה ניתוח סידרת מודלים ליניאריים זו חודשיים באמצעות הרצת 65 מודלים $(T = 67, \tau = 2)$ של רגרסיה ליניארית זו חודשיים בשיטת חודש נע. באופן כזה המודל הראשון התבסס רק על הנתונים של חודשי ינואר ופברואר 1999, המודל הבא על נתוני פברואר ומרס 1999, וכך הלאה עד שהמודל האחרון הותאם רק עבור הנתונים המתייחסים לחודשים של מאי ויוני 2004.

4 ממצאים

4.1 המשתנים המסבירים במודל רגרסיה הדונית

בטבלה מס' 5 מוצגים המשתנים המסבירים של המודל שנתקבלו לאחר השלב של סינון וגריעת משתנים אי-רלוונטיים כפי שתואר לעיל. במודלים של רגרסיה מופיעים שלושה סוגים של משתנים מסבירים

4.1.1 משתנים מסבירים רציפים

מקדמי המשתנים הרציפים: לוג-גיל הדירה, לוג-שטח הדירה (LOGAGE, LOGSHMD מצביעים על עוצמת האפקט הראשי או השפעה של משתנה מסביר על המשתנה מחיר הדירה המבוטאת במונחי שינוי במחיר הדירה כתוצאה משינוי ביחידה אחת של מאפיין הדירה.

4.1.2 משתנים דיכוטומיים וקטגוריים

רשימה זו כוללת את כל משתני האיכות הנמדדים בסולם שמי עם או ללא היררכיה. ביניהם משתנים דיכוטומיים המקבלים ערכים 0 ו-1 בהתאם לקטגוריה ומשתנים קטגוריים המקבלים יותר משני ערכים שמיים. האחרונים הוחלפו למספר משתנים דיכוטומיים השווה למספר הקטגוריות של המשתנה במקורי פחות 1 (עקב דרגת חופש הנוצרת כאשר כל המשתנים הדיכוטומיים מקבלים 0, אזי ברור כי נבחרה הקטגוריה שעבורה לא הוגדר משתנה דיכוטומי).

א. המשתנים הדיכוטומיים במודל:

RAVKOM – (ראה הגדרה 2.3.3)

HADASH – (ראה הגדרה 2.3.5)

MEGDUM – משתנה הנוצר על בסיס משתנה CODMEG (ראה הגדרה 2.3.2), והוא מקבל ערך של 1 אם סוג המגורים הוא בית פרטי ו-0 אם היא בבניין דירות.

משמעותם של משתנים הללו הינה שינוי במחיר כתוצאה ממעבר מקטגוריה 0 לקטגוריה 1. לדוגמה, מקדמו של משתנה CODMEG מציין בכמה מייקר/מוזיל המעבר בין דירה בבניין לבית פרטי.

ב. המשתנים הקטגוריים שהוחלפו בדיכוטומיים

המשתנה הקטגורי KVUDIR (ראה הגדרה 2.3.1) הוחלף ב-3 משתנים דיכוטומיים

2,3,4, kvudum_2, kvudum_3, kvudum_4. כל אחת מהמשתנים הללו מקבל ערך 1 אם הדירה היא בת 2,3,4 חדרים בהתאמה ו-0 אחרת. דירות בנות חדר 1 לא נכללו בניתוח. חשוב לציין כי מספר המשתנים הדיכוטומיים הוא מספר הקטגוריות פחות 1 של המשתנה הקטגורי המקורי. ברור, כי כאשר שלושת המשתנים שווים ל-0 אזי הדירה במקרה זה היא בת 5- חדרים או יותר (ראה טבלה 2).

4.1.3 אינטראקציות בין המשתנים

מלבד האפקטים הראשיים של המשתנים המסבירים במודל רגרסיה נבדקים גם אפקטים של אינטראקציה כפולית בין המשתנים השונים, המציינים את מידת השינויים של האפקט הראשי של משתנה כלשהו כתוצאה מרמות שונות של ערכי משתנה אחר איתו הראשון באינטראקציה.

טבלה 5: משתני איכות של דירה שנכללו במודל רגרסיה הדונית לאמידת שינויים במחירי דירות בחיפה

LOGSHMD	הלוגריתם הטבעי של משתנה שטח הדירה SHETAH_MADAD
LOGAGE	הלוגריתם הטבעי של משתנה גיל הדירה AGE ע"פ הגדרה 2.3.4
RAVKOM	משתנה דיכוטומי המקבל ערך 1 אם הדירה נמצאת בבניין רב קומות ו-0 אחרת
RAVHAD	משתנה דיכוטומי שהוא אינטראקציה כפולית של משתנה RAVKOM ו-HADASH (רלוונטי לגבי המודל השנתי בלבד)
LSHMDAGE	משתנה דיכוטומי שהוא אינטראקציה כפולית בין משתנה לוג-גיל הדירה ולוג-שטח הדירה AGE*SHETAH_MADAD
HADASH	משתנה שייכות הדירה לדירות חדשות ע"פ ההגדרה ב- 2.3.5
MEGDUM	משתנה דיכוטומי המקבל ערך 1 אם הדירה היא בבניין מגורים ו-0 אחרת
KVUDUM_2	משתנה דיכוטומי המקבל ערך 1 אם הדירה שייכת לקבוצה 2 (היא בת 1.5-2 חדרים) ו-0 אחרת
KVUDUM_3	משתנה דיכוטומי המקבל ערך 1 אם הדירה שייכת לקבוצה 3 (היא בת 2.5-3 חדרים) ו-0 אחרת
KVUDUM_4	משתנה דיכוטומי המקבל ערך 1 אם הדירה שייכת לקבוצה 4 (היא בת 3.5-4 חדרים) ו-0 אחרת
EZORDUM_1	משתנה דיכוטומי המקבל ערך 1 אם הדירה היא באזור ארנונה 1 ו-0 אחרת
EZORDUM_2	משתנה דיכוטומי המקבל ערך 1 אם הדירה היא באזור ארנונה 2 ו-0 אחרת
KVUEZDUM_21	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ואזור ארנונה 1
KVUEZDUM_31	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ואזור ארנונה 1
KVUEZDUM_41	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ואזור ארנונה 1
KVUEZDUM_22	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 2 ע"פ מספר חדרים ואזור ארנונה 2
KVUEZDUM_32	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 3 ע"פ מספר חדרים ואזור ארנונה 2
KVUEZDUM_42	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות לקבוצת דירות 4 ע"פ מספר חדרים ואזור ארנונה 2
NUMMON2	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-2 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON3	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-3 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON4	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-4 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON5	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-5 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON6	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-6 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON7	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-7 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON8	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-8 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON9	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-9 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON10	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-10 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON11	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-11 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON12	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-12 מתוך 12 החודשים הנכללים במודל רגרסיה שנתי בלבד
NUMMON	משתנה דיכוטומי המקבל ערך 1 אם העסקה בגין מכירת דירה נעשתה בחודש ה-2 מתוך 12 החודשים הנכללים במודל רגרסיה 12 חודשית בלבד
EZORAVDUM_1	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות הדירה לבניין רב קומות לבין שייכות הדירה לאזור ארנונה 1
EZORAVDUM_2	משתנה דיכוטומי שהוא אינטראקציה כפולית בין שייכות הדירה לבניין רב קומות לבין שייכות הדירה לאזור ארנונה 2

האינטראקציות בין המשתנים הרציפים, כגון LSHMDAGE (ראה טבלה 5) שולבו בפרוצדורת בחירת משתנים רלוונטיים על סמך מידע מוקדם אודות מקדם המתאם ביניהם ומתוך הנחה כי השפעת שטח הדירה על מחירה עשוי להשתנות ככל שגיל הדירה הולך וגדל. כמו כן נבדקו אינטראקציות בין המשתנים הרציפים למשתנים קטגוריים, וזאת במטרה לבחון שינויים של השפעת משתנה רציף כתוצאה ממעבר בין רמה אחת לשנייה במשתנה קטגורי. כך למשל קבוצת משתנים KVVUAGE_2, KVVUAGE_3, KVVUAGE_4 שהם אינטראקציה כפלית בין משתנה לוג-גיל הדירה (LOGAGE) וקבוצת משתנים KVVUDUM_2, KVVUDUM_4, KVVUDUM_3, באה לבחון את שינוי השפעת גיל הדירה על המחיר במעבר בין הרמות השונות של מספר חדרים בדירה. כמו כן הוגדרו מספר אינטראקציות בין מספר משתנים דיכוטומיים בודדים כגון RAVHAD ובנוסף בין קבוצות משתנים דיכוטומיים כגון קבוצת משתני KVUEZDUM (ראה טבלה 5).

האינטראקציות שנכללו במודל:

$$א. LSHMDAGE = \log(\text{SHETAH_MADAD}) \times \log(\text{AGE})$$

זהו משתנה שהוא אינטראקציה כפלית בין משתנה לוג-גיל הדירה ולוג-שטח הדירה. משמעותו של משתנה זה הינו כיצד משתנה (גדלה/קטנה) ההשפעה של שטח הדירה במ"ר על מחיר הדירה כתוצאה משינוי בגיל הדירה בשנה אחת. יצוין כי ייתכן מצב שבו שני המקדמים של המשתנים גיל ושטח הדירה הם בעלי השפעה חיובית על מחיר הדירה, אך מקדמו של משתנה האינטראקציה יהיה שלילי, כלומר באינטראקציה עם גיל הדירה כיוון האפקט הראשי של שטח הדירה משתנה.

ב. אינטראקציה בין אזור ארנונה ובין מספר חדרים בדירה, KVUEZDUM: משתנים דיכוטומיים המציינים את השינוי של השפעת רמות שונות של מספר חדרים בדירה על המחיר כתוצאה ממעבר בין אזורי ארנונה שונים.

ג. אינטראקציה בין אזור ארנונה ושייכות הדירה לבניין רב קומות, EZORAVDUM: משתנים דיכוטומיים הבודקים את השינוי של השפעת אזור הדירה על המחיר במעבר בין הדירה הנמצאת בבניינים עד 4 קומות לדירה הנמצאת בבניין בן 5 קומות ומעלה.

4.2 בדיקת הנחות הרגסיה ליניארית

4.2.1 בדיקות נורמליות של משתנה LOGMEHIR

בחינת התפלגות משתנה LOGMEHIR בכל חודש על פני התקופה הנחקרת 1999-2004 בוצעה במסגרת בדיקת הנחות הרגסיה העיקריות: נורמליות של המשתנה המוסבר וקיום הנחת הומוסקדסטיות (שוויון שונות של טעויות לאורך התצפיות) ממצאי בדיקת נורמליות למשתנה LOGMEHIR מצביעים על כך כי רק ב-7 מקרים מתוך 67 השערת הנורמליות נדחת (ראה תוצאות מודגשות בטבלה 6) ברמת מובהקות של 0.01 ובשאר המקרים הנחת הנורמליות מתקבלת באותה רמת מובהקות.

טבלה 6: ערכי הסתברות של מבחני נורמליות (Shapiro-Wilk) למשתנה LOGMEHIR ע"פ חודש ושנה

שנה	1999	1999	1999	1999	1999	1999	1999	1999	1999	1999	1999	1999
חודש	1	2	3	4	5	6	7	8	9	10	11	12
p-value	0.0252	0.0420	0.0076	0.0720	0.0292	0.0658	0.0005	0.0997	0.0301	0.1945	0.0083	0.0302
שנה	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000	2000
חודש	1	2	3	4	5	6	7	8	9	10	11	12
p-value	0.0300	0.0595	0.0123	0.0002	0.6513	0.1857	0.7812	0.1370	0.2983	0.0660	0.5587	0.0722
שנה	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001	2001
חודש	1	2	3	4	5	6	7	8	9	10	11	12
p-value	0.0494	0.7696	0.9215	0.3835	0.1455	0.6646	0.1356	0.0393	0.5822	0.0139	0.0809	0.7721
שנה	2002	2002	2002	2002	2002	2002	2002	2002	2002	2002	2002	2002
חודש	1	2	3	4	5	6	7	8	9	10	11	12
p-value	0.0380	0.3477	0.3586	0.1902	0.1515	0.1262	0.6192	0.5212	0.2614	0.8985	0.6105	0.3522
שנה	2003	2003	2003	2003	2003	2003	2003	2003	2003	2003	2003	2003
חודש	1	2	3	4	5	6	7	8	9	10	11	12
p-value	0.0716	0.0970	0.1040	0.0211	0.0956	0.0005	0.6869	0.0020	0.0055	0.0880	0.0044	0.0567
שנה	2004	2004	2004	2004	2004	2004						
חודש	1	2	3	4	5	6						
p-value	0.0040	0.0050	0.0242	0.0102	0.0718	0.1962						

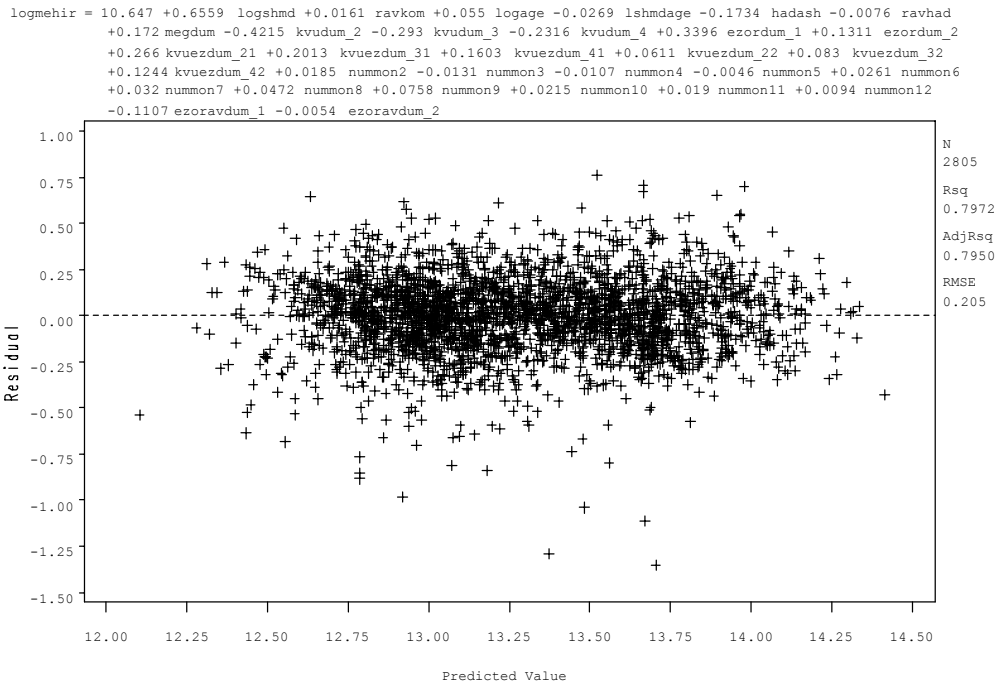
4.2.2 בדיקת שוויון שונויות של רעשים – הנחת הומוסקדסטיות

עבור כל המודלים של רגרסיה ליניארית בסדרה השנתית ובסדרה הדו חודשית בוצעו מבחני White(1980) הבודקים האם הרעשים ברגרסיות הם שווה שונויות ובלתי תלויים במשתנים המסבירים של המודלים. השערת האפס במבחן זה מדברת על קיום הומוסקדסטיות. בשליש מהמקרים במודלים השנתיים ההשערות של שוויון שונויות נדחו ברמת מובהקות של 0.05 אך הן לא נדחו ברמת מובהקות של 0.01. בשליש מהמקרים ההשערה לא נדחתה ברמת מובהקות 0.05. בשאר המקרים השערת האפס נדחתה ברמת מובהקות 0.01. בהתחשב בעובדה כי מספר התצפיות בכל מודל הוא מעל 2000 חשוב לציין כי ישנה הסתברות גבוהה לדחות את ההשערה אף כי לא נראה בתרשימים של רעשים כי ישנה מגמת פיזור כלשהי (ראה תרשימים 1, ו-2) במודלים הדו חודשיים למעט מקרים בודדים כל ההשערות בדבר שוויון שונויות לא נדחו ברמת מובהקות 0.05.

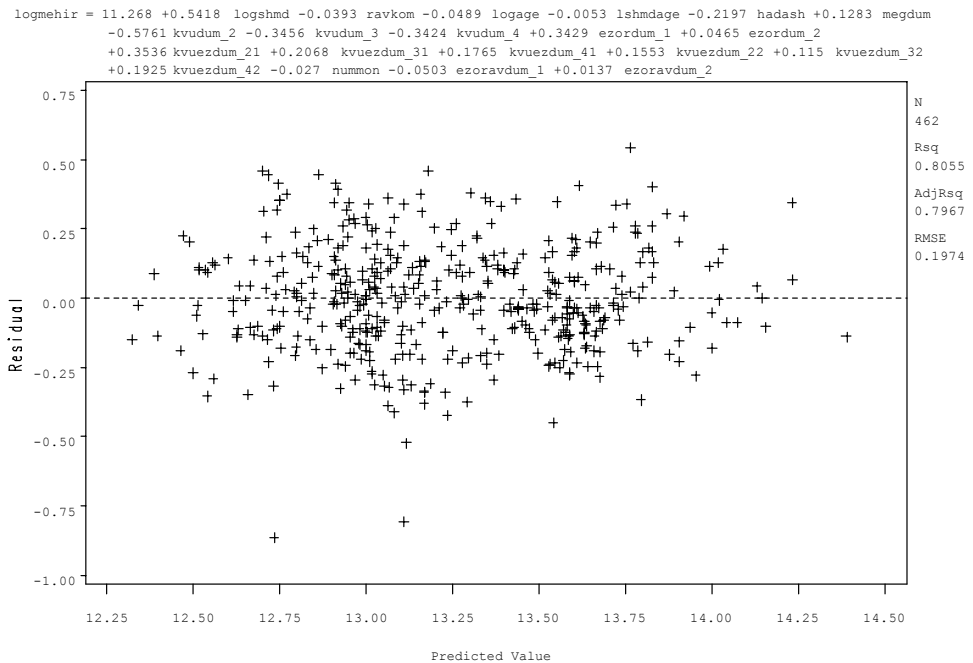
4.2.3 טיב ההתאמה

המבחנים למובהקות המודלים (כל המקדמים שווים ל-0 כנגד לפחות אחד מהם אינו שווה ל-0) הראו כי כל המודלים מובהקים. מקדם המתאם המרובה של שתי סדרות המודלים נע בין 0.73-0.82, דבר המצביע על טיב התאמה משביע רצון. רמות המובהקות המינימאליות (p-values) לדחיית ההשערה כי מקדם מסוים במודל שווה ל-0 מוצגות בטבלאות 4.2.1, 4.2.2, ו-4.4 (ראה נספח).

תרשים 1 : פיזור רעשים על פני תחזיות במודל שנתי של הרגסיה ליניארית (ינואר – דצמבר 1999)



תרשים 2 : פיזור רעשים כפונקציה על פני תחזיות במודל דו-חודשי של הרגסיה ליניארית (יולי 2000-יוני 2001)



4.3 חישוב מדד דיור

ע"פ הממצאים שנתקבלו מהרצת שתי סדרות מודלים, השנתית והדו-חודשית, נתקבלו ערכי שינויים במחירי דירות, כפי שמופיעים בשתי השורות האחרונות של הטבלאות 5.1.1, 5.1.2, 5.1.3 (המודל השנתי) וטבלאות 5.2.1, 5.2.2 (ראה נספח). השורה הלפני האחרונה מכילה ערכים ממוצעים של אומדני השינויים במחירי דירות במונחי פרופורציות שינוי ע"פ (3.2) ו-(3.3). בשורה האחרונה חושבו ערכי המדד במונחי מדד ייחוס 200 באופן הבא:

$$I_{01.1999} = 204.3$$

הינו 1999 ינואר

מתוך הנתון שנתקבל מגף מחירים לצרכן המדד לחודש ינואר 1999, המחושב באמצעות אומדן שינוי במחירי הדירות, שנתקבל כתוצאה ממיצוע אומדני שינויים, שנתקבלו ממודל רגרסיה הדונית לתקופה של 12 חודשים, כלומר:

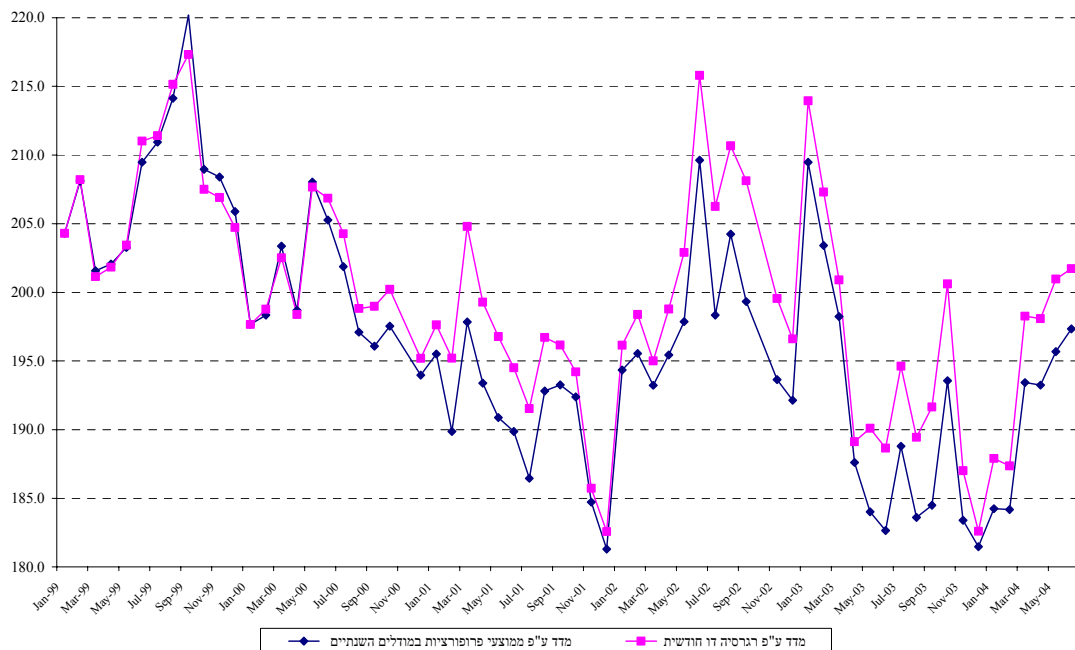
$$I_{02.1999}^{12} = \bar{R}_{02.1999} I_{01.1999}^{12}$$

ובאופן כללי, מחושב המדד לחודש $t+1$, $t = 1, K, T$ באופן הבא:

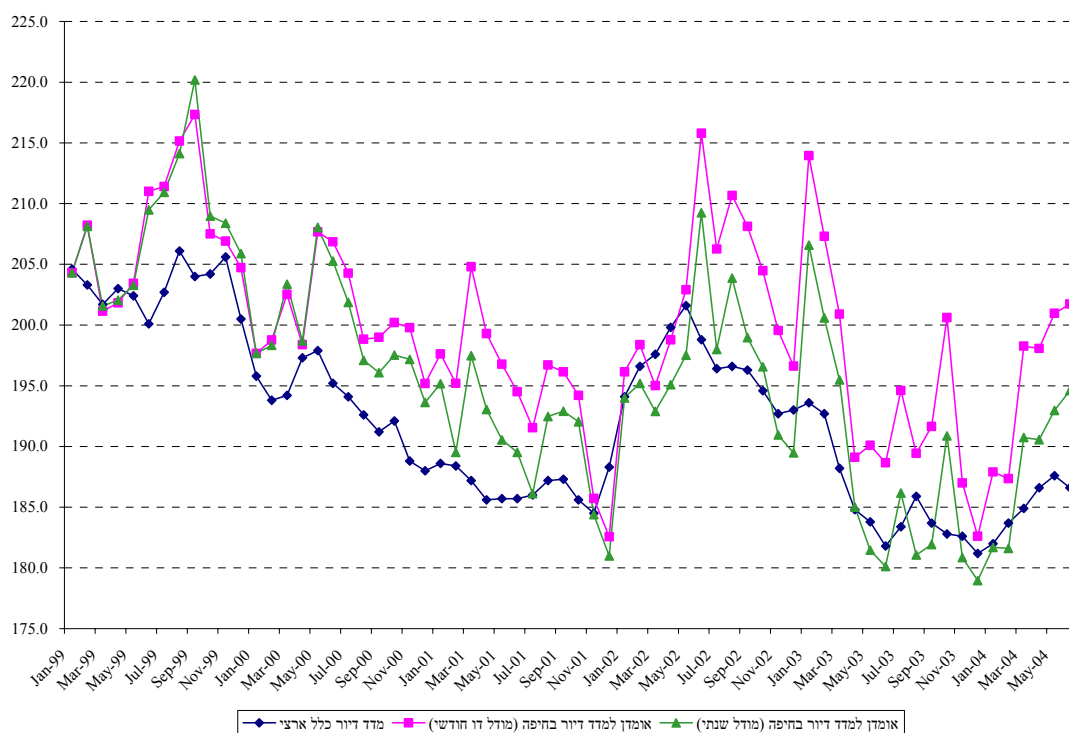
$$I_{t+1}^{\tau} = \bar{R}_{t+1} I_t^{\tau}$$

בתרשים 3 ניתן לראות את ההשוואה בין ערכי מדד הדיור ע"פ שני המודלים: $\tau = 12$ ו- $\tau = 2$. נראה כי אומדני מדד דיור המחושבים באמצעות המודל השנתי נמוכים יותר מאלו המחושבים באמצעות המודל הדו-חודשי.

תרשים 3: מדד מחירי דירות - חיפה - השוואה בין המודל השנתי והדו-חודשי



תרשים 4 : ביצועי מדד כלל ארצי למחירי דירות בהשוואה לאומדני מדד למחירי דירות בחיפה



תרשים 4 משווה בין שני האומדנים שהתקבלו בעבודה זו ומדד דיוור ארצי כפי שחושב באגף המחירים. נראה כי הגרפים של המדד הכלל הארצי והאומדנים של חיפה נעים ביחד בתאום על פני התקופה נחקרת של 1999-2004. כמו כן, ניתן לזהות קרבה גבוהה יותר של ביצועי המדד המחושב באמצעות המודל השנתי לביצועי מדד הכלל הארצי בהשוואה לפער גדול יותר שבין ביצועי המדד, המחושב באמצעות המודל הדו השנתי וביצועי מדד דיוור הכלל ארצי.

5 סיכום

עבודה זו עסקה בטיפול בקובץ נתונים של דירות בחיפה, השלמת נתונים חסרים, בדיקות לוגיות וזקיפות של נתונים בלתי תואמים וחסרים, ניתוח סטטיסטי של מחירי הדירות ומאפייניהן, הגדרת משתני עזר לצורכי שילובם במודלים, התאמת מודלים של רגרסיה ליניארית בשיטת החודש הנע והישוב אומדני שינויים במחירי דירות ע"פ מודל רגרסיה הדונית שנתית ודו-חודשית. יצוין, כי הממצאים שנתקבלו בעבודה זו יפים רק עבור הנתונים של הדירות בחיפה. על מנת להרחיב את המודל למימד כלל ארצי, יש לבחון מספר יישובים נוספים ולנתח עבורם את התנהגות מחירי הדירות ומאפייניהן.

