The Sensitivity of the Wage Equation to the Regression Method

Pavel Jelnov, Edna Schechtman and Shlomo Yitzhaki

The method of ordinary least squares (OLS) is the most frequently used method in regression analysis. In this method the sum of squared deviations from the assumed regression line is minimized. An alternative method, based on the Gini Mean Difference, is very similar in its properties to the method of least squares, but allows for additional inspections with respect to the robustness and the quality of the estimators obtained. The focus of this paper is on the question whether the method used to estimate the wage equation affects the estimates, where the methods being compared are the ordinary least squares and the Gini-based regressions. The wage equation is based on Klinov's (2004) specification and the data set is the income survey of 2004. Our main conclusion is that there is no significant difference between the estimates of the regression coefficients obtained by the Gini-based and OLS methods. However, the Gini based regression allows the user to reveal that the behavior of the coefficient of hours of work is not constant over the entire age range. More specifically, among women we identify a negative relationship between earnings and age that occurs between ages from mid-thirties to mid-forties, which we attribute to the effect of "back to work" due to raising children. This change in the sign of regression coefficient does not show up among men. This kind of analysis cannot be performed automatically when using ordinary least squares. Sensitivity analyses strengthened this finding about the difference between men and women.

רגישות משוואת השכר לשיטת הרגרסיה[1]

פאבל ז'לנוב, שלמה יצחקי ועדנה שכטמן

שיטת הרגרסיה המקובלת היא שיטת הריבועים הפחותים. בשיטה זו ממזערים את סכום ריבועי הסטיות מקו הרגרסיה. שיטה אחרת לרגרסיה, רגרסיה המבוססת על מדד ג'יני, דומה מאד בתכונותיה לשיטת הריבועים הפחותים, אך מאפשרת גם לבצע בדיקות נוספות לגבי טיבם ויציבותם של האומדנים המתקבלים. מאמר זה מתמקד בבדיקה האם שיטת הרגרסיה המשמשת באמידת משוואת שכר משפיעה על האומדנים, כאשר השיטות המושוות הן רגרסיה בשיטת הריבועים הפחותים ורגרסיה בשיטת ג'יני. ההשוואה מיושמת על משוואת שכר של מינצר המופעלת על נתוני סקר ההכנסות של הלשכה המרכזית לסטטיסטיקה (2004). משוואת השכר שאמדנו לקוחה מתוך עבודתה של רות קלינוב על הבדלי שכר בין נשים וגברים בישראל (2004). משוואה זו שונה במקצת מהמודל הקלאסי של מינצר. בפרט, משתנה נסיון התעסוקה בו מוחלף במשתנה גיל. מסקנתנו העיקרית היא כי אין הבדל מובהק בין מקדמי הרגרסיה המתקבלים בשיטת ג'יני לבין אלה המתקבלים בשיטת הריבועים הפחותים, לפחות כשמדובר במשוואת השכר. ואולם, שיטת ג'יני מאפשרת לגלות כי התנהגות משוואת השכר אינה אחידה על פני ציר הגיל. בפרט, אצל נשים ישנה תופעה של"חזרה לעבודה" מאמצע שנות השלושים עד אמצע שנות הארבעים לחייהן, המתבטאת בהתהפכות סימנו של מקדם הרגרסיה שקושר בין שכר לגיל, תופעה שאינה מתגלה אצל הגברים. תופעה זו אינה מתגלה בשיטת הריבועים הפחותים באופן מיידי. ניתוחי רגישות שביצענו חיזקו ממצא זה על ההבדל בין המינים.

[1] תודתנו לרות קלינוב ולשני קוראים אלמונים על הערות שאפשרו לנו לשפר את המאמר.

א. רקע

משוואת השכר היא המשוואה המסבירה את גובה שכרו של עובד שכיר כפונקציה של השכלתו, מאפייניו האישיים ותכונות מקום העבודה שלו. מאז המודל של הון אנושי שפותח על ידי מינצר (Mincer, 1974), בקר (Becker, 1967) ואחרים, הרי שלא נגזים אם נאמר שמשוואת השכר היא אחת המשוואות שנאמדה מספר הפעמים הרב ביותר. אולם הדיון סביב ההחזר השולי על ההשכלה עודנו נמשך. משוואתו של מינצר בצורתה הבסיסית מוגדרת:

$$\ln(E) = \beta_0 + rS + \beta_1 N + \beta_2 N^2 + \varepsilon \tag{1}$$

כאשר E = שכר, S = מספר שנות לימוד, N = מספר שנות ניסיון בעבודה (מינצר דיבר על משתנה זה כגיל פחות S+6), ε = שארית, שמניחים שהיא בעלת התפלגות הומוסקידסטית נורמלית בעלת תוחלת אפס. לכך ניתן להוסיף משתנים שונים.

הויכוח המתלווה לניתוח התוצאות, והמסבך את הדיון בו, הוא המידה שבה האומדים מוטים. ההטיה המיוחסת לאומדים תלויה במטרת האמידה ובשימושים שמשתמשים באומדנים, בצורת ניסוח הצורה הפונקציונלית של המשוואה, במשתנים המשמשים במשוואה ובקשרים ביניהם, בתקופת המדידה, במידה שבה לוקחים בחשבון את מסלול החיים והתמורות להשכלה ובגורמים רבים אחרים. תמצית מרוכזת של השימושים השונים והבעיות השונות ניתן למצוא במאמר הסקירה של Heckman, Lochner and Todd (2003) המאבחן את הבעיות השונות והשימושים השונים של המשוואה. אחד השימושים שהם מונים הוא אומדן התשואה לרמות ההשכלה השונות, אומדן שצריך לקחת בחשבון את עלויות רכישת ההשכלה, את התקופה שמתכננים לעבוד ואת הציפיות להתפתחות השכר על פני זמן, את אי הוודאות הטמונה בציפיות אלו, את העובדה שבדרך כלל אין הבחנה אמפירית ברורה בין הניסיון בעבודה והגיל, את המיסוי על השכר והשתנותו על פני זמן, את טיב ההשכלה הנרכשת וכמובן, צורת הקשר המתמטי המונחת בבסיס משוואת הרגרסיה. בנוסף, קיימות בעיות של אנדוגניות. תמצית הבעיה של אנדוגניות היא הטענה שבעלי היכולת האישית נוטים הן ללמוד יותר שנים והן להרוויח יותר, מה שגורם להערכת יתר באומדן השפעת ההשכלה על שכרו של אדם. חלק ניכר מהספרות המודרנית מנסה למצוא את משתנה העזר שינטרל מתאם חיובי זה.

במאמר זה לא נעסוק בבעיות הללו שהזכרנו לעיל ונתמקד באבחנה בסיסית יותר. עניינו הוא לבדוק באיזו מידה האמידה של משוואת השכר בשיטת הריבועים הפחותים, שזו שיטת הרגרסיה המקובלת, רגישה לשיטת הרגרסיה. כלומר, באיזו מידה האומדנים המתקבלים בשיטה זו אכן יציבים. לצורך ההשוואה נשתמש בשיטת רגרסיה אלטרנטיבית, שיטת ג'יני. במידה שיסתבר שהאומדנים המתקבלים אכן שונים ומושפעים משיטת הרגרסיה הרי שצריך להכניס ממד נוסף לביקורת על השיטה ולמבחנים המלווים את התוצאות והוא – מדוע נבחרה שיטת רגרסיה מסוימת. במחקר זה איננו מתעניינים במשמעויות הכלכליות אלא באיזו מידה שינוי בשיטת האמידה משנה את התוצאות המתקבלות. כדי לבדוק את ההשפעה של שיטת האמידה נשווה את האומדנים המתקבלים בשיטת הג'יני לאלו המתקבלים בשיטת הריבועים הפחותים. במידה שהאומדנים יציבים נסיק ששיטת האמידה אינה משפיעה על האומדנים המתקבלים. אולם במידה שהאומדנים משתנים במידה רבה הרי שיתעורר החשש ששיטת האמידה אכן משפיעה על האומדנים המתקבלים, ואם כך האומדנים אינם יציבים. בנוסף לאמידת הפרמטרים מאפשרת לנו שיטת הג'יני לעמוד על היציבות של המקדמים לאורך עקום הרגרסיה. מכאן שגם אם המקדמים של הרגרסיה הם שווים בין שתי השיטות ניתן לעמוד על ההשתנות של המקדם לאורך קו הרגרסיה, ואם כך ניתן לבדוק באיזו מידה ההנחה של מקדם רגרסיה קבוע אכן מייצגת את המשתמע מהנתונים. הבסיס להשוואה הוא מודל הלקוח מתוך מאמרה של רות קלינוב (2004). הנתונים במחקר הם של סקר הכנסות של הלשכה המרכזית לסטטיסטיקה (2004). כדאי להעיר שכל הבעיות הנדונות בספרות שהעלינו לעיל לגבי ההטיות בין תוצאות המודל לבין הפרשנות שמייחסים לתוצאות בשימושים השונים של המודל אינן נעלמות והן מתקיימות גם בשיטת הרגרסיה שאנו מציעים. ההבדל היחיד שייתכן שקיים הוא שהיות ושיטת הרגרסיה שאנו מציעים רגישה פחות לתצפיות קיצוניות הרי שייתכן שהשפעות הפתרונות לבעיות שהועלו בספרות תהיינה שונות בעוצמתן בשיטות הרגרסיה השונות. אולם זו השערה בלבד, ורק היישום יאפשר לדעת אם אכן השערה זו מתקיימת.

על מנת להתעלם מהבעיות והשימושים השונים הקיימים בספרות, התרכזנו בעבודה זו במבנה משוואת השכר של הגברים לעומת מבנה משוואת השכר של הנשים. בעיה זו פשוטה יחסית: במידה שאין למין השפעה על משוואת השכר הרי שנקבל מקדמים דומים אצל הגברים והנשים; במידה שלמין יש השפעה על צורת הפונקציה הרי שנקבל אומדנים שונים.

מטרת העבודה היא להציג את הרגרסיה המתבססת על מדד ג'יני כשיטה הדומה מחד גיסא בתכונותיה לשיטת הריבועים הפחותים ומאידך גיסא יש לה תכונות מספר שאינן קיימות בשיטת הריבועים הפחותים. במיוחד, שיטת הרגרסיה על פי ג'יני יוצרת אומדים עמידים (ROBUST) שרגישים פחות לתצפיות קיצוניות מאשר אומדני הריבועים הפחותים, וכן מאפשרים לאתר מקומות שבהם יש חשש וצורך לבדוק את המונוטוניות של מקדם הרגרסיה, הן של רגרסיית הג'יני והן של הריבועים הפחותים, לאורך עקום הרגרסיה. בניגוד למרבית שיטות הרגרסיה האחרות ובדומה לשיטת הריבועים הפחותים, כל האומדים יכולים להיכתב בצורה מפורשת ועל כן האינטואיציה שפותחה עבור הריבועים הפחותים תופשת לגבי אומדני רגרסיית הג'יני. המבנה של האומדים בשיטת הג'יני דומה במהותו למבנה בשיטת הריבועים הפחותים, להוציא ההבדלים הבאים: כל אימת שבמשוואה כל שהיא מופיעה שונות, היא מוחלפת במדד ג'יני בריבוע. כל אימת שיש מקדם של שונות משותפת – הוא מוחלף במקבילה של "ג'יני משותף", וכל אימת שיש מקדם מתאם של Pearson – הוא מוחלף במקדם המתאם על פי ג'יני. ההבדלים נובעים רק מכך שהתכונות של המדדים השונים בתנאים מסוימים עשויות להיות שונות. בנספח מס' 1 אנו מציגים דיון אינטואיטיבי על ההבדלים בין התכונות ומסבירים ממה הם נובעים. הקורא המעוניין בהתעמקות בתחום הטכני מופנה לספרות שבה התפרסמו ההבדלים הטכניים. (קיימות תכונות נוספות לרגרסיית הג'יני שאינן קיימות ברגרסיית הריבועים הפחותים ושתשמשנה אותנו במאמר זה, אולם הן תוצגנה רק לאחר הצגת ממצאים אמפיריים.) כמו כן, הקשרים בין המקדמים השונים ברגרסייה על פי ג'יני זהים בצורת הקשר ביניהם לקשרים הקיימים ברגרסיית הריבועים הפחותים: כך למשל ניתן לתאר את הקשר בין מקדמי הרגרסיה ברגרסיה הפשוטה למקדמי הרגרסייה ברגרסייה המרובה כזהים, ניתן להגדיר מקדמי מתאם חלקיים וכדומה. המאפיין את רגרסיית הג'יני הוא שבדומה לשיטת ריבועים פחותים היא ניתנת להצגה כפתרונן של משוואות לינאריות, כאשר המקדמים הם פרמטרים במשוואות. כמו כן, המקדמים גם בשיטת ג'יני וגם בשיטת הריבועים הפחותים ניתנים להצגה כממוצעים משוקללים של השיפועים בין תצפיות עוקבות של המשתנה הבלתי תלוי. לכן ניתן "לפרק" את הטווח של המשתנה הבלתי תלוי לחלקים ולנתח את התנהגות המקדם בכל חלק בנפרד. בנוסף, מקדם הרגרסיה בשיטת ג'יני ניתן להצגה בעזרת עקומות הנגזרות מעקומת הלורנץ – עקומות המאפשרות ללמוד על המונוטוניות של מקדם הרגרסיה לאורך המשתנה הבלתי תלוי.

כדאי להעיר שבמידה שהמודל המשמש את הרגרסיה נכון, והנכונות שאליה אנחנו מתייחסים אינה הנכונות במובן הכלכלי אלא הנכונות מבחינת הקיום של ההנחות הסטטיסטיות הגלויות והסמויות שעומדות מאחוריו, הרי שאין בהחלפת שיטת הרגרסיה כדי לשנות את תוצאות האמידה.[2] לעומת זאת, אם אחת ההנחות אינה מתקיימת הרי שתוצאות האמידה עלולות להשתנות כתוצאה מהחלפת השיטה.

מבנה העבודה הוא כדלהלן: בסעיף 2 נציג את הממצאים האמפיריים, סעיף 3 מציג כלי נוסף ייחודי לשיטת הרגרסיה של הג'יני המאפשר לאתר איזורים שבהם יש לבצע בדיקת מונוטוניות של קשרי הרגרסיה, בדיקה שתופסת גם לרגרסיית הג'יני וגם לרגרסיית ריבועים פחותים. סעיף 4 משתמש בכלי שמוצג בסעיף 3 לניתוח הממצאים. מהממצאים עולה שכאשר משווים את השכר של גברים לשכר של נשים ביחס לגיל, מתגלה שבעוד שהשכר של גברים עולה בתחילה עם הגיל ולאחר מכן יורד הרי שאצל הנשים מונוטוניות זו אינה מתקיימת, ובין הגילים של 33 עד 44 ניתן להבחין בירידה של השכר. אנחנו משערים שתוצאה זו נובעת מהקדשת זמן לגידול הילדים. בסעיף 5 אנו בודקים את הממצא על ידי בדיקת הקשר שבין הגיל לבין משתנים אחרים המעידים על השקעת זמן בעבודה: השתתפות בכח העבודה ושעות עבודה. הממצאים מחזקים את השערתנו שהתנהגות הנשים שונה מהתנהגות הגברים. סעיף 6 מכיל סיכום הממצאים.

ב. ממצאים אמפיריים

במאמרה על פערי השכר בין גברים לנשים בישראל מציגה קלינוב (2004) את אומדני רגרסיית השכר שבה המשתנה המוסבר הוא לוג השכר. המודל (להלן – "מודל קלינוב") שונה מהמודל הקלאסי של מינצר. במודל זה, במקום משתני נסיון תעסוקה ונסיון תעסוקה בריבוע מופיעים גיל וגיל בריבוע. הנימוק לכך הוא שלדעתה של קלינוב אין צורת המדידה של מינצר מתאימה למשק הישראלי. מינצר מודד ניסיון בעבודה על ידי הנוסחה גיל פחות שנות השכלה פחות קבוע (6 או 7 או 8). לטענת קלינוב הניסיון בעבודה עולה, ולא יורד, עם שנות ההשכלה, מאחר שהיציבות בעבודה היא פונקציה חיובית של שנות ההשכלה והיא חשובה יותר מכניסה מוקדמת לשוק העבודה. כמו-כן, תלמידים רבים במוסדות להשכלה גבוהה עובדים תוך כדי לימודיהם. אין עדיין מדידה מדויקת, ולכן קלינוב הניחה שרירותית

[2] בהנחות הסטטיסטיות הכוונה היא שאכן קיים קו רגרסיה (ולא עקום של רגרסיה), ושהשארית הבלתי מוסברת אכן בלתי תלויה סטטיסטית במשתנים הבלתי תלויים.

שהשפעת ההשכלה מתבטלת, והנסיון בעבודה הוא פונקציה של הגיל בלבד. המשתנים האחרים הם משתנים שמופיעים ושידוע שמשפיעים על שכר כגון מצב משפחתי, מספר הילדים, מוצא והענף הכלכלי. זאת מאחר שנמצא במחקרים שונים שנשים נשואות ובעלות ילדים משתכרות פחות מנשים יחידות, כנראה מאחר שבעלות המשפחה מקדישות לה יותר זמן. קלינוב רצתה לבדוק זאת, ולא מצאה עדות לכך. כמו-כן נבדקה ההשערה שהמגזר הציבורי מפלה נשים פחות מהמגזר הפרטי, מאחר שאין לו "דעות קדומות". קלינוב מצאה שההפך הוא הנכון. מאחר שהעניין שלנו הוא בבחינת התכונות הסטטיסטיות של המודל, אנו מאמצים את המודל של קלינוב כמות שהוא. בדומה למודל של קלינוב, האוכלוסייה מחולקת לשתי קבוצות – גברים ונשים.[3] רשימת המשתנים של מודל קלינוב ותיאורם מובאים בנספח 2.

הניתוח המובא בהמשך הוא בעל המבנה הבא: נסקור את תוצאות הרגרסיה בשיטת הריבועים הפחותים ותוצאות הרגרסיה בשיטת ג'יני, נראה את הדמיון והשוני וננסה להסביר אותם.

לפני שנתאר את הממצאים נציין הערות מספר:

א. שיטת ג'יני אינה מגדירה את הקבועים שדרכם תעבור הפונקציה. על-מנת שההשוואה לשיטת הריבועים הפחותים תהיה "נקייה" ככל היותר, נקבע כי הרגרסיה תעבור דרך ממוצעי המשתנים (כפי שנעשה בשיטת הריבועים הפחותים).

ב. טעויות התקן של אומדים בשיטת הג'יני מחושבות בשיטת ה- Jackknife ‏(Efron, 1982). שיטת ה- Jackknife היא שיטה עתירת חישוב. הרעיון, באופן כללי, הוא ליצור n ערכים של האומד על ידי לקיחת n מדגמים בגודל (n-1) מתוך המדגם המקורי מגודל n, כאשר המדגם ה- i מכיל את כל התצפיות חוץ מהתצפית ה- i. בהינתן n הערכים של האומד אפשר לחשב סטיית תקן.

ג. במשוואת הרגרסיה של קלינוב מופיעים שני משתנים שאחד מהם הוא טרנספורמציה מונוטונית של השני: גיל וגיל בריבוע. רגרסיה בשיטת ג'יני מתבססת על דירוגים של המשתנים הבלתי תלויים. על כן, משתנה שהוא טרנספורמציה מונוטונית של משתנה אחר יקבל אותם הדירוגים, ויגרום למטריצת הדירוגים של הנתונים להיות סינגולרית. לכן, כדי להריץ משוואות רגרסיה זהות בשתי השיטות הגדרנו ברגרסיה של ג'יני משתנה חדש המבוסס על גיל ועל גיל בריבוע.

נסמן את המקדמים של גיל וגיל בריבוע (A, A^2) בשיטת הריבועים הפחותים כ- b_2 , b_1 בהתאמה. במשוואת

[3] למעט שני משתני דמה המופיעים אצל קלינוב ואינם מופיעים אצלנו. קלינוב השתמשה בנתונים דו-שנתיים, עבור מספר זוגות של שנים. במחקר זה הנתונים הם חד-שנתיים, לקוחים מתוך סקר ההכנסות של למ"ס לשנת 2004. על כן, לא יופיע אצלנו משתנה דמה של השנה (אצל קלינוב זה 1 עבור 2000 ו- 0 עבור 1999), ומשתנה שמבטא את היקף התעסוקה בתקופות אחרות, שלא הופיע בסקר שבו השתמשנו.

רגרסיית ג'יני נגדיר משתנה חדש $A+\frac{b_2}{b_1}A^2$ ונשתמש בו כמשתנה בלתי תלוי. משמעות הדבר היא כי אנו מקבעים את היחס בין המקדמים שנקבל בג'יני עבור גיל וגיל בריבוע להיות b_2/b_1. לכן המקדם של גיל בריבוע בשיטת ג'יני הוא המקדם של גיל בשיטת ג'יני כפול יחס המקדמים b_2/b_1 משיטת הריבועים הפחותים, כלומר מקדם של גיל בריבוע בשיטת הריבועים הפחותים כפול היחס בין מקדמים של גיל בשתי השיטות. למעשה קיבענו את הקשר שבין A ו- A^2 שיהיה זהה בשתי הרגרסיות. על מנת להבהיר נראה זאת בעזרת משוואות:

נניח שאנו מעוניינים להכניס את הקשר: $y=a+bx+cx^2$ שאותו קיבלנו מרגרסיית הריבועים הפחותים לתוך רגרסיית הג'יני. נגדיר משתנה חדש Z כך שכאשר נכניס אותו לרגרסיה – הוא יכפה את הצורה הפונקציונלית שביקשנו – בצורה לינארית. כלומר נגדיר $z=x+\frac{c}{b}x^2$ כאשר המקדמים b ו- c ידועים לנו מריבועים פחותים. ברגרסיית הג'יני אומדים את המשוואה: $y=a'+d\,z$. במידה שמקדם הרגרסיה של X המתקבל בג'יני זהה למקדם הריבועים הפחותים – כלומר b=d, הפרשנות של המקדם זהה לפרשנות שניתנת בריבועים פחותים. במידה שהם שונים אזי רגרסיית הג'יני משנה גם את המקדם של הגורם הריבועי.

זאת כי $y=a'+d(x+\frac{c}{b}x^2)=a'+dx+\frac{dc}{b}x^2$

והמשמעות היא שכפינו על רגרסיית הג'יני צורה ריבועית והיא תיקנה לנו את המקדם של המשתנה בריבוע. כלומר, כפינו על הרגרסיה צורה פונקציונלית מסויימת על הקשר בין X ל X^2 ותחת מגבלה זו קיבלנו מקדמים.

לוח מס. 1 מציג את ההבדלים במקדמים שהתקבלו בשתי השיטות הן עבור גברים והן עבור נשים. הצגה מלאה של תוצאות הרגרסיה מופיעה בנספח 3.

המסקנה המתקבלת היא שאין הבדל של ממש בין מקדמים בשיטת ג'יני לאלו בשיטת הריבועים הפחותים. בשתי השיטות, על פי התוצאות המוצגות בנספח, כל המשתנים מובהקים, ואפילו המקדם של גיל בריבוע, שבשיטת ג'יני אינו מופיע ישירות ברגרסיה, זהה בשתי השיטות. כל ההפרשים להוציא אחד הם בסִפרה השלישית אחרי הנקודה העשרונית, ורק במקרה אחד – נשים, מקדם 45–49 שעות עבודה בשבוע, ההפרש הוא בסִפרה השנייה אחרי הנקודה העשרונית. המסקנה שלנו היא שהתוצאות למעשה זהות.

לוח 1: ההבדלים בין מקדמי הרגרסיה בין שיטת הריבועים הפחותים לשיטת הג'יני:

סקר הכנסות 2004*

נתוני שנת 2004	נשים (4796 תצפיות)	גברים (6354 תצפיות)
	b(G)-b(OLS)	b(G)-b(OLS)
מספר נפשות במשפחה	-0.0023	0.0009

	נשוי	0.0002	0.0076
	לא יהודי	0.0031	0.0059
	עולה	-0.0001	0.0017
מספר שנות לימוד	**עד שמונה**	-0.0009	0.0030
	תשע או עשר	0.0000	0.0049
	13 עד 15	0.0000	0.0062
	שש עשרה ומעלה	-0.0001	0.0042
ענף תעסוקה	**חקלאות, תעשייה ובנייה**	-0.0005	0.0068
	חשמל ומים, בנקאות, ביטוח ופיננסים, שירותים עסקיים	0.0002	0.0079
	מנהל ציבורי, חינוך, בריאות, שירותים חברתיים וקהילתיים	-0.0007	0.0085
מספר שעות עבודה בשבוע	**39–35**	0.0001	0.0097
	49–45	-0.0100	0.0098
	לפחות 50	-0.0002	0.0096
	גיל	0.0012	0.0016
	גיל בריבוע	0.0000	0.0000

*המשתנים "החסרים" – 12–11 שנות לימוד, 40–44 שעות עבודה בשבוע, הנם משתני בסיס. משתנה הבסיס עבור ענף תעסוקה כולל תחבורה, אחסנה ותקשורת, מסחר סיטוני וקמעוני ותיקונים, שירותי אירוח ואוכל ושירותים פרטיים למשק הבית. הסבר המשתנים האחרים מופיע בנספח 1. טבלה מלאה עם המקדמים וסטיות התקן מופיעה בנספח 2.

המסקנה המתקבלת היא שלגבי אמידת משוואת שכר אין בשיטה המוצעת על ידי ג'יני כדי לשנות את האומדנים.

ג. שימוש בעקומות ריכוז לזיהוי מונוטוניות

ניתן לפרש מקדמי רגרסיה המתקבלים בשיטת הריבועים הפחותים או בשיטת ג'יני כשיפוע ממוצע של השיפועים המתקבלים בין כל שתי תצפיות סמוכות, כאשר כל שיפוע משוקלל בפונקציה של המרחק בין התצפיות הסמוכות של המשתנה הבלתי תלוי (ראה Yitzhaki, 1996). שיפוע ממוצע יכול להסתיר את העובדה שייתכן שלאורך קטעים שלמים של עקום הרגרסיה סימני השיפוע שונים – כלומר בקטעים מסוימים השיפוע הוא חיובי ובקטעים אחרים השיפוע הוא שלילי. שיטת הרגרסיה של הג'יני מאפשרת לבצע בדיקה ויזואלית כך שניתן לבדוק אם השיפוע של עקום הרגרסיה מחליף סימן לאורך קטעים של עקום הרגרסיה. במידה שמצאנו החלפת סימן אזי החלפת הסימן תתקיים הן ברגרסיית הג'יני והן בריבועים הפחותים. מכאן שניתן להעזר בכלים המסופקים על ידי הג'יני לבדיקת יציבות מקדמי הריבועים הפחותים. בסעיף זה נציג את הצורה שבה נאתר האם קיימת החלפת סימן של שיפוע ברגרסיה ובסעיף 4 נדגים את השימוש בה.

הגדרה: שיפוע של קו רגרסיה ייקרא מונוטוני אם לא ניתן למצוא קטע לאורך עקום הרגרסיה שלו היינו מריצים רגרסיה על אותו קטע בלבד היינו מקבלים סימן אחר למקדם הרגרסיה.

לצורך הצגת הגישה נשתמש בהצגה אחרת של מדד ג'יני המתבססת על עקומת הלורנץ.[4] במקביל להצגת הג'יני בעזרת עקומת לורנץ ניתן להציג גם את השונות המשותפת על פי ג'יני בעזרת עקומה. עקומה זו נקראת עקומת הריכוז (ACC – Absolute Concentration Curve) וההבדל בינה לבין עקומת לורנץ הוא שבעוד שבעקומת לורנץ התצפיות מסודרות לאורך הציר האופקי באחוזונים מתוך האוכלוסייה ואילו לאורך הציר האנכי מוצג ההצבר של המשתנה באחוזים, הרי שלאורך עקומת הריכוז (ACC) הסידור של האחוזונים לאורך הציר האופקי הוא על פי משתנה אחד (למשל X), ואילו ההצבר של המשתנה על הציר האנכי הוא של משתנה אחר (למשל Y). קיים פער גדול בין הידע הנדרש לגזירת התכונות של העקומות לבין הידע הנדרש לשימוש בהן. לכן, בהמשך סעיף זה ניתן את עיקרי הגזירה של התכונות ואילו בתחילת הניתוח של הנתונים נציג את הידע הנדרש מהמשתמש בגישה. לכן, מי שאינו מעוניין בפיתוח יכול לדלג לניתוח הנתונים בסעיף הבא.

עקומות ACC ו- LMA

כאשר עובדים עם המקבילה לשונות המשותפת בשיטת ג'יני, ניתן להציג אותה כשטח הכלוא בין שתי עקומות: קו אי התלות (LOI – Line of Independence) ו- ACC (Absolute Concentration Curve) – יוגדרו בהמשך. עובדה זו מאפשרת לנתח את התרומה של קטעים של התחום שבו מוגדר המשתנה X לשונות המשותפת cov(Y,F(X)) ולעקוב אחר התנהגות מקדם הרגרסיה לאורך התחום של המשתנה המסביר. השימוש בעקומות יאפשר לנו לבדוק את המונוטוניות של קשרי הרגרסיה עבור המשתנים הבלתי תלויים גיל והשכלה במודל קלינוב.

העקומות מוגדרות באופן הבא (Yitzhaki, 2003):

ACC: על ציר ה- X מוצגת ההתפלגות המצטברת של X. על ציר ה- Y מוצג ההצבר של המשתנה Y הנלווה לתצפיות X. העקומה מתחילה ב- (0,0), וכאשר F(x) מגיע ל- 1, ACC מגיעה לממוצע של Y בכל האוכלוסייה.

באופן פורמלי:

[4] למדד ג'יני יש מעל לתריסר הצגות שונות המכבידות מחד גיסא על השימוש בו, אולם מאידך גיסא מאפשרות לאתר מאפיינים של הנתונים שקשה לקבלם בשיטות אחרות (ראה Yitzhaki, 1998).

$$ACC_{Y.X}(p) = \int_{-\infty}^{X_p} g(t)\ dF_X(t) \qquad (9)$$

כאשר $g(x)$ מוגדרת כתוחלת מותנית $g(x) = \mu_{Y.X} \equiv E_Y(Y \mid X = x)$ ואילו X_p מוגדר על-ידי

$$. \ p = \int_{-\infty}^{X_p} dF_X(t)$$

LOI זהו קו ישר המתחיל ב- $(0,0)$ ומסתיים ב - $(1,\mu_Y)$ (כפונקציה של $F(x)$ שתסומן ב p):

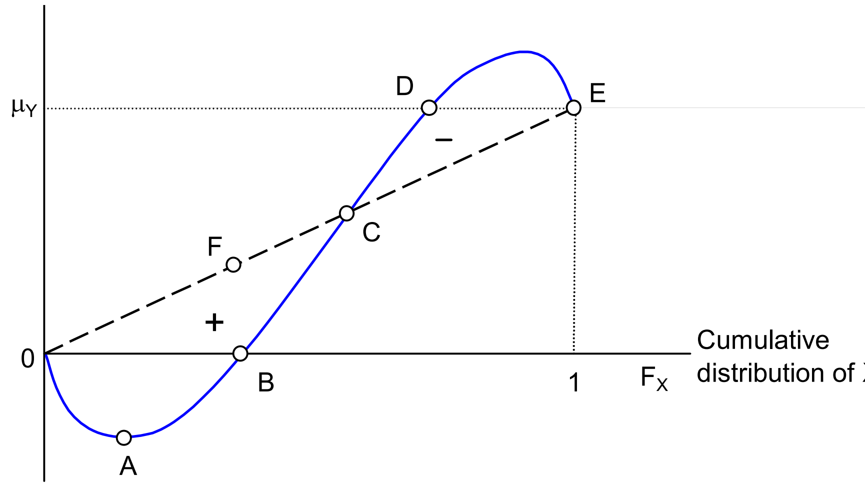
$$. \ LOI_{Y.X}(p) = \mu_Y . p$$

המשמעות של ה- LOI היא שאנו מציירים ACC תחת הנחה של אי תלות. במקרה זה, התוחלת המותנית של Y בתנאי X היא קבועה, ועל כן ה- ACC בהנחת אי התלות היא קו ישר. מכאן שהשטח הכלוא בין ה- LOI ל- ACC (שכפי שנטען בהמשך שווה בערכו לשונות משותפת על פי ג'יני) הוא הפער שבין ACC בפועל לבין ACC תיאורטית תחת הנחת אי תלות.

לצרכים שלנו, קל יותר להראות את עקומת הריכוז בציור – כפי שמוצגת בציור 1.

בציור 1 מוצגת עקומת ריכוז דמיונית (העקומה 0ABCDE) בין משתנה Y לבין משתנה X (קיימת גם עקומה בין משתנה X ל- Y שצורתה יכולה להיות אחרת ממה שמוצג בציור). האלכסון (0FCE) נקרא LOI (Line of Independence) שמשמעותו היא **שאילו** היו המשתנים Y ו- X בלתי תלויים אזי העקומה ACC הייתה מתלכדת עם הקו LOI. השטח הכלוא בין העקומה לבין ה- LOI הוא cov(Y, F(X)), כלומר המקבילה של השונות המשותפת על פי ג'יני. במידה שה- ACC נמצאת מתחת ל- LOI אזי התרומה לשונות המשותפת על פי ג'יני של אותו קטע היא חיובית.

ציור 1: עקומת הריכוז וקו אי התלות



לעומת זאת אם העקומה ACC נמצאת מעל ה- LOI, אזי התרומה לשונות משותפת על פי ג'יני היא שלילית. הסימן של השונות המשותפת על פי ג'יני יהיה תלוי בסכום השטחים הכלואים בין העקומה לקו.

בציור 1 יש קטע ראשון OC שבו הקו הנו מעל העקום ולכן ה- COVARIANCE הוא חיובי בתחום, ותחום שני CE שבו השטח הכלוא הוא שלילי ולכן ה- COVARIANCE בתחום הוא שלילי. סך כל השטח הכלוא שווה ל- cov(Y, F(X)).[5]

קיימות מספר תכונות של העקומה שחשובות לענייננו.

(א) במידה שמצאנו שהעקומה אינה חותכת את האלכסון אזי טרנספורמציה מונוטונית של המשתנה X לא תוכל לשנות את סימנו של cov(Y,F(X)), שהרי על הציר האופקי מופיע F(X) שאינו מושפע מטרנספורמציות מונוטוניות של X.

(ב) נניח שזרקנו תצפיות קיצוניות של X מהרגרסיה – כלומר תצפיות עם הערכים הגבוהים ביותר של X או תצפיות קיצוניות נמוכות של X, או קצצנו את המדגם על ידי זריקת תצפיות קיצוניות של X משני הצדדים, אזי עקומות ACC וה- LOI החדשות יתחילו ויסתיימו בנקודות הזריקה (מנורמלות לאפס ואחד), אולם צורת הקטע הפנימי, הקטע שנותר, לא תשתנה. לכן, אם העקומה הנה קעורה (או קמורה) לכל אורכה, אזי לא ניתן לשנות את סימן מקדם הרגרסיה על ידי זריקת תצפיות קיצוניות של X.

[5] מקרה פרטי של ACC הוא $A_{X,X}(p)$ (*Absolute Lorenz Curve*).

(ג) נניח שהיינו מציבים על ציר X את X עצמו במקום את F(X). ניתן להראות שבמקרה כזה היה השטח שבין הקו לעקומה שווה ל- cov(Y,X), כלומר היה שווה למונה של מקדם הרגרסיה בשיטת הריבועים הפחותים. הבעיה שהיתה מתעוררת היא שהקו LOI היה מפסיק להיות קו ישר[6] (כי צורתו תהיה תלויה בהתפלגות של X) והעקומה ACC תאבד מצורתה ונצטרך לחשב שטחים בין שתי עקומות. אולם מאחר שהזזה אופקית של עקומה לא תשנה את העובדה שעקומות נחתכות או אינן נחתכות, הממצאים שאנו מוצאים על פי (א) ו- (ב) תקפים גם לריבועים הפחותים (ראה Yitzhaki, 1990).

בטרם נפנה לניתוח הממצאים נוסיף עוד הערה אחת. על מנת לפשט את הדיון, העקומה שתוצג בניתוח האמפירי היא עקומה המתארת את ההפרש בין הקו LOI ל- ACC ונקרא לה עקומת ה- LMA (Line Minus ACC), ואת הקורא המתעניין במלוא תכונותיה אנו מפנים ל- Yitzhaki, 2003; Yitzhaki and Schechtman, 2004.

ד. ממצאים לגבי מונוטוניות של קשרים

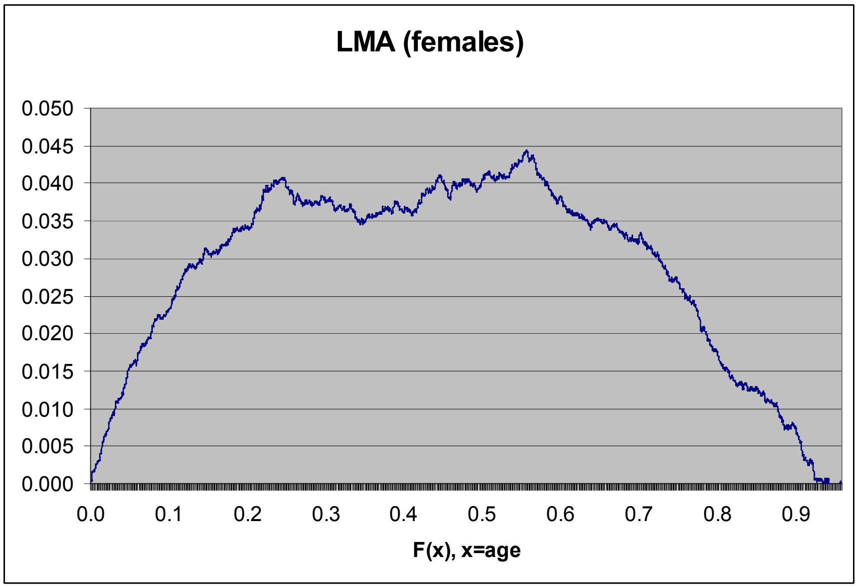
בסעיף 3 הראינו שניתן להציג את השונות המשותפת על פי ג'יני כשטח שנמצא בין עקומת LMA לציר האופקי. שטח חיובי משמעותו מקדם רגרסיה חיובי באותו קטע, ושטח שלילי מעיד על מקדם שלילי. להלן נראה כיצד ניתן להשתמש בעובדה זו בדוגמת גיל והשכלה כמשתנים בלתי תלויים.

גיל כמשתנה בלתי תלוי

ציור 2 מציג את עקומת ה- LMA של הנשים ואילו ציור 4 מציג את עקומת ה- LMA של הגברים. על הציר האופקי מוצגת ההתפלגות המצטברת של הגיל, ואילו הציר האנכי מייצג את הסטייה המצטברת של לוג השכר ממצב של אי תלות בגיל. טווח הגילים של הנשים והגברים הוא 25–70. ניתן לראות מהעקומה עבור נשים שבין אחוזון 25 לבין אחוזון 55 העקומה הופכת לקמורה, מכאן שבתחום גילים זה מקדם הרגרסיה הן של הריבועים הפחותים, הן של הג'יני והן מקדם רגרסיה של כל טרנספורמציה מונוטונית של גיל יראו מקדם רגרסיה שלילי. כלומר, בתחום גילים זה השכר יורד עם עליית הגיל. בדיקה בנתונים מראה שמדובר בתחום הגילים 33–44. (קיימת גם קמירות קטנה בתחומים אחרים, כגון בין אחוזון 55 ל- 70, אבל היא לאורך תחום קטן על כן יתכן שמקורה בטעויות דגימה.)

[6] להוציא מקרה שבו X מתפלג התפלגות אחידה, שאז ישאר LOI קו ישר.

ציור 2: עקומת LMA בין הלוגריתם של השכר לגיל – נשים



המסקנה מהגרף היא כי בטווח הגילים 33–44, בגלל קמירות העקומה, מקדם הרגרסיה של לוג השכר על גיל אצל נשים הוא שלילי! ניתן לבדוק את הטיעון על-ידי הרצת רגרסיה (בשתי השיטות) רק על טווח גילים זה. המודל הפעם יכלול רק את הגיל כמשתנה בלתי תלוי של ln(E). בטבלה הבאה מובאות תוצאות הרגרסיות:

לוח 2: מקדם הרגרסיה של גיל נשים כמשתנה בלתי תלוי יחיד בטווח הגילים 33–44

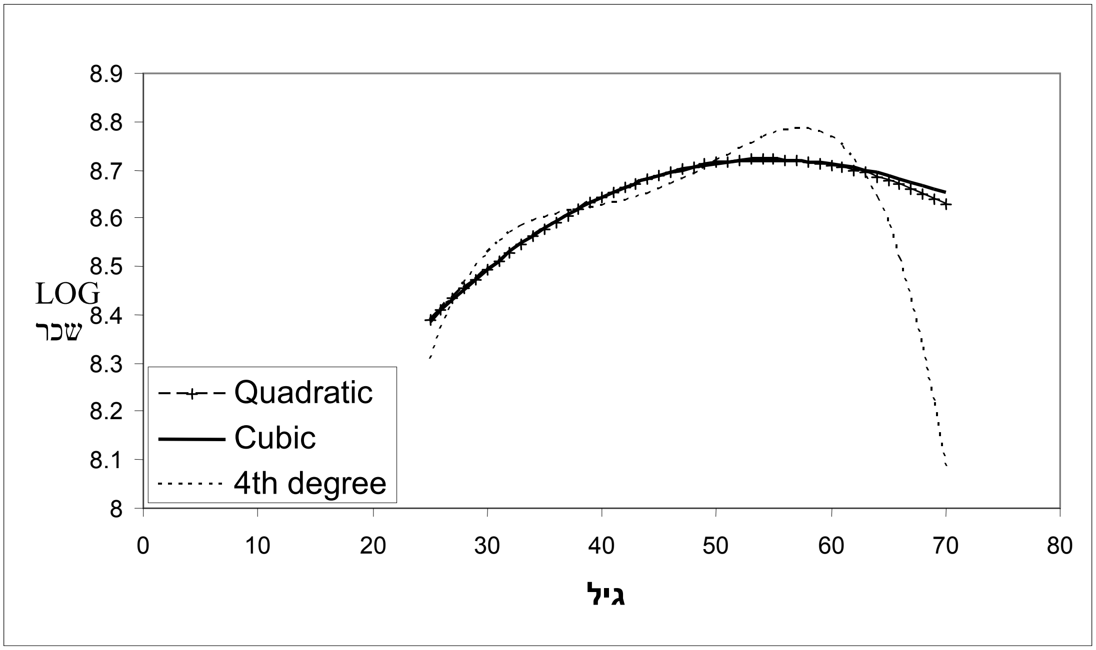
שיטת רגרסיה	B(age)	Std(b)
Gini	-0.0085	0.0050
OLS	-0.0085	0.0050

אכן, קיבלנו מקדם שלילי עבור טווח הגילים שבו עקומת LMA קמורה, ותוצאות זהות בשתי שיטות הרגרסיה. ללוח 2 כדאי להעיר שתי הערות: הראשונה היא שגם כאשר נרחיב את טווח הגילים נמשיך לקבל מקדמים שליליים, זאת מאחר שהוספת תחום קטן יחסית של מקדמים חיוביים עדיין יכולה להשאיר את מקדם הרגרסיה הכולל את התחום הנוסף שיהיה שלילי. אולם במקרה זה, מאחר שאנו עוסקים בסיכום שטחים חיוביים ושליליים אין לנו ביטחון לגבי

התוצאה שתתקבל אם נבצע טרנספורמציה מונוטונית של משתנה הגיל או לגבי הסימן של מקדם הרגרסיה אם נעבור מרגרסיית ג'יני לרגרסיית ריבועים פחותים. ההסבר לממצא זה נעוץ כפי הנראה בתופעה של יציאת נשים מכוח העבודה עקב גידול הילדים ולחזרתן אליו לאחר שהילדים גדלו (התופעה של יציאת הנשים מכוח העבודה וחזרתן אליו בתום תקופת גידול הילדים מתועדת במאמר של Ben-Porath and Gronau (1985). הערה שנייה היא שממצא זה של יציאת נשים מכוח העבודה וחזרתן אליו יכול לשמש אותנו להערכת "השנים האבודות" מבחינת ההשתתפות בשוק העבודה של הנשים, הערכה שגם יכולה לאפשר לנו לכמת את השפעת "השנים האבודות" על הפערים בשכר בין נשים וגברים. Berman and Klinov (1997) מתעדים תופעה של יציאה מכוח העבודה של גברים. כימות של השפעת ההתנהגות השונה של גברים ונשים על פני טווח הגילים יכול לתת הסבר מסוים להפרש השכר המתגלה בחתך רוחב בין גברים ונשים – שהרי אצל הגברים מתועדת יציאה מכוח העבודה של המשתכרים שכר נמוך ואילו אצל הנשים מתועדת יציאה וחזרה לכוח העבודה, יציאה שמדרך הטבע כרוכה בירידה בניסיון הנצבר ובשכר. הערכה מסוג זה היא מעבר לתחום הדיון של העבודה הזו כי היא תחייב מערכת אומדנים מסוג אחר. מה שביקשנו להראות בסעיף זה הוא השתנות הסימן של מקדם הרגרסיה בתחום מסוים של המשתנה הבלתי תלוי, ואת זאת הראינו בעזרת עקומת ה- LMA המאפשרת ניתוח של התרומות של הקטעים השונים של המשתנה הבלתי תלוי למקדם הרגרסיה, הן של הג'יני והן של הריבועים הפחותים.

אחד המבקרים האנונימיים של העבודה הציע לבדוק האם פולינום מדרגה 3 בשיטת OLS, כאשר המשתנה המסביר הוא גיל, יזהה את היפוך הסימן ש"נחשף" בשיטת ג'יני. למען השלמות הותאמו שלושה פולינומים – מדרגות 2, 3 ו- 4. התוצאות מוצגות בציור 3. כפי שאפשר לראות, התוצאות שונות זו מזו, ואי לכך קיימת סכנה שבחירת הדרגה של הפולינום תקבע את המסקנות (בעוד שבשיטת ג'יני התוצאות נקבעות באופן אוטומטי וללא התערבות החוקר). כמו כן ניתן לראות שפולינום מדרגה 3 אינו מזהה את התופעה המדוברת.

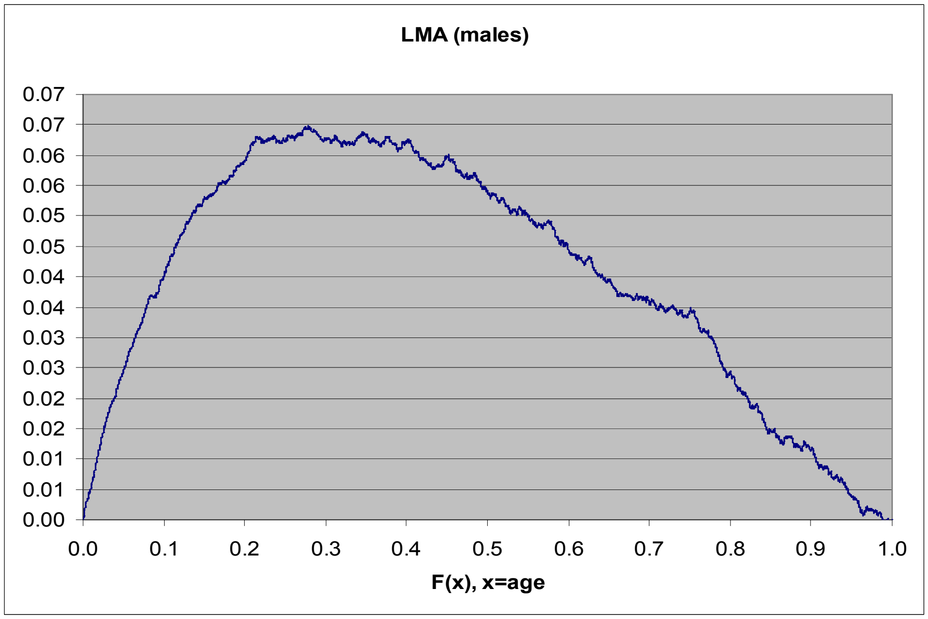
ציור 3: נשים בנות 25–70 – לוג שכר כפונקציה של גיל



הסבר לממצא זה ניתן למצוא ב- Yitzhaki (1996) שהראה שמקדם הרגרסיה הוא ממוצע משוקלל של שיפועים סמוכים, והשקלול נקבע על ידי התפלגות המשתנה הבלתי תלוי. במקרה של משוואה לינארית הרי שהשקלול נקבע על ידי מבנה השונות של המשתנה הבלתי תלוי, ובמקרה של משוואה ריבועית הרי שמדובר בשונות של ריבוע הגיל (כלומר בהתפלגות המומנט הרביעי של הגיל) ועל כן הדגשים הניתנים לשיפוע של לוג השכר כפונקציה של הגיל משתנים בצורה דרסטית. כדאי להעיר שהציורים מראים את מה שהיינו מצפים מהתיאוריה של הון אנושי, שבשנים הראשונות להצטרפות לשוק העבודה ישנה צבירת ניסיון ואילו בשנים המאוחרות יותר מתגברת השפעתו של הפחת על ההון האנושי.

נעבור עתה לגברים. ציור 4 מציג את עקומת LMA של הגברים: העקומה קמורה לכל אורכה! משמעות הדבר היא שסימן מקדם הרגרסיה בין לוג שכר לגיל אינו משתנה אצל הגברים, וזאת בניגוד למתרחש אצל הנשים.

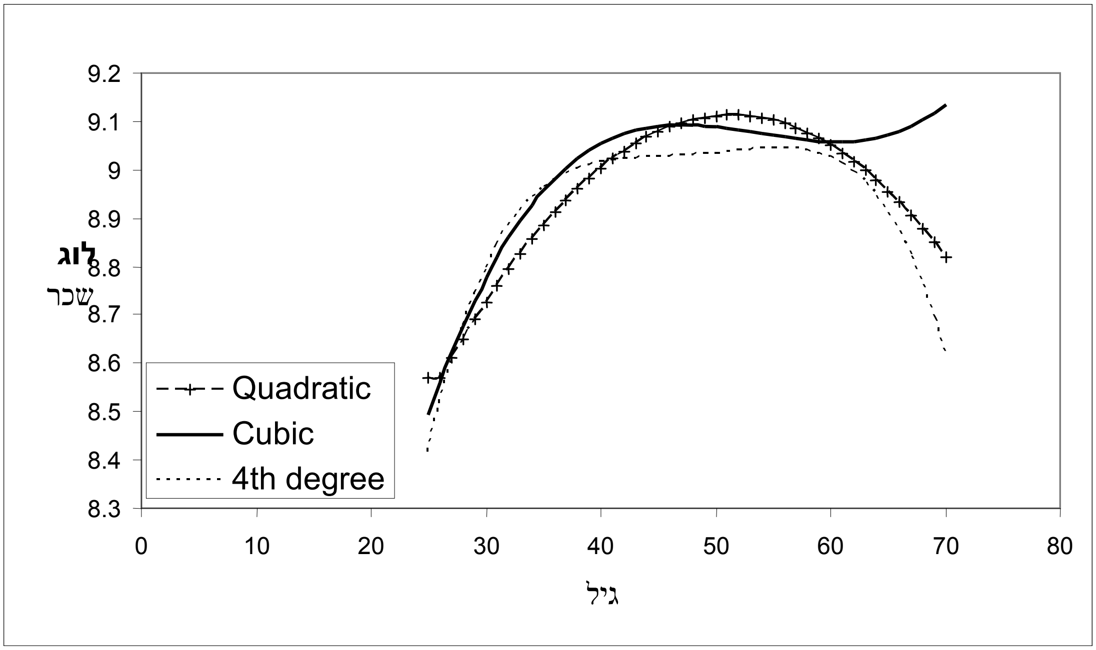
ציור 4 : עקומת LMA של לוג שכר מול גיל, גברים



ציורים מהצורה של ציור 2 ו- 4 יכולים גם לסייע בהבנת המרכיבים של מקדמי הרגרסיה בשיטת הריבועים הפחותים. כפי שהראינו לעיל – כאשר מוצאים החלפת סימן של מקדם רגרסיה – החלפת סימן זו תופסת גם לגבי מקדמי הריבועים הפחותים. כדי להתאים את הציור לחקירת מקדמי הרגרסיה של הריבועים הפחותים כל שעלינו לעשות הוא להחליף את הציר האופקי כך שבמקום שעל הציר האופקי יהיה F(X) מציינים על הציר האופקי את המשתנה X (מותר לעשות זאת כי הפונקציה F היא פונקציה מונוטונית עולה של X). כשעושים החלפת ציר מסוג זה אזי הקו הישר של LOI מתחלף בעקומה שצורתה תלויה בהתפלגות של X, והקלות בהסקת מסקנות מהציור נעלמת. אולם עדיין ניתן להשתמש בציור להבנת התרומה של חלקי ההתפלגות למונה של מקדם הריבועים הפחותים.

לצורך השלמת התמונה אנו מציגים בציור 5 את הצורות הפונקציונליות המתקבלות כאשר מכניסים למשוואת הרגרסיה של לוג שכר כפונקציה של גיל, גיל בריבוע, גיל בשלישית וגיל ברביעית. כמו במקרה של נשים הצורות שמתקבלות במשוואה ריבועית הן בהתאם לצפי המניח שפועלים שני כוחות – צבירת ניסיון מחד גיסא ופחת של הון אנושי מאידך גיסא, ואילו כאשר מסתכלים על משוואה מחזקה שלישית מקבלים תמונה שונה. ייתכן שהדבר נובע מהטיית מדגם – כלומר מי שממשיך לעבוד בגילים המאוחרים שכרו גבוה יחסית.

ציור 5: גברים בני 25–70 – לוג שכר כפונקציה של גיל

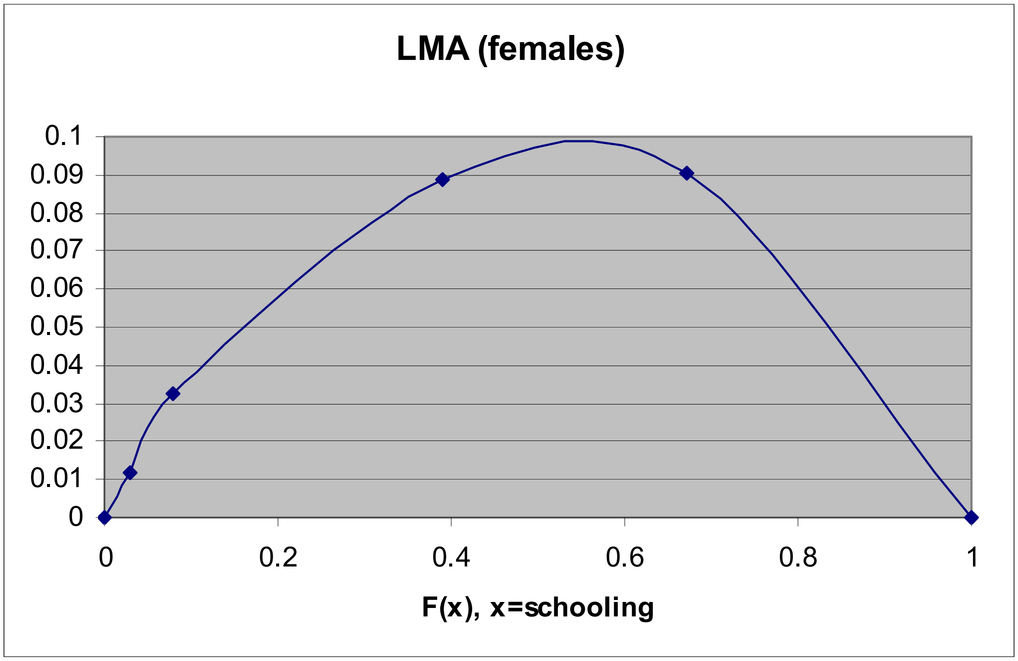


השכלה כמשתנה מסביר

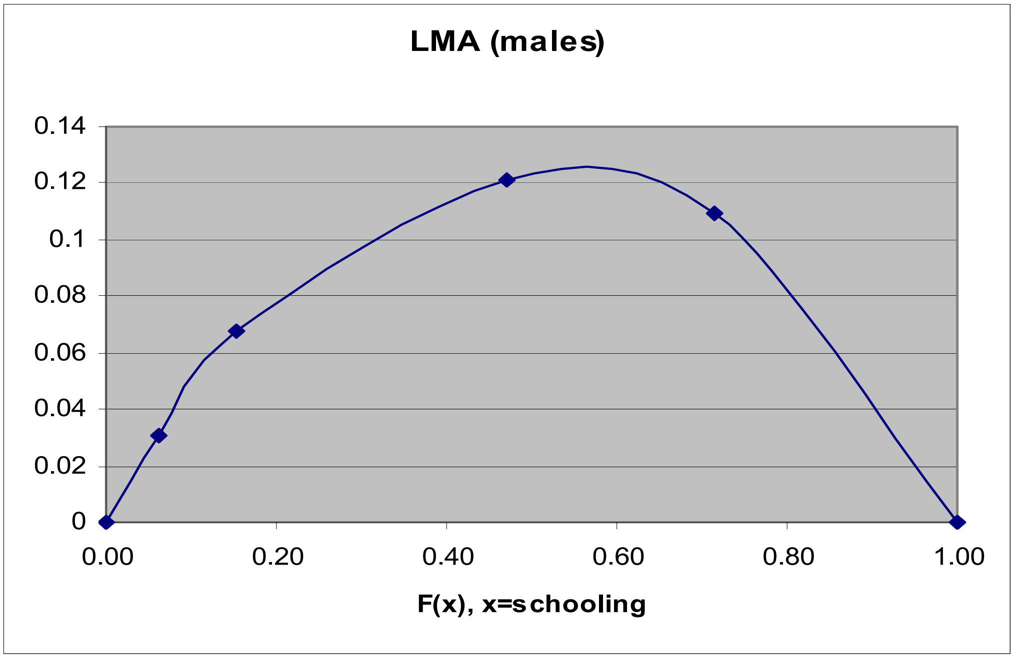
במודל של קלינוב אין משתנה רציף של שנות לימוד, אלא ההשכלה מחולקת ל- 5 קבוצות (4 משתני דמה + בסיס): 0–8 שנות לימוד, 9–10 שנות לימוד, 11–12 שנות לימוד (בסיס), 13–15 שנות לימוד, 16 שנות לימוד או יותר.

על מנת לבנות את עקומת LMA עבור ההשכלה, לקחנו את מספר התצפיות שעבורן ערך משתנה הדמה הוא 1 בכל אחת מהקבוצות. עבור קבוצת הבסיס מספר התצפיות הוא התצפיות שעבורן ערך כל משתני הדמה הוא 0. אחר-כך בנינו את ההתפלגות המצטברת F(x) לפי מספר התצפיות בכל קבוצה, ואז את פונקציות LOI ו- ACC. לבסוף חישבנו את LMA. להלן עקומות LMA שמתקבלות:

ציור 6: עקומת LMA של שכר לעומת השכלה, נשים


LMA (females)
0.1
0.09
0.08
0.07
0.06
0.05
0.04
0.03
0.02
0.01
0
0
0.2
0.4
0.6
0.8
1
F(x), x=schooling

ציור 7: עקומת LMA של שכר לעומת השכלה, גברים



הנקודות על העקומה מציינות כל אחת מהקבוצות. עבור שתי קבוצות ההשכלה הנמוכות ביותר יש מעט תצפיות, יחסית לאחרות, לכן הנקודות שלהן קרובות.

כצפוי, עבור שני המשתנים LMA חיובי לכל F(x), זאת אומרת, הקשר בין השכלה לשכר הוא מונוטוני לאורך כל התחום לנשים ולגברים. אולם חשוב לציין שייתכן שההקבצה של המשתנה השכלה העלימה תחומים של מתאם שלילי שבין השכלה ושכר.

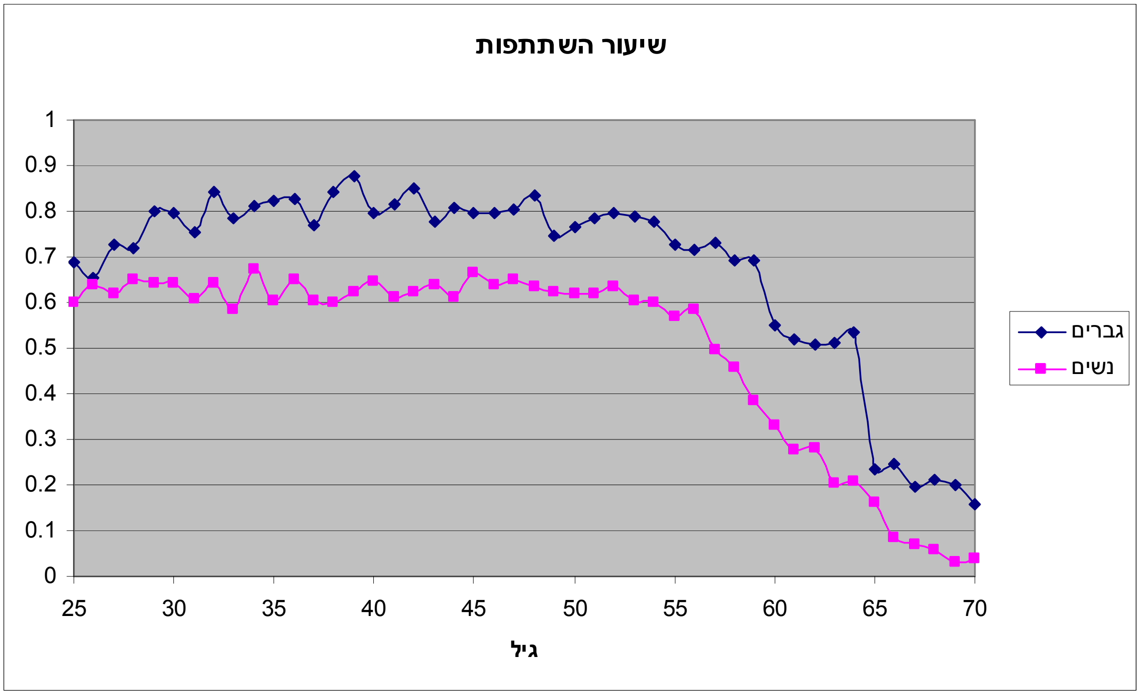
ניתן להסיק שגם כשמדובר במשתנה דמה, בתנאי שהתכונה שמדובר בה מיוצגת על ידי כמה משתני דמה ושיש ביניהם סדר אורדינלי כך שיש לנו נקודות מספר להצגה, עקומת LMA מאפשרת לחלץ מידע על השונות המשותפת.

ה. משתנים נוספים

הממצא העיקרי שלנו בסעיפים הקודמים היה שכאשר מסתכלים על מקדמי הרגרסיה של משוואת השכר אין הבדלים מהותיים בין תוצאות הרגרסיה המתקבלות בעזרת רגרסיית הג'יני לבין רגרסיה בשיטת הריבועים הפחותים, אולם כאשר מסתכלים בעזרת האמצעים הנוספים ששיטת הג'יני מעמידה לרשות החוקר, ואמצעים אלו הם עקומות הריכוז המאפשרות לראות את השינויים במקדם הרגרסיה לאורכו של המשתנה הבלתי תלוי, אנחנו מגלים הבדלים מהותיים בקשר של שכר לגיל בין גברים ונשים. בעוד ששכר הגברים מתנהג בצורה סדירה – עולה עם הגיל ולאחר מכן יורד, הרי ששכר הנשים עולה ויורד ולאחר מכן שוב עולה ויורד. ההסבר האינטואיטיבי הוא שהגורם לכך הנו תקופת גידול הילדים שבה, יש לשער, מוטל עומס גידול הילדים בין המינים בצורה לא שוויונית ועל כן מחייב נשים, יותר מאשר גברים, להקטין את שעות העבודה ואת הנכונות להשקיע בעבודה. כדי לבחון הסבר אינטואיטיבי זה, נסתכל על ההבדל בין המינים במשתנים קשורים שאינם מופיעים במשוואת השכר שהצגנו. המשתנה הראשון שנסתכל עליו הוא שיעור ההשתתפות בכוח העבודה כפונקציה של גיל. זאת מאחר שאם ההסבר הוא נכון, וחלק מהנשים מקדישות יותר זמן מגברים לגידול הילדים, הרי שיש לצפות שהדבר יתבטא גם בהקטנת שיעור ההשתתפות בכוח העבודה וגם בהקטנת שעות העבודה.

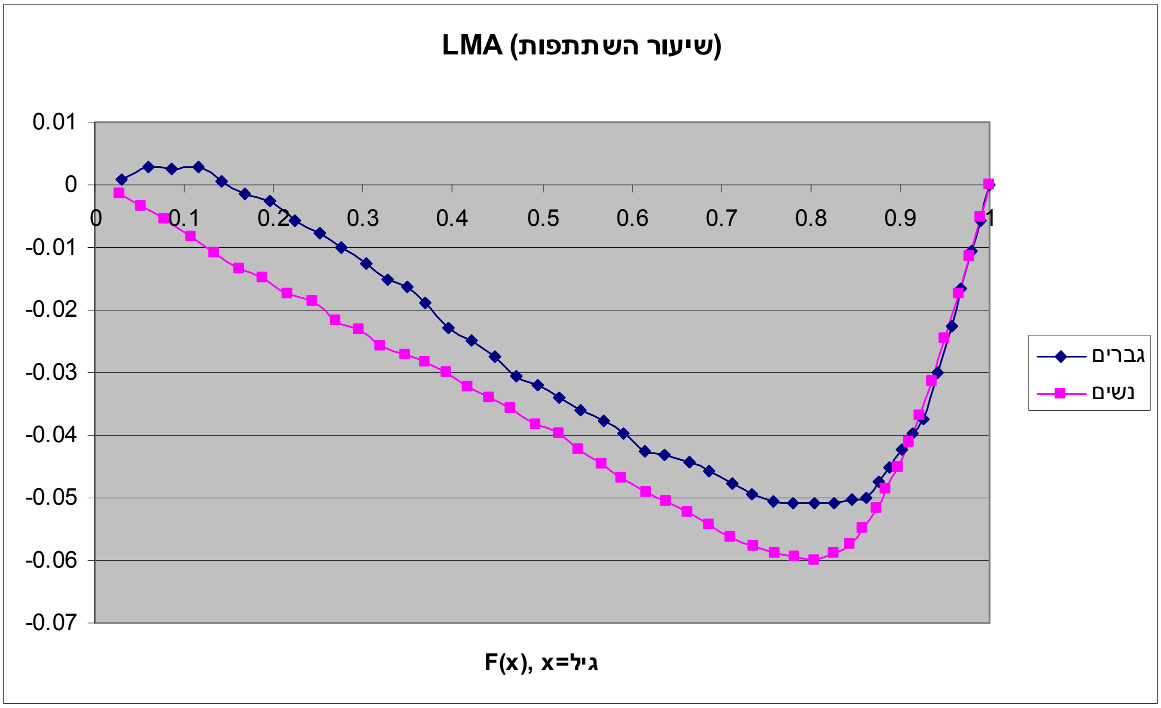
ציור 8 מציג את שיעור ההשתתפות של גברים ונשים כפונקציה של גיל. שיעור ההשתתפות הוגדר כאחוז הפרטים שעבדו כמות חיובית של שעות. מהגרף ניתן לראות ששיעור ההשתתפות של גברים עולה בגילים הצעירים מסדר גודל של 70 אחוז בגילים שמתחת לשלושים עד לסדר גודל של כמעט 90 אחוזים בגיל ארבעים ולאחר מכן מתחיל לרדת עם ירידה חדה בגיל שישים וחמש, שנת הפרישה לגמלאות בשנת הסקר. לעומת זאת, אצל הנשים איננו רואים עליה משמעותית בשיעור ההשתתפות בהתאם לגיל, ואם כך קיים שיעור השתתפות קבוע בשנות העבודה הראשונות ולאחר מכן אנו רואים הקטנה בשיעור ההשתתפות. מכאן שהירידה בשכר שהבחנו בה אצל הנשים בשנות השלושים אינה נסתרת כאשר מסתכלים על שיעור ההשתתפות בכוח העבודה, אם כי הצורה משתנה: בעוד שאצל הגברים מזהים עליה בשיעור ההשתתפות הרי שאצל הנשים אין מזהים עליה.

ציור 8: שיעור ההשתתפות בכוח העבודה של גברים ונשים בהתאם לגיל



ציור 9 מציג את עקומת ה- LMA של הנתונים שהוצגו בציור 8, זאת על מנת לאפשר לנו לראות את המעבר מציור אחד לשני. אצל הגברים, ניתן לזהות מתאם חיובי שיתקבל בהתבסס על ארבעים אחוז מהתצפיות של הגילים הנמוכים ומתאם שלילי שמתקבל מ- 60 אחוז התצפיות של הגילאים הגבוהים – אולם בסך הכל אנו מגלים שמקדם הרגרסיה של שיעור ההשתתפות הוא שלילי. לעומת זאת, אצל הנשים מהווה עקומת ה- LMA בשנים הנמוכות קו ישר, מה שמעיד על שיעור השתתפות קבוע ולאחר מכן ירידה בשיעור ההשתתפות. אצל הנשים הקשר של שיעור ההשתתפות כפונקציה של גיל הוא קשר מונוטוני שלילי, ולא תיתכן טרנספורמציה מונוטונית שתשנה סימנו של קשר זה. לעומת זאת, אצל הגברים אם נקטין את ערכם של הגילים הגבוהים הרי שנוכל למצוא טרנספורמציה מונוטונית שתגרום שהקשר יהיה חיובי, אם כי, מכיון שהתחום עם המתאם החיובי הוא קטן, הרי שנדרשת טרנספורמציה אגרסיבית מאד כדי להשיג שינוי סימן של מקדם הרגרסיה.

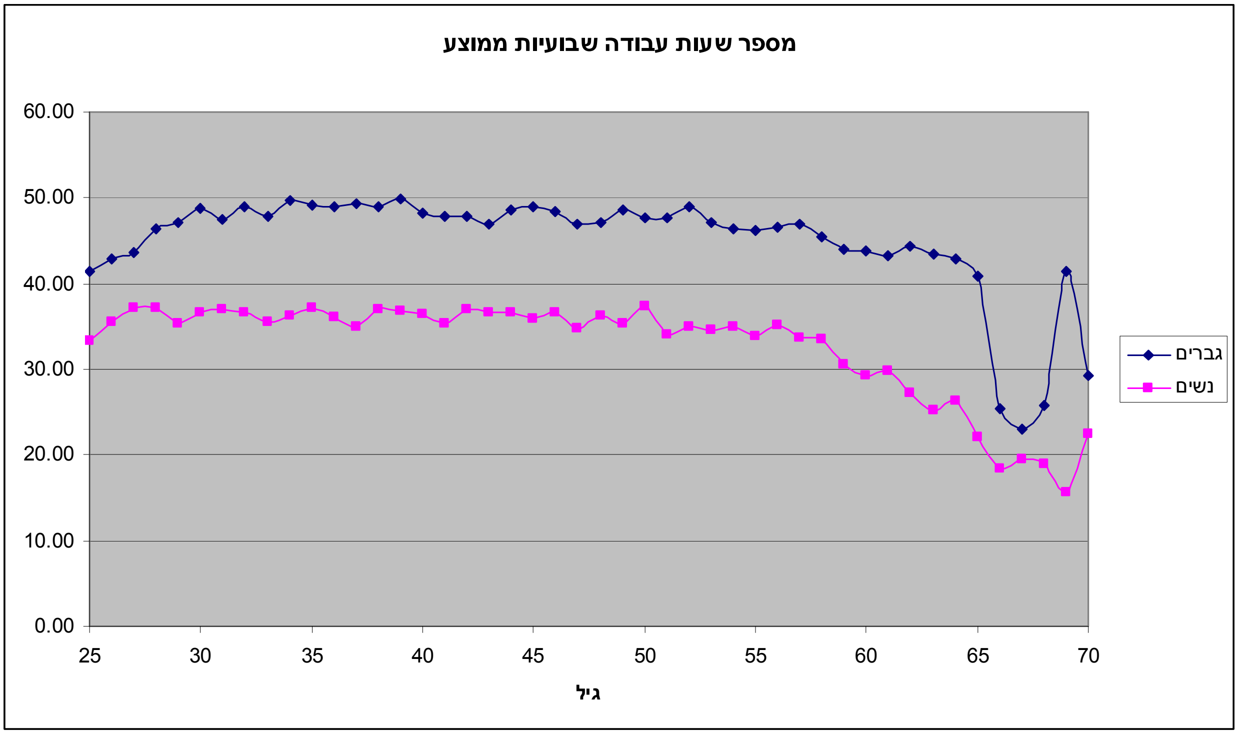
ציור 9: עקומת LMA של שיעור השתתפות של גברים ונשים



הגדרת ההשתתפות בכוח העבודה בציור 8 ובציור 9 סטתה מהגדרת ההשתתפות שבמודל של קלינוב, ושעליו התבססנו במאמר זה. במודל של קלינוב הוגבלו התצפיות לאנשים שעבדו 35 שעות ומעלה לשבוע, ואילו בציורים 8 ו- 9 לא הגבלנו את מספר שעות העבודה. כאשר חוזרים על הציורים על פי ההגדרות של קלינוב, הרי שההבדלים שהצבענו עליהם רק מתחזקים. על מנת לא להלאות את הקורא לא חזרנו על הציורים המגדירים השתתפות על פי קלינוב.

ציור 10 מציג את שעות העבודה הממוצעות בשבוע לפי גיל לגברים ולנשים. גם בציור זה אנו רואים ממצא דומה לממצא לגבי השתתפות בכוח העבודה. אצל הגברים קיימת עליה בשעות העבודה השבועיות מסביבות של 40 שעות עבודה לשבוע לסביבות של 50 שעות עבודה בין הגילים 25 עד 30, ולאחר מכן ממוצע שעות העבודה נשאר קבוע עד סביבות אמצע שנות החמישים וירידה בגילים מבוגרים יותר. אצל הנשים העליה היא מתונה יותר, והירידה בשעות העבודה הממוצעות מתחילה בסביבות שנות החמישים. ציור 11 מציג את עקומת ה- LMA.

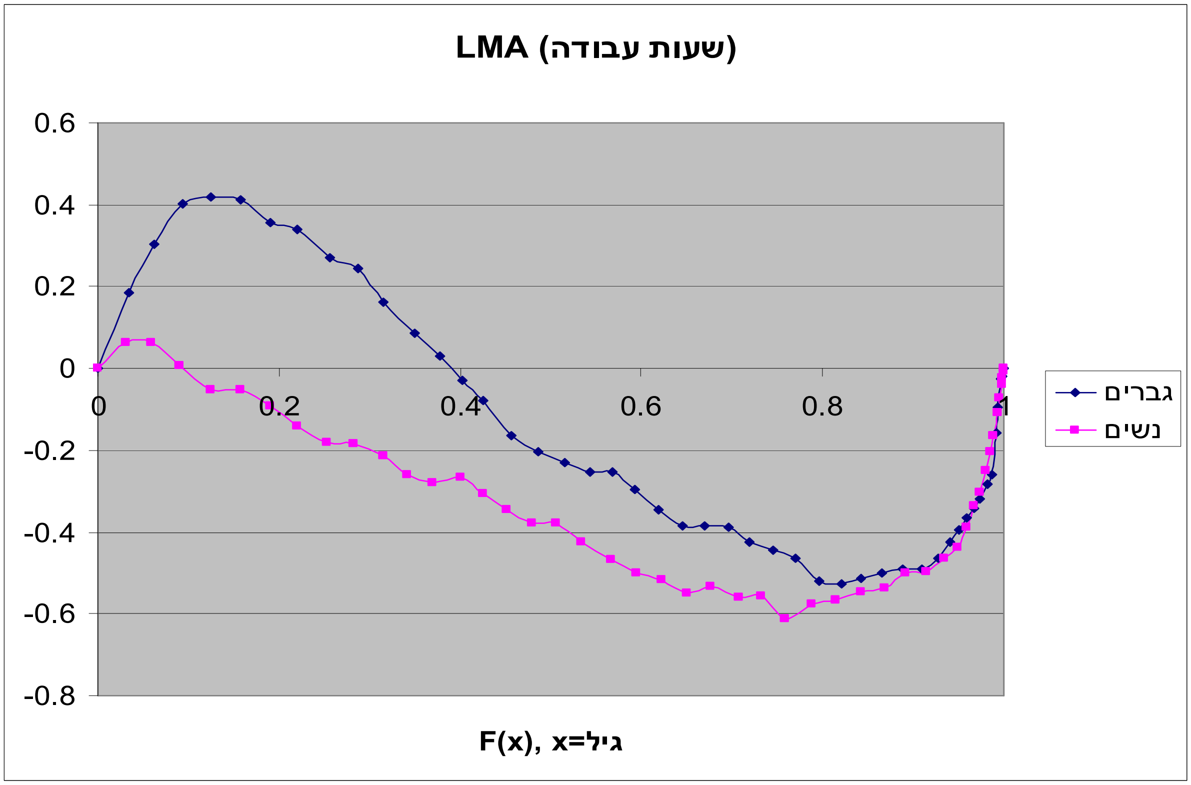
ציור 10: מספר שעות עבודה ממוצע בשבוע של נשים וגברים בהתאם לגיל



מציור 11 ניתן לראות שאצל הגברים קיים מתאם חיובי בין שעות עבודה וגיל. כך, למשל, לו היינו מתבססים רק על 60 אחוז של התצפיות בגילים הנמוכים הרי שאצל הגברים היינו מקבלים מקדם רגרסיה חיובי בהרצת רגרסיה של שעות עבודה על גיל, ואילו מארבעים אחוז מהתצפיות המתבססות על גיל גבוה היינו מקבלים מקדם רגרסיה שלילי. מקדם הרגרסיה על כלל האוכלוסייה הינו שלילי. אצל הנשים אנו מגלים תופעה דומה, אם כי שיעור העליה בשעות עבודה והמתאם החיובי נמוכים הרבה יותר מאשר אצל הגברים.

גם במקרה זה חזרנו על הציורים על פי הגדרתה של קלינוב, כאשר אנו מגבילים שעות עבודה ל- 35 שעות ומעלה. אין בממצאים מהציורים כדי לשנות את התמונה.

ציור 11: עקומת LMA של מספר שעות עבודה של נשים וגברים בהתאם לגיל



ו. סיכום

השוואה בין מקדמי משוואת השכר שהורצו בשיטת הריבועים הפחותים לאלה שהתקבלו בשיטת הג'יני מגלה שהמקדמים אינם רגישים לשיטת הרגרסיה, מה שמעיד על ניסוח טוב של המודל. אולם שימוש בקשר בין מקדם הרגרסיה על פי ג'יני לעקומת הריכוז מגלה מידע נוסף מהנתונים ששיטת הריבועים הפחותים אינה מאתרת. ראינו כי על-ידי הצגת עקומת LMA אנו מגלים שאצל הנשים הקשר בין גיל לשכר אינו קשר מונוטוני לאורך כל תחום הגילים וקיים תחום שבו עם עליית הגיל השכר יורד. השערתנו היא שכך נחשף "אפקט חזרה לעבודה" אצל הנשים, זאת אומרת התהפכות הסימן של מקדם הרגרסיה של הגיל במשוואת השכר בגילים 33–44. כדי לבדוק את המידה שבה הממצאים שלנו על ההבדל בין גברים ונשים מתקיימים גם אם מגדירים את המושג של שיעור השתתפות בצורה שונה. חזרנו ובדקנו את ההתנהגות של הנשים והגברים על פני הגילים השונים על ידי הסתכלות על שיעור ההשתתפות בכוח העבודה ועל ידי הסתכלות על שעות העבודה השבועיות הממוצעות. ההגדרות השונות לא סתרו את הממצא שקיימת

התנהגות שונה של נשים על פני הגילים מזו של הגברים. בעוד שהגברים מגבירים את השתתפותם בכוח העבודה בשנות השלושים לחייהם הרי שאין רואים אצל הנשים את הממצא הזה. אנו רואים בכך חיזוק להשערה שגידול ילדים, שנופל כנראה בעיקרו על הנשים, גורם להשתלבות נמוכה יותר בשוק העבודה ועל כן מתבטא בתמורה לעבודה. בנוסף, ראינו שגם כאשר המשתנה הבלתי תלוי הוא קטגורי, ומיוצג על ידי קבוצה של משתני דמה ולא משתנה רציף, ניתן לבנות את עקומת LMA ובלבד שיהיה משתנה סדר (אורדינלי) ושיהיו לפחות שלוש נקודות (זאת אומרת לפחות שני משתני דמה שבאים במקום משתנה רציף). ראינו זאת בדוגמת השכלה כאוסף של משתני דמה. במקרה של השכלה, לא נמצאו הבדלים בולטים בין נשים לגברים.

מקורות

קלינוב ר' (2004), "מה שעל פני השטח ומה שמתחתיו: פערי שכר בין נשים וגברים", 1970–2000, המרכז לפיתוח על-שם פנחס ספיר ליד אוניברסיטת תל-אביב, נייר דיון מס' 2004-2 .

Becker G.S. (1967), Human Capital and the Personal Distribution of Income, *W.S. Woytinsky Lecture No. 1*, Ann Arbor, University of Michigan.

Ben-Porath Y. and Gronau R. (1985), "Jewish Mother Goes to Work: Trends in Labor Force Participation of Women in Israel, 1955–1980", *Journal of Labor Economics* 3(1), Part 2, S310–S327.

Berman E. and Klinov R. (1997), "Human Capital Investment and Nonparticipation: Evidence from a Sample With Infinite Horizon" (Or: "Jewish Father Stops Going to Work"), *Discussion Paper* no. 97.05 , Jerusalem.

Efron B. (1982), "The Jackknife, the Bootstrap and Other Resampling Plans", *Society for Industrial and Applied Mathematics*, 27–35.

Heckman J.J., Lochner L.J. and Todd P.E. (2003), "Fisty Years of Mincer Earning Equations", *National Bureau of Economic Research*, WP no. 9732, Cambridge, MA.

Mincer J. (1974), "Schooling, Experience and Earnings", *National Bureau of Economic Research*, New York.

Olkin I. and Yitzhaki S. (1992), "Gini Regression Analysis", *International Statistical Review* 60(2), 185–196.

Schechtman E. and Yitzhaki S. (1987), "A Measure of Association Based on Gini's Mean Difference", *Communications in Statistics, Theory and Methods* 16, 207–231.

Schechtman E. and Yitzhaki S. (1999), "On the Proper Bounds of the Gini Correlation", *Economics Letters* 63(2), 133–138.

Schechtman E. and Yitzhaki S. (2006), "Gini's Multiple Regression", Mimeo.

Schechtman E., Yitzhaki S. and Artsev Y. (2008), "Who Does Not Respond in the Household Expenditure Survey: An Exercise in Extended Gini Regressions", *Journal of Business & Economic Statistics* , 26(3), 329-344.

Yitzhaki S. (1990), "On the Sensitivity of a Regression Coefficient to Monotonic Transformations", *Econometric Theory* 6(2), 165–169.

Yitzhaki S. (1996), "On Using Linear Regression in Welfare Economics", *Journal of Business & Economic Statistics*", 14(4), 478–486.

Yitzhaki S. (1998), "More Than a Dozen Alternative Ways of Spelling Gini", *Research on Economic Inequality* 8, 13–30.

Yitzhaki S. (2003), "Gini's Mean Difference: A Superior Measure of Variability for Non-normal Distributions", *Metron* LXI(2), 285–316.

Yitzhaki S. and Schechtman E. (2004), "The Gini Instrumental Variable" or "The 'Double Instrumental Variable' Estimator", *Metron* LXII(3), 287–313.

נספח -1

ההבדלים בין שתי שיטות הרגרסיה יוצגו בעיקר בצורה אינטואיטיבית ומילולית. הקורא המעוניין בפיתוחים התיאורטיים מופנה לסדרת מאמרים שבה פותחה הרגרסיה: Olkin and Yitzhaki (1992) פיתחו את רגרסיית הג'יני הפשוטה, Schechtman and Yitzhaki (2006) ו- Schechtman, Yitzhaki and Artzev (2008) מגדירים את תכונות הרגרסיה המרובה ומפתחים את תכונות האומדים.

על מנת לתאר את ההבדלים בין שתי שיטות הרגרסיה כדאי להתחיל במדדי הפיזור שעליהם הן נשענות. את השונות, שעל סמך תכונותיה ניתן לקבל את רגרסיית הריבועים הפחותים, ניתן לכתוב בצורה הבאה:

יהיו X_1 ו- X_2 שני משתנים מקריים בלתי תלויים ובעלי התפלגות זהה למשתנה X. אזי השונות של X ניתנת להיכתב כ:

$$\sigma_X^2 = 0.5E\{(X_1 - X_2)^2\}. \qquad (1)$$

לעומת השונות, מדד הפיזור על פי ג'יני יכול להיכתב כ:

$$G_X = E\{|X_1 - X_2|\}. \qquad (2)$$

השוואה בין (1) ל- (2) מגלה שני עקרונות המנחים את מדידת הפיזור: האחד הוא לקיחת ההפרשים בין כל שתי תצפיות שנבחרו בצורה אקראית והוא זהה לשני המדדים, והשני הוא השיטה שבה משתמשים למדידת המרחק בין נקודות: כאשר משתמשים בשונות מדידת מרחקים בין נקודות מתבססת על הגישה האוקלידית שמשמעותה היא שכל הנקודות הנמצאות במרחק שווה מנקודה נתונה נמצאות על מעגל מסביב לה, לעומת מדידת המרחק שקיימת בג'יני המכונה city block שמשמעותה היא שמדידת המרחק היא כמו בהליכה בעיר שחצויה בכבישי שתי וערב – ניתן לנוע בה רק לכיוון מזרח/מערב או צפון/דרום (אין קיצורי דרך על ידי הליכה באלכסון), ואז הנקודות הנמצאות במרחק שווה מנקודה נתונה נמצאות על היתר של משולש ישר זווית ושווה שוקיים ואילו הנקודה שממנה מודדים את המרחקים נמצאת על הקודקוד ישר הזווית. (להסבר מפורט על התכונות של מדד ג'יני ראה [Yitzhaki, 2003].) לו היינו עוסקים במרחקים בים או באוויר הרי שברור שמדידת המרחקים הנכונה היא זו שעומדת מאחרי השונות, אולם מאחר שאנו עוסקים במדידת מרחקים בין תצפיות במדגם הרי שהתשובה לשאלה מה נכון יותר אינה מובנת מאליה,

והיא תישפט לאור התכונות הנוספות שנגזרות מההבדלים במדידת המרחקים ולאור השימושים המיועדים בתוצאות הרגרסיה.

השלב הבא, החיוני להבנת ההבדל בין הרגרסיות, הוא הצגת השונות והג'יני בעזרת נוסחת השונות המשותפת (Covariance). הסיבה לכך היא שהשונות המשותפת מגדירה את מקדם המתאם ואת נוסחת הפירוק של הפיזור של קומבינציה לינארית של משתנים מקריים החיונית לרגרסיה. את נוסחאות (1) ו- (2) ניתן להציג כ:

$$\sigma_X^2 = \mathrm{cov}(X, X) \tag{3}$$

$$G_X = 4\,\mathrm{cov}(X, F(X)) \tag{4}$$

כאשר F(X) היא ההתפלגות המצטברת של המשתנה X.[7] מאחר שההתפלגות המצטברת הינה משתנה טהור חסר ממדים, היחידות של הג'יני הן היחידות של המשתנה, וזאת בניגוד לשונות שהיחידות שלה הן היחידות של המשתנה בריבוע.

בעזרת נוסחאות (3) ו- (4) ניתן להגדיר את הפרמטרים החיוניים לרגרסיה – השונות המשותפת ומקדם המתאם. ניתן לראות מנוסחה (4) שבשיטה המבוססת על ג'יני ניתן להגדיר שתי מקבילות לשונות המשותפת, והן: cov(Y,F(X)) וכן cov(X,F(Y)) כאשר X ו- Y הם משתנים מקריים. מכאן שבניגוד לשונות שנגזר ממנה מקדם מתאם אחד, הרי שבגישת הג'יני קיימים שני מקדמי מתאם בין כל שני משתנים, ותכונותיהם מצויות ב- Schechtman and Yitzhaki (1987, 1999). מקדמי המתאם בשיטת הג'יני שמוגדרים בין שני משתנים מקריים אינם חייבים להיות שווים ואף אינם חייבים להיות בעלי אותו סימן, ועל כן הם מאפשרים לקבל יותר מידע על ההתפלגויות. כלומר, בעוד שהשונות המשותפת הנגזרת מהשונות היא סימטרית בשני המשתנים אפילו אם אין קשר סימטרי בין ההתפלגויות, הרי שהשונות המשותפת על פי ג'יני אינה סימטרית, ורק כאשר מתקיימים תנאי סימטריה בין ההתפלגויות תהיה השונות המשותפת סימטרית.

לאחר שהגדרנו את המקבילה לשונות המשותפת, ניתן להגדיר את מקדם הרגרסיה ברגרסיה פשוטה. מקדם הרגרסיה הפשוטה על פי הריבועים הפחותים הוא:

$$b_{Y.X}^{OLS} = \frac{\mathrm{cov}(Y, X)}{\mathrm{cov}(X, X)} \tag{5}$$

[7] ההכפלה ב- 4 במשוואה (4) נועדה להבטיח ניסוח מתמטי נכון. בהמשך נתעלם ממנה.

ואילו מקדם הרגרסיה של הג'יני יכול להיכתב בצורה דומה כ:

$$b^G_{Y.X} = \frac{\mathrm{cov}(Y, F(X))}{\mathrm{cov}(X, F(X))} \tag{6}$$

מאחר ש- F(X) הינו משתנה חסום בין אפס לאחד, הרגישות של מקדם הרגרסיה של הג'יני לתצפיות קיצוניות קטנה יותר מאשר הרגישות של מקדם הרגרסיה של הריבועים הפחותים. שני מקדמי הרגרסיה המוצגים ב- (5) ו- (6) ניתנים להצגה כממוצעים משוקללים של השיפועים המוגדרים בין בין כל שתי נקודות סמוכות של המשתנה הבלתי תלוי (ראה Yitzhaki, 1996). חשיבותה של פרשנות זו היא בכך שהיא משחררת את המשתמש ברגרסיה מהצורך להניח קו ישר באוכלוסייה וניתן להסתפק בגישה האומרת שאנו מתאימים קירוב ליניארי לעקום הרגרסיה.

מעבר מרגרסיה פשוטה לרגרסיה מרובת משתנים הוא מסובך, ונפנה את הקורא המעוניין למאמרים שבהם פותחה הרגרסיה המרובה: Schechtman and Yitzhaki (2006) ו- Schechtman, Yitzhaki and Artzev (2008). לענייננו ניתן להסתפק באמירה הבאה: נסתכל על מקדמי הרגרסיה בשיטת הריבועים הפחותים. ניתן לנסח אותם כפתרונות של משוואות שמכילות שונויות ושונויות משותפות. כל אשר עלינו לעשות הוא להחליף כל שונות בג'יני בריבוע וכל שונות משותפת במקבילה של הג'יני, וכך נקבל נוסחאות של מקדמי רגרסיה המבוססות על ג'יני והדומות בצורת הצגתן למקדמי רגרסיה של ריבועים פחותים.[8] פורמלית, מקדמי הרגרסיה של הג'יני ניתנים להצגה הבאה:

$$\beta^G = (E(F'X))^{-1} E(F'Y) \tag{7}$$

כאשר המטריצה E[F'X] בנויה מ- $Cov(x_j, F(x_i))$ ובאופן דומה גם מוגדרת E[F'Y]. את האומדנים לפרמטרים נמצא על ידי החלפת ההתפלגות המצטברת בדרגה של כל משתנה במדגם, כאשר דרגה מסומנת ב- r:[9]

$$.\, b^G = \left[r^t X\right]^{-1} r^t Y \tag{8}$$

בדומה לרגרסיית ג'יני הפשוטה – גם ברגרסיית הג'יני המרובה אין צורך להניח מודל ליניארי באוכלוסייה. כל שעושים הוא – אומדים **קירוב** ליניארי למודל, קירוב שניתן לבחון סטטיסטית את טיבו. רגרסיית הג'יני דומה מבחינה צורנית

[8] קיומן של שתי מקבילות לשונות המשותפת רומז לכך שתהיינה שתי רגרסיות המבוססות על רגרסיית הג'יני, ואמנם כך הדבר. במאמר זה אנו מציגים את הרגרסיה הדומה לריבועים הפחותים.

[9] דרגה מתקבלת על ידי מיון התצפיות בסדר עולה והענקת מספר טבעי עולה לכל תצפית.

לרגרסיית הריבועים הפחותים, מלבד מגבלה אחת: על מנת שהמטריצה במשוואה (8) תהיה הפיכה, אסור שמשתנה בלתי-תלוי אחד יהיה פונקציה מונוטונית של אחר, כי אחרת, מכיון שערכי פונקציות ההתפלגות שלהם זהים לכל התצפיות תהיה תלות לינארית בין שורות המטריצה כך שלא ניתן להפכה. כדאי להעיר שלמרות ש (8) לא הושגה על ידי מינימיזציה הרי שהטעויות המוגדרות כסטייה של המשתנה התלוי מהמודל הלינארי, $e = Y-bX$, יוצרות, בדומה לריבועים הפחותים, מתאם אפס עם המשתנה הבלתי תלוי (מתקבלים תנאים המקבילים למשוואות הנורמליות) – אולם מאחר שקיימים שני מתאמים לג'יני בין שני משתנים – רק אחד מהם מתאפס – כלומר $\mathrm{cov}(e, F(X)) = 0$. השונות המשותפת השנייה תתאפס גם היא אם המודל הוא ליניארי – תכונה המאפשרת לבדוק את הנחת הליניאריות בעזרת רגרסיית הג'יני. כלומר, כאשר פותרים את משוואה (8) מקבלים שמתקיים $\mathrm{cov}(e, F(X)) = 0$ מתוך הגדרה לכל משתנה בלתי תלוי. אם, לעומת זאת, היינו ממזערים את הג'יני של הסטייה מקו הרגרסיה היינו מקבלים $\mathrm{cov}(X, F(e))= 0$. כאשר קיים באוכלוסייה קו ישר (כלומר: כאשר המודל הליניארי מתאים) הרי ששתי המשוואות תתקיימנה, ואילו כאשר עקום הרגרסיה הוא קמור או קעור הרי שרק משוואה אחת תתאפס – בהתאם לשיטה שבה נקטנו.

נספח 2 – פונקציית שכר עבור עובדים שכירים ,בגילים 70–25

<u>קבוצת האוכלוסייה:</u>

עובדים משרה מלאה, ועבדו כל השנה או חלק ממנה.

<u>משתנה מוסבר:</u>

LN(E) – לוג השכר כאשר הבסיס הוא e.

משתנים מסבירים:

Sex – גברים: 1, נשים: 2.

Age – גיל.

Age2 – גיל בריבוע.

H – משתנה דמי עבור שעות עבודה (הבסיס – 40–44 שעות עבודה).

Married – משתנה דמי עבור נשוי.

Total – מספר הנפשות במשק הבית.

Nonjew – משתנה דמי עבור האוכלוסייה הלא יהודית.

S – משתני דמי עבור השכלה (בסיס: 11–12 שנות לימוד).

Whole work – משתנה דמי עבור העובדים בשאר התקופות.

Olim – משתנה דמי עבור עולים חדשים (אחרי 1990).

ענפים:

P_tech- משתני דמי עבור חקלאות, תעשייה ובנייה.

S_h_tech – משתני דמי עבור חשמל ומים, בנקאות, ביטוח ופיננסים, שירותים עסקיים.

Public – משתני דמי עבור מנהל ציבורי, חינוך, בריאות, שירותים חברתיים וקהילתיים.

Low-tech (בסיס) – תחבורה, אחסנה ותקשורת, מסחר סיטונאי וקמעונאי ותיקונים, שירותי אירוח ואוכל ושירותים פרטיים למשק הבית.

נספח 3 – תוצאות ההרצות של הרגרסיות.

לוח 1: משוואת שכר של גברים, סקר הכנסות (הערכים המסומנים ב * חושבו בנפרד בהתאם להגדרותיהם. הם אינם מופיעים כחלק מהפלט של התוכנה).

		Gini			OLS		
		B	std(b)	p-value	B	std(b)	p-value
	Intercept	7.2220*			7.2686	0.1166	0.0000
מס' נפשות במשפחה	**total**	-0.0183	0.0000	0.0000	-0.0192	0.0056	0.0000
נשוי	**married**	0.1591	0.0005	0.0000	0.1515	0.0201	0.0000
לא יהודי	**nonjew**	-0.1764	0.0005	0.0000	-0.1823	0.0231	0.0000
עולה חדש	**olim**	-0.3449	0.0003	0.0000	-0.3466	0.0177	0.0000
שנות לימוד	**e0_8**	-0.2716	0.0011	0.0000	-0.2746	0.0323	0.0000
	e9_10	-0.1630	0.0006	0.0000	-0.1679	0.0266	0.0000
	e13_15	0.2323	0.0004	0.0000	0.2261	0.0192	0.0000
	e16	0.4826	0.0004	0.0000	0.4784	0.0191	0.0000
תחום עיסוק	**P-tech**	0.0495	0.0004	0.0000	0.0427	0.0185	0.0212
	S_h_tech	0.3040	0.0013	0.0000	0.2961	0.0359	0.0000
	Public	0.1084	0.0004	0.0000	0.0999	0.0193	0.0000
מס' שעות עבודה בשבוע	**h35_39**	-0.2279	0.0027	0.0000	-0.2376	0.0431	0.0000
	h45_49	0.1711	0.0004	0.0000	0.1613	0.0190	0.0000
	h50	0.3656	0.0004	0.0000	0.3560	0.0179	0.0000
גיל	**age**	0.0546	0.0000	0.0000	0.0530	0.0057	0.0000
	age^2	-0.0005*			-0.0005	0.0001	0.0000

לוח 2: משוואת שכר לנשים, סקר הכנסות (הערכים המסומנים ב * חושבו בנפרד בהתאם להגדרותיהם. הם אינם מופיעים כחלק מהפלט של התוכנה).

		Gini			OLS		
		b	std(b)	p-value	B	std(b)	p-value
	Intercept	7.6343*			7.1290	0.1376	0.0000
מס' נפשות במשפחה	**total**	-0.0265	0.0073	0.0003	-0.0242	0.0071	0.0007
נשוי	**married**	0.1340	0.0188	0.0000	0.1338	0.0183	0.0000
לא יהודי	**nonjew**	-0.1880	0.0330	0.0000	-0.1911	0.0402	0.0000
עולה חדש	**olim**	-0.2310	0.0190	0.0000	-0.2309	0.0188	0.0000
שנות לימוד	**e0_8**	-0.2000	0.0441	0.0000	-0.1991	0.0494	0.0001
	e9_10	-0.2020	0.0334	0.0000	-0.2020	0.0384	0.0000
	e13_15	0.1950	0.0208	0.0000	0.1950	0.0210	0.0000
	e16	0.4160	0.0205	0.0000	0.4161	0.0206	0.0000
תחום עיסוק	**P-tech**	0.1530	0.0281	0.0000	0.1535	0.0269	0.0000
	S_h_tech	0.5050	0.0338	0.0000	0.5048	0.0339	0.0000
	Public	0.1860	0.0224	0.0000	0.1867	0.0210	0.0000
מס' שעות עבודה בשבוע	**h35_39**	-0.0775	0.0227	0.0006	-0.0776	0.0236	0.0010
	h45_49	0.1500	0.0218	0.0000	0.1600	0.0210	0.0000
	h50	0.3260	0.0255	0.0000	0.3262	0.0244	0.0000
גיל	**Age**	0.0458	0.0033	0.0000	0.0446	0.0069	0.0000
	age^2	-0.0004*			-0.0004	0.0001	0.0000