

סדרת ניירות עבודה
WORKING PAPER SERIES

מס' 21

**בדיקת נתוני שכר במפקד 1995
באמצעות קובץ שכר של ביטוח לאומי**

**Analysis of Wage Data from the 1995 Census by using
Wage File of the National Insurance Institute**

דמיטרי רומנוב* ואורלי פורמן*
Dmitri Romanov* Orly Furman*

תמוז תשס"ו, יולי 2006

***הלשכה המרכזית לסטטיסטיקה – אגף מדען ראשי**

***Central Bureau of Statistics – Chief Scientist Department**

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 13015, ירושלים 95464 טלפון: 02-6592666, פקס' 02-6521340

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

תקציר

נתוני הכנסה – משכר, עבודה עצמאית וחמש קצבאות של המוסד לביטוח לאומי – שנאספו במפקד 1995 היוו נתון מפתח לאין ספור מחקרים בזכות גודל המדגם, יכולת קישור רשומות ברמת הפרט למפקד 1983 ומקורות מזוהים אחרים, ושילוב בין נתונים כלכליים לבין מאפיינים דמוגרפיים, תעסוקתיים וחברתיים.

לצורך טיוב והשלמת נתוני שכר שנאספו במפקד התקבל מהמוסד לביטוח לאומי קובץ שכר לשנת 1995, הכולל מידע על שכר שנתי של 1.82 מיליון שכירים בכל המשרות שדווחו לביטוח לאומי, ופירוט של חודשי עבודה במהלך השנה. קובץ זה שימש אותנו לבדיקת נתוני שכר שדווחו על ידי הנפקדים ולבחינת נכונות של תיקונים וזקיפות שנעשו לשכר בשלבי עיבוד נתוני המפקד.

השכר של 307 אלף שכירים שרשום בקובץ המפקד נמוך ב-16.5 אחוזים בממוצע משכר של אותם השכירים בקובץ ביטוח לאומי. עריכת נתוני שכר כללה חמישה סוגי תיקונים, כאשר השכיחים ביניהם היו תיקון על-ידי רגרסיה (21 אחוזים) ושתילת ערכים מתוך קובץ ביטוח לאומי (8.6 אחוזים). רק ל-68 אחוזים מהשכירים, שרשום להם שכר במפקד, הנתון המקורי של שכר לא תוקן – עבור אלה הנתון המדווח היה נמוך ב-15.6 אחוזים בממוצע מהשכר המצוי בקובץ ביטוח לאומי.

השוואה בין שכר מדווח לשכר תחשיבי מקובץ ביטוח לאומי ברוטו ובניכוי תשלומי חובה (נטו) מעלה כי בחצי התחתון של התפלגות ההכנסות נתון של שכר מדווח קרוב יותר לשכר תחשיבי נטו מאשר לשכר תחשיבי ברוטו. לעומת זאת, עבור החצי העליון של התפלגות ההכנסות ניתן לקבוע כי נתון הכנסה שנאסף במפקד הוא לרוב שכר ברוטו שדווח בחסר.

תיקון נתוני שכר על-ידי שתילת ערכים מתוך קובץ של ביטוח לאומי היה לא מדויק בחלק מהמקרים. שימוש ברגרסיה לצורך זקיפת נתוני שכר היה מיותר משום שניתן היה להיעזר בפונקציות של מס הכנסה, דמי ביטוח לאומי ודמי ביטוח בריאות כדי לגזור אומדן סביר של שכר ברוטו משכר נטו, כאשר האחרון דווח על ידי הנפקד והראשון לא דווח.

ברצוננו להודות לאביבה בטאט (ז"ל) על הערותיה המועילות.

בדיקת נתוני שכר במפקד 1995 באמצעות קובץ שכר של המוסד לביטוח לאומי

א. הגדרות

מטרת הבדיקה היא לבחון נתוני שכר ותעסוקה במדגם 20 אחוזים של מפקד האוכלוסין והדיוור 1995 (להלן, המדגם) באמצעות השוואתם למידע המצוי בקובץ שכר של המוסד לביטוח לאומי עבור שנת מס 1995 (להלן, קובץ שכר). במסגרת הבדיקה ייבחנו הנתונים הבאים:

- הכנסת שכיר (ברוטו) לחודש ספטמבר 1995 (להלן, שכר ספטמבר);
- מעמד שכיר בעבודה בחודש ספטמבר 1995.

בקובץ המדגם שהועמד לצורך הבדיקה חלק מנתוני שכר ספטמבר מקוריים (אלה שנאספו במהלך המפקד) נמחקו, תוקנו או הוחלפו בנתון אחר – מתוך קובץ שכר או תוך זקיפת ערך תחשיבי על בסיס גרסיה. במקרים אלה לא ידוע לנו מה היה הנתון המקורי. לכן נתונים של שכר ספטמבר ייבחנו בשני אופנים: עבור נתונים מקוריים נבדוק את התאמתם לנתוני שכר בקובץ שכר; עבור נתונים שתוקנו נבדוק את סבירות התיקון.

קובץ שכר, שמקורו בדיווח שנתי של מעסיקים למוסד לביטוח לאומי על עובדים שכירים שמועסקים על ידם, כולל משתנים דמוגרפיים בסיסיים (מין, מצב משפחתי, תאריך לידה, תאריך עליה), מידע מינהלי (זיהוי מעביד וסיווג המבוטח – שכיר, עוזרת בית, פנסיונר וכדומה), שכר שנתי ורשימת חודשי עבודה במהלך שנת המס. כל רשומה בקובץ מזוהה לפי מספר זיהוי של העובד – מה שישמש כמפתח זיווג עם נתוני המדגם. יש להדגיש כי השכר בקובץ ביטוח לאומי הינו השכר החייב בדמי ביטוח לאומי, ויש תשלומים מסוימים שמשלמים במסגרת יחסי עובד-מעביד שלא חייבים בדמי ביטוח לאומי, כגון פיצויי פיטורין, פדיון חופשה עם פרישת העובד, ההוצאות שהותרו לעובד לניכוי ועוד. רשימת חודשי עבודה בקובץ שכר מפרטת עבור אילו חודשים שולם שכר, או, ליתר דיוק, עבור אילו חודשים עובד מבוטח בענפי הביטוח הלאומי.

משרה	שכר שנתי (ש"ח)	חודשי עבודה/ביטוח (משמאל לימין)	שכר תחשיבי לצורך ההשוואה לשכר ספטמבר
א	12000	111111111111	$12000/12=1000$
א	10000	001100111000	$10000/5=2000$
א	10000	111110000000	0
א	3000	000110000000	0
ב	10000	011111111110	$10000/10=1000$
סה"כ	13000	011111111110	$1000+0=1000$
א	3000	000000111000	$3000/3=1000$
ב	8000	000000011110	$8000/4=2000$
סה"כ	11000	000000111110	$1000+2000=3000$

משום שבמדגם נתון השכר מתייחס לחודש ספטמבר 1995, נגדיר שכר חודשי בקובץ שכר כסך תשלום שנתי חלקי מספר חודשי עבודה בשנה – בתנאי שהפרט עבד בחודש ספטמבר, או אפס אם לא עבד בספטמבר. בלוח לעיל מודגם החישוב (נתון של ספטמבר מודגש).

המקרה הרביעי בלוח מתאר מצב של שכיר שעבד בשני מקומות עבודה במהלך השנה: חודשיים במשרה א' ועשרה חודשים במשרה ב'. משום שחודשי עבודה במשרה א' אינם כוללים ספטמבר, שכר תחשיבי של הפרט שווה לשכר תחשיבי במשרה ב'. המקרה האחרון בלוח מביא דוגמה של חישוב עבור שתי משרות כאשר בשתייהן הפרט עבד בספטמבר. מכאן ברור הכלל: שכר תחשיבי של הפרט מתקבל כסכום של שכר תחשיבי לספטמבר בכל המשרות, ולא כיחס בין שכר שנתי בכל המשרות לבין מספר חודשי עבודה/ביטוח בשנה.

לצורך בדיקת מעמד שכיר בעבודה נגדיר שכיר בקובץ שכר כפרט שהיה מבוטח כשכיר, על פי סיווג המוסד לביטוח לאומי, בספטמבר 1995 באחת ממשרות עבודה שהחזיק.

חודשי עבודה בקובץ שכר מוגדרים כאיחוד חודשי עבודה/ביטוח בכל המשרות של שכיר. מספר זה שונה מסך חודשי עבודה בכל המשרות. כך, בשתי הדוגמאות האחרונות בלוח לעיל מספר חודשי עבודה לצורך החישוב הוא 10 ו-5, כאשר סכימה הייתה נותנת 12 ו-7 חודשים, בהתאמה. ככלל, הבדיקות ייעשו רק לפרטים שעבורם קיים מספר זיהוי תקין במפקד, הרי אם ערך חסר לא ניתן לאתר פרט בקובץ שכר.

המשך העבודה מאורגן כדלקמן. בפרק ב' תוצג השוואה של קובץ שכר לנתונים שוטפים של שכר ומספר משרות שכיר ב-1995 במטרה להעריך מידת השלמות של קובץ מינהלי שישמש לבדיקת נתוני מפקד. בפרק ג' תיערך בדיקת שכר שדווח במדגם 20 אחוזים של מפקד 1995 באמצעות הקבלת הרשומות, ברמת הפרט, לקובץ שנתי של שכירים המבוטחים במוסד לביטוח לאומי. בפרק ד' נבחן מהימנות דיווח על מעמד בעבודה. סיכום יחתום את העבודה.

ב. קובץ שכר של המוסד לביטוח לאומי ל-1995

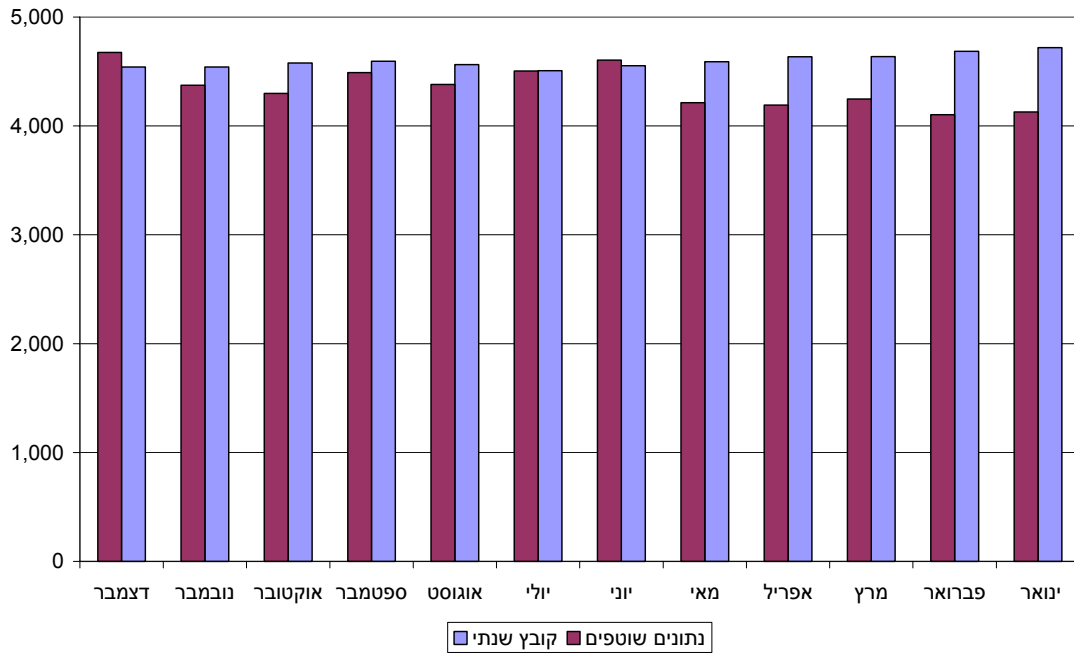
טרם ביצוע בדיקה שבה קובץ מינהלי משמש לאימות נתוני שכר המדווחים במפקד מתבקשת בדיקה לתקפות הנתונים בקובץ מינהלי עצמו. לא ניתן לעשות זאת באמצעות השוואה עם מקור בלתי תלוי, הרי דיווח מעסיקים למוסד לביטוח הלאומי הוא המקור היחיד לסטטיסטיקת השכר החודשי. השוואה בין סטטיסטיקה על שכר ומשרות שכיר ב-1995, שמתקבלת באופן שוטף ממדגם בדוק ומשוקלל של כ-9.5 אלף מפעלים (מעסיקים), לנתונים על כל אוכלוסיית השכירים בקובץ שנתי של המוסד לביטוח לאומי אמורה ללמד על שלמות הקובץ.

בגרף 1 מוצגת סדרה חודשית של שכר למשרת שכיר ומספר משרות שכיר במשק ב-1995. בממוצע שנתי, מספר משרות בקובץ שכר נמוך מהנתונים השוטפים ב-3.6 אחוזים (1,965 אלף לעומת 2,039 אלף), ואילו שכר למשרת שכיר בקובץ גבוה כ-5.6 אחוזים (4,596 ש"ח לחודש לעומת 4,355 ש"ח). בחודש ספטמבר, שעומד במוקד ההשוואה בין מפקד לקובץ שכר, הפערים דווקא קטנים יחסית: מספר משרות שכיר בקובץ שכר נמוך ב-3.1 אחוזים ממספר משרות על פי הנתונים השוטפים, בעוד שהשכר למשרת שכיר גבוה ב-2.3 אחוזים.

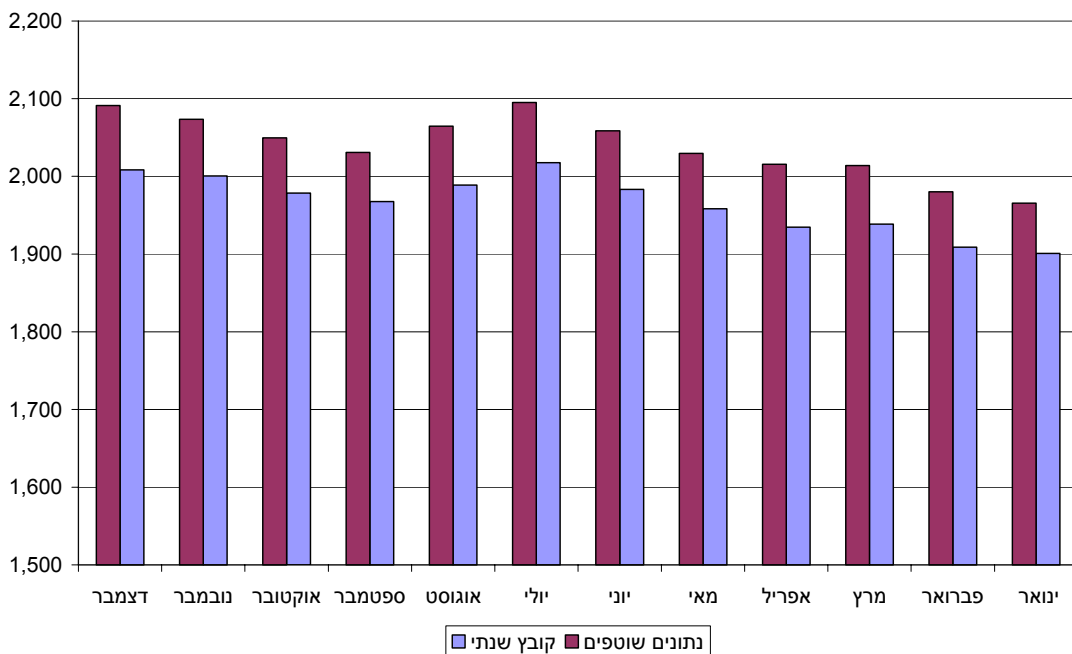
הפער במספר משרות שכיר, 74 אלף בממוצע שנתי, יכול לנבוע, בחלקו הארי, מכך שבקובץ שכר ששימש לבדיקת המפקד נכללים רק אזרחי ישראל, ואילו בנתונים שוטפים כלולים עובדים זרים שמדווחים לביטוח לאומי – מספרם ב-1995 עמד על 60 אלף איש בממוצע שנתי. אשר לפער בשכר למשרת שכיר, יש לקחת בחשבון כי שיטת החישוב בשני המקורות שונה. בקובץ שכר הנתון נגזר משכר שנתי, כפי שמתואר בהמשך, ועל כן הוא מוחלק על פני השנה.

גרף 1. שכר ממוצע למשרת שכיר ומספר משרות שכיר על פי נתונים שוטפים של המוסד לביטוח לאומי ומתוך הקובץ השנתי של 1995

א': שכר ממוצע למשרת שכיר, ש"ח



ב': משרות שכיר,



אלפים

הערה: נתונים שוטפים – ללא עובדים מיהודה, שומרון ועזה.

נסמן w_{it} שכר של הפרט במשרה i בחודש t , אזי השכר שיש בקובץ השנתי לכל משרה הוא

$$W_i = \sum_t w_{it}$$

שכר ממוצע למשרת שכיר במשק בחודש נתון מתוך הקובץ השנתי הוא

$$\sum_i (W_i / N_i) / \sum_i N_i$$

כאשר N_i מספר חודשי עבודה בשנה של הפרט במשרה i .

לעומת זאת, השכר הממוצע למשרת שכיר מדיווחים שוטפים של מעסיקים בחודש נתון t מחושב

$$\sum_i w_{it} / \sum_i N_{it}$$

באופן הבא: בנוסף לשוני שנובע מצורת החישוב, ההבדלים בהגדרת האוכלוסייה

שהתבטאו בפער במספר משרות שכיר אמורים להשפיע גם על השכר הממוצע למשרת שכיר (ולגרום להטיה כלפי מעלה בנתון מתוך הקובץ השנתי).

ג. בדיקת שכר ספטמבר

מהשוואת שכר ממוצע בין המפקד לקובץ שכר עולה כי השכר שדווח במפקד נמוך ב-16.5%

בממוצע משכר אותם שכירים בקובץ השכר. שיעור ההפרש מיוחס לשכר השכירים בקובץ שכר, שכר אותו מצאנו כגבוה יותר מהשכר המדווח ולכן שיעור ההפרש גבוה מכל השיעורים החלקיים שיוצגו בהמשך מאחר ואלה חושבו מתוך השכר המדווח.

להלן השוואת השכר הממוצע ומספר שכירים בין המפקד לקובץ שכר.

נתוני מפקד		קובץ שכר מזווג עם מפקד		קובץ שכר	
הפרש לעומת מפקד מזווג עם קובץ שכר, אחוז	סה"כ	הפרש לעומת קובץ שכר, אחוז	סה"כ	שכר ממוצע (ש"ח)	מספר שכירים ¹
-16.5	3,978	1.2	4,763	4,706	
-8.1	1,566.5	-6.1	1,705.0	1,815.1	

השכירים אשר עלו במדגם 20% של המפקד אמורים לייצג נאמנה את כלל אוכלוסיית השכירים כולל את שכרם. פרק זה מוקדש להבנת הסיבות לסטייה זו.

¹ מספר שכירים במדגם כפול 5.

נתונים מקוריים של שכר ספטמבר במדגם תוקנו בכמה שיטות. להלן רשימת סוגי תיקון ושכיחותם במדגם, כפוף לתנאי שפרט מזוהה ומעמד בעבודה שלו שכיר.

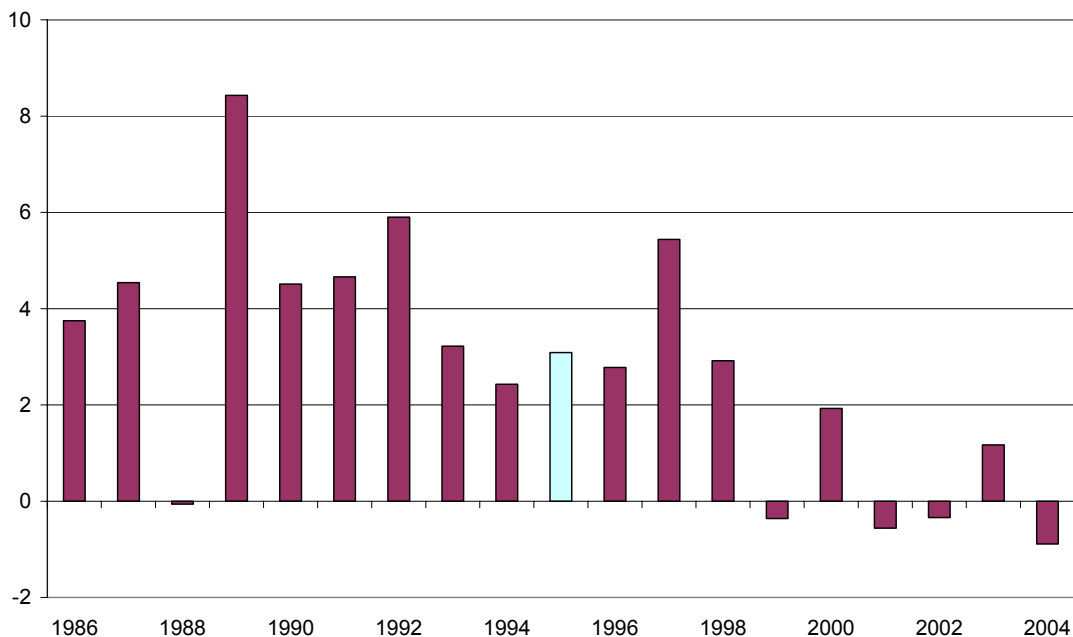
לוח 1. תיקונים שנעשו לנתון השכר במדגם

קוד	פירוש	מספר מקרים	אחוז מסה"כ
X	ערך שלא בוצע בו כל תיקון	211,678	68.9
r	חישוב שכר לפי רגרסיה	64,467	21.0
5	שתילת נתון מקובץ שכר	26,378	8.6
q	חלוקה ב-10.100 של הכנסות	3,784	1.2
1	ערך מחוק לריק שהיה נכון	692	0.2
z	שכפול נתונים	15	0.0
	סה"כ	307,014	100.0

אנו לא נעסוק בערכים מחוקים ואלה ששוכפלו. מראש יש להסתייג ולומר כי אין לצפות שדיווח הנפקד על שכרו (ברוטו) בחודש ספטמבר יהיה שווה לשכר תחשיבי בקובץ שכר, שהוא ממוצע לכל חודשי עבודה באותה השנה, משום שקיימת שונות בשכר בין חודשי השנה בגלל גידול טבעי של השכר, תשלומים עונתיים, תוספת יוקר, הסדרי תשלום מוסדיים וכיו"ב. גרף 2 מציג הפרש בין שכר ממוצע למשרת שכיר בחודש ספטמבר לבין שכר למשרת שכיר ממוצע באותה השנה (על פי דיווחים למוסד לביטוח לאומי, במחירים שוטפים). בגרף רואים כי בשנות ה-90 שכר בחודש ספטמבר היה לרוב גבוה מהשכר הממוצע בשנה, כאשר בשנת 1995 הפער עמד על 3.1 אחוזים. בשנים 2001 עד 2004 הפער כמעט נמחק (הן בגלל תהליך דיסאינפלציה ופירוק מנגנוני הצמדת שכר, והן בגלל שינויים במועדי תשלום והיקף תשלום תוספות במגזר הציבורי).

לאור עובדות אלה ניתן לצפות ששכר ספטמבר (לא מתוקן) במדגם יהיה גבוה בכמה אחוזים מהממוצע השנתי כפי שמופיע בקובץ שכר. בלוח 2 מוצגת השוואה שכזאת בנפרד לכל קבוצת נתונים לפי סוג תיקון – נתונים מקוריים סומנו בקוד תיקון X – עבור בעלי משרת שכיר יחידה במהלך 1995. בלוח 3 מוצגת אותה ההשוואה עבור שכירים שעבדו בכמה מקומות עבודה במהלך שנה. הפרדה בין שתי אוכלוסיות אלה – מחזיקי משרה אחת ומחזיקי משרות רבות – נחוצה בגלל השוני בתחשיב שכר בקובץ ביטוח לאומי אשר לכשעצמו עשוי לגרום לפער בין שכר ספטמבר לבין שכר תחשיבי.

גרף 2. הפרש בין שכר למשרת שכיר בספטמבר למוצע שנתי, 1986-2004, אחוזים



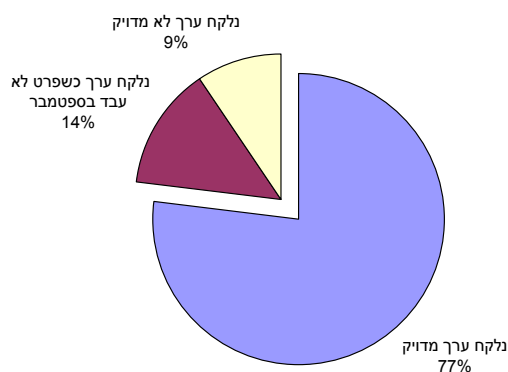
כעת נתבונן בשורה הראשונה של לוח 2 המתייחסת לאותם מקרים שבהם נתון השכר במדגם מקורי (לא תוקן): בממוצע, שכר תחשיבי היה גבוה ב-19 אחוזים משכר ספטמבר שדווח במדגם. לא רק שפער זה נפתח בכיוון הפוך לצפוי מתוך סטטיסטיקה מצרפית של שכר, גם גודלו לא ניתן להסבר על-ידי הבדלים קלים בין מה שנדגמים תופסים כ"שכר ברוטו" לבין השכר החייב בדמי ביטוח לאומי שמופיע בקובץ שכר. ההשערה היחידה שניתן להעלות להסבר פער שכזה מדברת על אפשרות שכתוצאה מטעות בקידוד השכר במדגם הוא נטו (בניכוי מס הכנסה, דמי ביטוח לאומי ודמי ביטוח בריאות) ולא ברוטו. טענה זו נבחנת בהמשך.

הקבוצה הגדולה של ערכים מתוקנים, כ-40 אלף מקרים שהם 20 אחוזים ממדגם שכירים בעלי משרה אחת במהלך 1995, היא תוצאה של זקיפה באמצעות רגרסיה. זקיפה זו נדונה בהרחבה בהמשך, ואילו עכשיו נעבור לתיקון באמצעות שתילת נתונים מקובץ ביטוח לאומי (תיקון בקוד 5) הנוגע ל-8.4 אחוזים מהנדגמים, שכן לפי הגדרה ערכים אלה חייבים להיות זהים לנתוני קובץ. כפי שהוצג בגרף 3, בשלושה רבעים מהמקרים אכן שתלו ערך מתאים; ב-2,260 מקרים נלקח ערך שכר כאשר הפרט לא עבד בחודש ספטמבר; יש עוד 1,562 מקרים בהם נלקח נתון אחר והסטייה המוחלטת בינו לבין הנתון שמצאנו בקובץ עומדת על 547 ש"ח בממוצע.

לוח 2. השוואת שכר ספטמבר לשכר ממוצע ב-1995 לשכירים שהייתה להם משרה אחת במהלך השנה, לפי סוג תיקון נתונים במדגם

קוד תיקון	מספר מקרים	שכר ספטמבר במדגם ממוצע (סטיית התקן), ש"ח לחודש	שכר תחשיבי בקובץ שכר ממוצע (סטיית התקן), ש"ח לחודש	הפרש ממוצע (אחוז משכר ספטמבר)
X	140,192	4,514 (4,201)	5,373 (5,486)	19.0
r	39,952	3,754 (2,803)	4,244 (4,171)	13.0
5	16,663	6,199 (8,964)	5,778 (6,895)	-6.8
q	2,372	5,370 (4,844)	5,718 (5,202)	6.5
סה"כ	199,179	4,513	5,184	14.9

גרף 3. התפלגות ערכים שנשתלו על בסיס קובץ שכר של ביטוח לאומי לשכירים שהייתה להם משרה אחת במהלך 1995



בלוח 3 מוצגת השוואה עבור אותם פרטים שעבדו בכמה משרות – במקביל או זו אחרי זו – בשנת 1995. ניתן לראות כי בקבוצת ערכים לא מתוקנים (קוד תיקון X) הפער לעומת שכר תחשיבי מקובץ ביטוח לאומי עומד על כ-17 אחוזים, בדומה למה שראינו בלוח 2. בקבוצת ערכים שתולים מתוך קובץ שכר (קוד תיקון 5) שוב אין זהות לשכר תחשיבי (גרף 4), אך להבדיל מהמצב בקרב בעלי משרת שכיר אחת שראינו בגרף 3, בגלל ריבוי משרות רק ב-7 אחוזים מהמקרים נלקח ערך מתאים; ב-75 אחוזים מהמקרים מקובץ שכר נלקח נתון לא מתאים כך שהסטייה המוחלטת הממוצעת הגיעה ל-1,475 ש"ח, לעומת 547 ש"ח במקרה של משרה אחת. בממוצע, השכר במדגם בעלי משרות רבות גדול ב-13.4 אחוזים מהממוצע התחשיבי בקובץ ביטוח לאומי, בהשוואה לפער של 14.9 אחוזים בקרב שכירים שעבדו במשרה אחת בלבד.

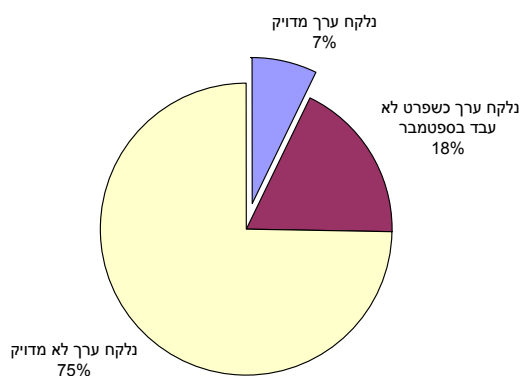
לוח 3. השוואת שכר ספטמבר לשכר ממוצע ב-1995 לשכירים שהיו להם כמה משרות במהלך השנה, לפי

סוג תיקון נתונים במדגם

קוד תיקון	מספר מקרים	שכר ספטמבר במדגם ממוצע (סטיית התקן), ש"ח לחודש	שכר תחשיבי בקובץ שכר ממוצע (סטיית התקן), ש"ח לחודש	הפרש ממוצע (אחוז משכר ספטמבר)
X	58,664	3,251 (3,656)	3,789 (5,205)	16.6
r	19,462	3,010 (2,294)	3,061 (3,305)	1.7
5	9,254	3,721 (6,614)	3,965 (6,431)	6.6
q	1,346	2,828 (4,157)	5,672 (7,232)	100.5
סה"כ	88,726	3,241	3,676	13.4

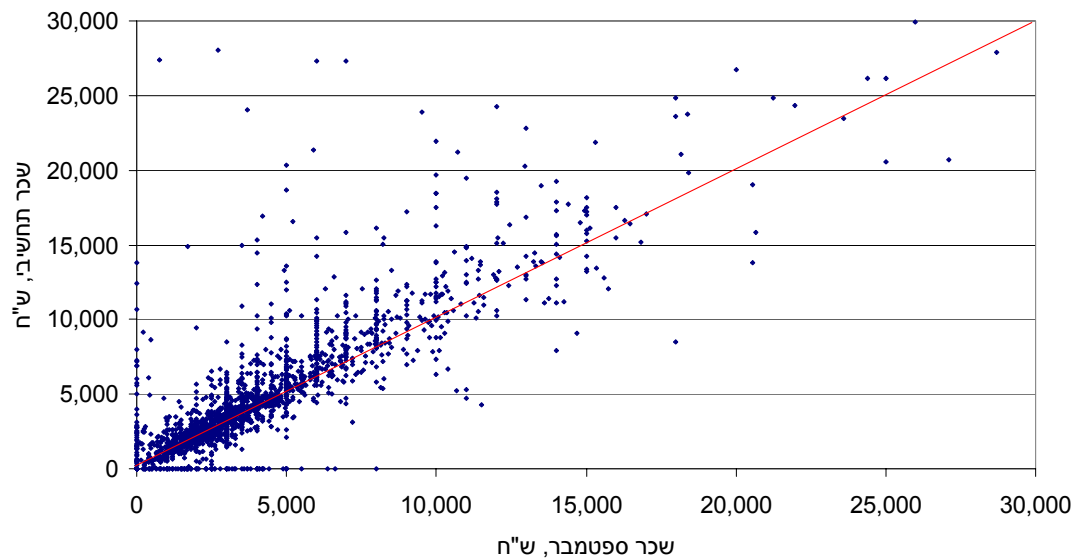
גרף 4. התפלגות ערכים שנשתלו על בסיס קובץ שכר של ביטוח לאומי לשכירים

שהייתה להם יותר ממשרה אחת במהלך 1995



הפער הממוצע בין השכר ברוטו שדווח במפקד לבין נתון השכר מקובץ ביטוח לאומי (בקבוצת מקרים שלא תוקנו) נע, כאמור, בין 17 ל-19 אחוזים (מתוך הנתון המדווח). בגרף 5 מוצג שכר מדווח ושכר תחשיבי עבור מדגם מיקרי של כ-2000 פרטים (כשלושים תצפיות הושמטו מהגרף כי אחד מהערכים היה מעל 30 אלף ש"ח). הקו האדום של אלכסון מציין, כמובן, צפי לדיווח חסר הטייה על השכר – וניתן לראות בבירור שפיזור התצפיות מסביב לאלכסון אינו מאוזן: שכר תחשיבי גדול, בממוצע, משכר ספטמבר. עוד רואים בגרף 5 כי ישנם מקרים לא מעטים (5.5 אלפים מתוך 199 אלף) בהם פרטים דיווחו על שכר ספטמבר אפס כאשר שכר תחשיבי חיובי, וההיפך (3.12 אלף מקרים) כאשר פרטים לא עבדו בספטמבר על פי נתוני ביטוח לאומי אך דיווחו במדגם על שכר כלשהו.

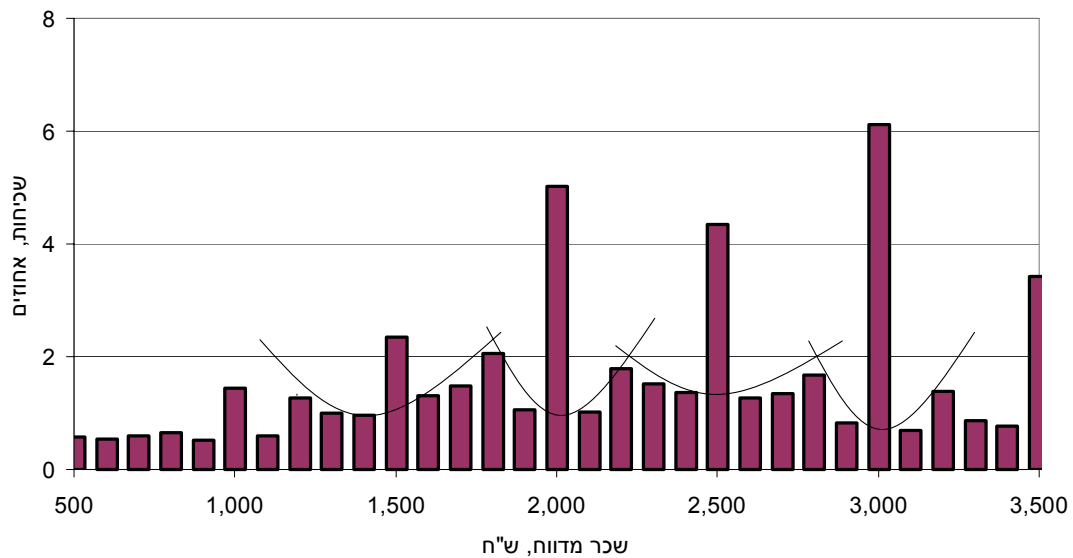
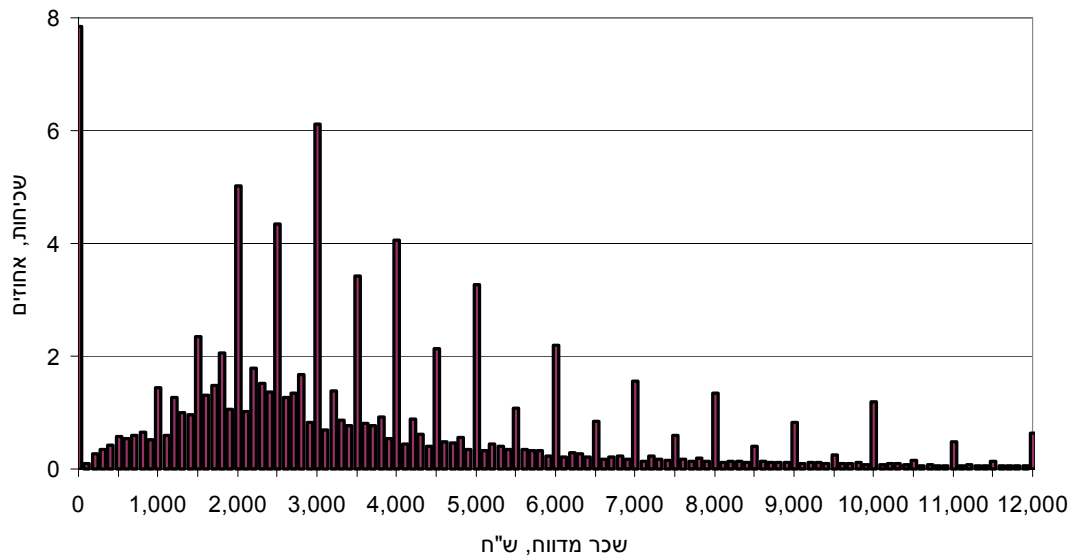
גרף 5. שכר ספטמבר שדווח במדגם ושכר חודשי תחשיבי מקובץ ביטוח לאומי, ש"ח



בגרף 5 ניכרת תופעה של עיגול השכר המדווח, שמשקפת ב"פסים" אנכיים בטווח של עד 10,000 ש"ח. בגרף 6 התופעה מוצגת בבירור. שכיחות חורגת של השכר המדווח במספר עגול של אלפים וחצי אלפים מוכיחה שהפרטים אכן נוטים לעגל את הכנסתם. לאור עובדה זו נשאלת השאלה האם מדובר בעיגול על פי כללי החשבון (למשל, גם 2,900 וגם 3,100 מעגלים ל-3,000), או הפרטים "מושכים" את סכומי השכר שלהם כלפי מעלה עד למספר העגול הקרוב? לשתי שיטות העיגול הללו השלכות שונות על מהימנות הדיווח על השכר. במקרה הראשון לא עשויה להיווצר הטייה באומדן השכר הממוצע, בעוד שהשיטה השנייה כן תתבטא בהטייה שהולכת ופוחתת יחסית לרמת השכר האמיתי של הפרט.² בפאנל התחתון של גרף 6 רואים כי מסביב לכל נקודה של שכיחות חורגת בסכום עגול של השכר (בטווח ההכנסה עד 3,500 ש"ח לחודש) נוצר "שקע" סימטרי. הדבר מעיד על כך שעיגול הסכומים על-ידי הנדגמים נעשה על-פי כללי החשבון ואין בו, אפוא, סיבה להטייה.

² עיגול הסכומים הינו רק אחד הגורמים להטייה אפשרית בנתוני ההכנסה. סיבות נוספות להטייה הן בלבול בין הכנסה נטו לברוטו, בין הכנסה כוללת להכנסה מעבודה, דיווח כוזב בתכלית וכו'.

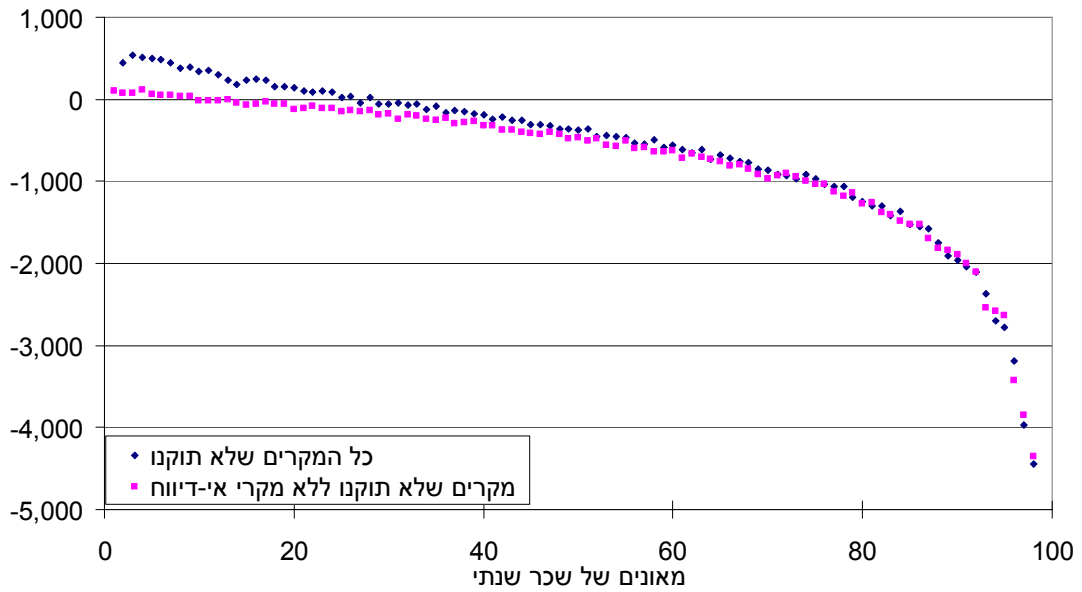
גרף 6. התפלגות שכר ספטמבר שדווח במדגם



הערה: נתונים עבור 197.6 אלף פרטים שעבורם שכר ספטמבר לא תוקן.

מתוך הסתכלות בגרף 5 קשה להעריך האם גודל הפער בין השכר המדווח לבין השכר התחשיבי ושונות של הפער תלויים ברמת ההכנסה של הפרט. כדי לענות על השאלה זו יש להפריד מקרים של אי-דיווח (שכר מדווח אפס) ממקרים של דיווח. כמו כן, יש להוציא מקרים של דיווח כשאין רישום שכר בקובץ ביטוח לאומי. שתי קבוצות אלה מונות, כאמור, כ-17.8 אלף איש. גרף 7 מציג פער בין שכר ספטמבר ושכר תחשיבי ממוצע (הפאנל העליון) וסטיית התקן של הפער (הפאנל התחתון), עם ובלי מקרי אי-דיווח באחד משני המקורות.

גרף 7. פער בין שכר ספטמבר המדווח במדגם לבין שכר חודשי תחשיבי (פאנל עליון),
 וסטיית התקן של הפער (פאנל תחתון), לפי מאונים של שכר שנתי בקובץ ביטוח לאומי, ש"ח

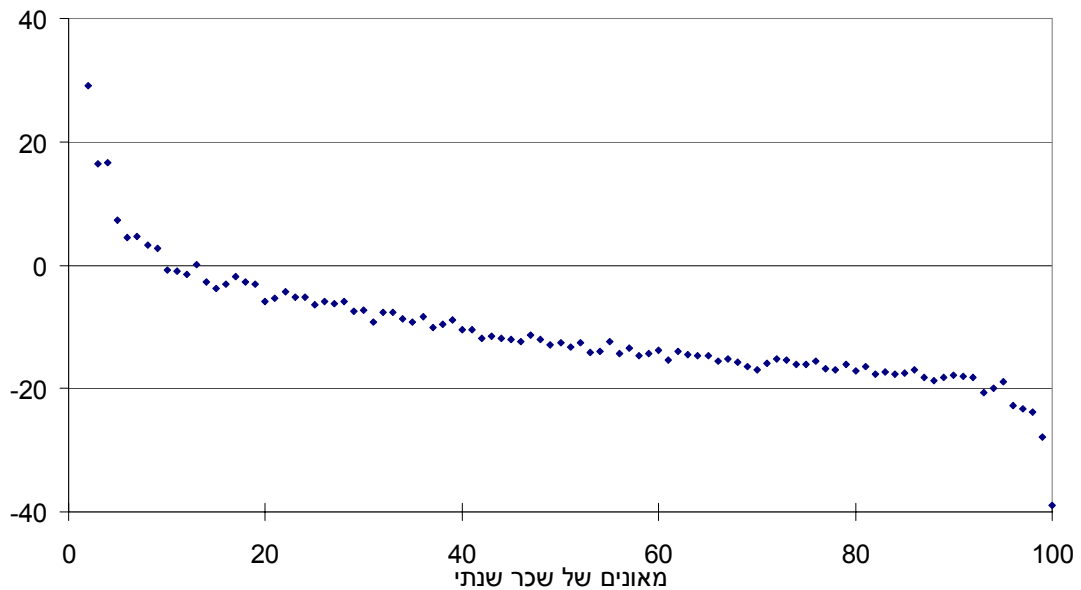


הערה: כל המקרים שלא תוקנו –199 אלף פרטים, ללא מקרי אי-דיווח – 181 אלף.

מגרף 7 עולות המסקנות הבאות:

- הפער בין השכר המדווח במדגם לבין השכר החודשי התחשיבי מקובץ ביטוח לאומי יורד עם ההכנסה, דהיינו בהכנסות נמוכות (27 מאונים ראשונים במדגם של 199 אלף פרטים או 9 מאונים ראשונים במדגם של 181 אלף פרטים ללא מקרים של אי-דיווח) ישנו דיווח יתר, בעוד שבהכנסות בינוניות וגבוהות הנדגמים מוסרים דיווח חסר על שכרם;
- תת-דיווח הטרוסקדסטי – שונות הפער הולכת וגדלה עם ההכנסה;
- חשוב להתייחס לבודד מקרי אי-דיווח על שכר (כ-18 אלף) משום שהשפעתם על גודל הפער ועל שונותו לא מבוטלת, אולם המקרים הללו אינם משנים את התוצאות באופן איכותי.

גרף 8. סטיית שכר ספטמבר המדווח במדגם משכר חודשי תחשיבי, לפי מאונים של השכר השנתי המדווח לביטוח לאומי, אחוזים משכר תחשיבי



הערה: נתונים עבור 181 אלף פרטים שדיווח על שכר ספטמבר שלהם במדגם לא תוקן, ללא מיקרי אפס שכר במדגם או בקובץ ביטוח לאומי.

בגרף 8 נבחן הגודל היחסי של הפער בין השכר המדווח לבין הנתון התחשיבי, כאחוז מהשכר התחשיבי. שיעור הסטייה הולך ויורד עם הכנסה, ועולה על 30 אחוזים בשני קצוות של התפלגות ההכנסות (בסימנים הפוכים, כמובן).

הסבר אפשרי לסטייה כה גדולה של השכר המדווח משכר תחשיבי יכול להימצא בבלבול בין הכנסה נטו וברוטו. כאן נבדוק טענה זו, אם כי ברור שהיא לא רלוונטית למצב בחלק תחתון של התפלגות ההכנסות, שבו נפוץ דיווח יתר וחבות המס זניחה (הכנסה מתחת לסף מס הכנסה, וחלים רק שיעורים מופחתים של ביטוח לאומי וביטוח בריאות). אשר לעצם האפשרות של בלבול בין הכנסה נטו וברוטו, ייתכן כי הבלבול בין נטו לברוטו קיים בקרב הציבור, אך לא מן הנמנע כי חלק מבלבול זה מקורו בטעות בקידוד ההכנסות בשלבי עיבוד נתוני המדגם. סברה אשר לא ניתן לאשש או לסתור בהתחשב בעובדה כי מידע על הכנסה נטו שנאסף במדגם לא נשמר בכל צורה שהיא.

על פי שאלון המדגם נוסח שאלה 34 הוא "מה הייתה הכנסתך **ברוטו** (לפני ניכויים) **ממשכורתך** בספטמבר 1995?" [הדגשה במקור]. ברם, אם לא ידועה לפרט הכנסה ברוטו ניתן היה לרשום הכנסה נטו, תוך סימול מתאים. על-פי מסמך פנימי של צוות המפקד, הרישום המקורי (לפני זקיפות, תיקונים ואובדן נתוני הכנסה נטו) היה כדלקמן:

מספר מקרים	הכנסה ברוטו	הכנסה נטו
71,798	אין	יש
152,326	יש	אין
56,747	יש	יש

כדי לבדוק האם ייתכן שמה שסומן בקובץ המפקד כהכנסה ברוטו מהווה למעשה הכנסה נטו, נערך חישוב ניכויי חובה – מס הכנסה, דמי ביטוח לאומי ודמי ביטוח בריאות – על בסיס שכר חודשי תחשיבי מקובץ ביטוח לאומי. לאחר מכן נשווה את התוצאה להכנסה שרשומה בקובץ המדגם, בהנחה שנתון מייצג הכנסה נטו. יש לסייג חישוב זה כי מדובר באומדן המיסוי הישיר הסטטוטורי שמתבסס על שיעורי מס והטבות מס בסיסיות בלבד, אלה שניתן לזקוף על פי המידע הקיים בקובץ המדגם, ללא התחשבות בהטבות ספציפיות רבות (בגין ילדים, ליישובים ואזורי ספר, עבור חסכון פנסיוני וכו'). יתרה מזאת, בסיס המס לחישוב – שכר תחשיבי – אינו תואם הכנסה חייבת בדמי ביטוח לאומי (שכר חודשי וחשוב חודשי של חבות המס), ולא בסיס מס הכנסה (שכר שנתי וחשוב שנתי של חבות המס). להלן שיעורים ומדרגות מס ששימשו לחישוב.

א. מס הכנסה (שיעורים ומדרגות שחלו בספטמבר 1995)

שיעור המס (אחוזים)	מדרגת המס (ש"ח)
15	2,900
30	7,640
45	13,850
50	כל שקל מעל 13,850

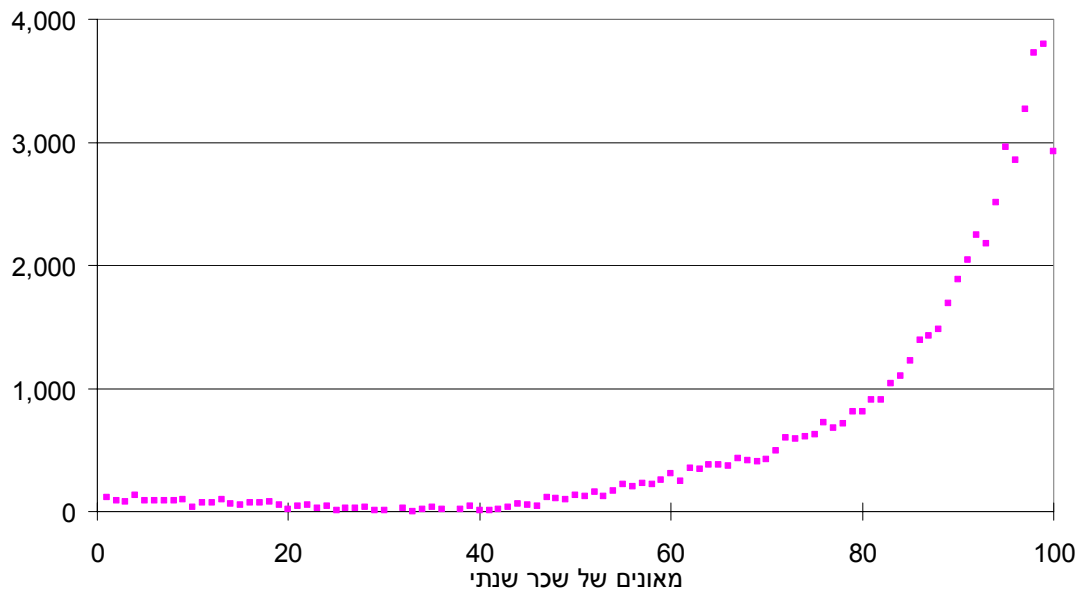
שווי נקודת זיכוי 125 ש"ח לחודש. נלקחו בחשבון הטבות מס הבאות: זיכוי לתושב (2.25 נ"ז), זיכוי לעולה חדש (לפי תאריך עלייה), זיכוי לאישה נשואה (1 נ"ז).

ב. ביטוח לאומי וביטוח בריאות (שיעורים ומדרגות שחלו בספטמבר 1995)

שיעור המס (אחוזים)	מדרגת המס (ש"ח)
0	1,041
3.71	2,081
9.7	16,648
0	כל שקל מעל 16,648

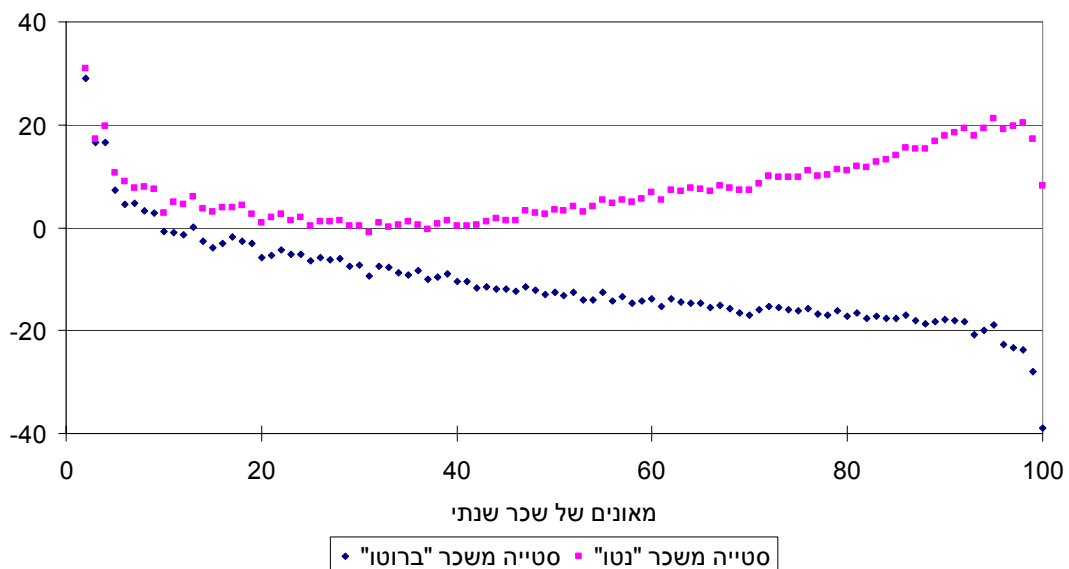
בגרף 9 מוצג פער בין הכנסה נטו תחשיבית – שכר תחשיבי מקובץ ביטוח לאומי בניכוי אומדני מס הכנסה, ביטוח לאומי וביטוח בריאות – לבין שכר שדווח במדגם, אשר אמנם סומן כ"ברוטו" אך אולי פירושו "נטו".

גרף 9. פער בין שכר ספטמבר המדווח במדגם לבין שכר חודשי תחשיבי בניכוי תשלומי חובה, לפי מאונים של שכר שנתי בקובץ ביטוח לאומי, ש"ח



הערה: נתונים עבור 181 אלף פרטים שדיווח על שכר ספטמבר שלהם במדגם לא תוקן, ללא מיקרי אפס שכר במדגם או בקובץ ביטוח לאומי.

גרף 10. סטייה של שכר ספטמבר המדווח במדגם משכר חודשי תחשיבי (ברוטו) ובניכוי תשלומי חובה (נטו), לפי מאונים של שכר שנתי, אחוזים משכר תחשיבי ברוטו



ניתן להסיק מגרף 9 מסקנות מעורבות. מצד אחד, עבור כמחצית מהמדגם – בהכנסות נמוכות, איפה שמעבר מברוטו לנטו לא אמור לחולל שינויים רבים – ההפרש בין אומדן תחשיבי של הכנסה נטו לדיווח במפקד לא עולה על מאה שקלים. מצד שני, בקרב בעלי הכנסות גבוהות ניכוי תשלומי חובה לא גרם למחיקת הפער בין השכר המדווח לבין תחשיב. גרף 10 מעמיד סטייה של השכר המדווח משכר תחשיבי ברוטו מול הסטייה משכר תחשיבי בניכוי תשלומי חובה – בהנחה שהשכר המדווח הוא למעשה הכנסה

נטו. מחד גיסא, בגרף 10 ניתן למצוא תמיכה בהשערה כי בהכנסות נמוכות ובינוניות (ממאון 10 עד מאון 50 בערך) השכר המדווח יכול להיות שכר נטו, משום שסטייתו משכר תחשיבי נטו קרובה מאוד לאפס. מאידך גיסא, התבדרות הקווים לשני צידי ציר 0 בחמשת העשירונים הגבוהים מעידה על כך שבקצה העליון של התפלגות ההכנסות כנראה נפוץ דיווח חסר על ההכנסות.

נחזור לתיקון נתוני שכר ספטמבר באמצעות רגרסיה (קוד תיקון z). זה התיקון השכיח שנגע לכל רשומה חמישית במדגם, דהיינו עבור כ-60 אלף פרטים שכר ספטמבר מהווה פרי זקיפה שרירותית. אי לכך, חשוב להבין את שיטת הזקיפה וניתן לעשות זאת מתוך מסמך פנימי של צוות המפקד הקרוי "עריכת הכנסות ברוטו/נטו מעבודה כשכיר". להלן, הקטע ממנו (איות של מקור):

"על סמך 56,747 רשומות בהם גם הכנסות נטו וגם הכנסות ברוטו של המפקד היו ידועות נקבעו

רגרסיות להמרת הכנסות נטו להכנסות ברוטו. רגרסיות חושבו ב-4 טווחי ההכנסות:

עד 1000	$Bruto=1.175*neto-16.12$
1001 עד 3000	$Bruto=1.337*neto-273.25$
3001 עד 7500	$Bruto=2.013*neto-2471.27$
מעל 7500	$Bruto=1.544*neto+2208.23$

על סמך הרגרסיות הנ"ל בוצעה המרת הכנסות נטו בקבוצת מקרים שבהם הכנסת נטו מלאה

והכנסת ברוטו ריקה".

ראוי להדגיש כי בקובץ אין מידע על הכנסה נטו (בהנחה שכל מה שסומן כברוטו אכן מתייחס

לברוטו ולא לנטו, כפי שנדון לעיל), למרות שמידע זה נאסף בשאלה 34 של השאלון ושימש לרגרסיה ולזקיפה עצמה, כך שלא ניתן לאמת נתון זה. אשר לשיטת הזקיפה, הכנסה "ברוטו" והכנסה "נטו" של הפרט קשורות דרך פונקצית מס שידועה לכל שנת מס, לכל חודש, וניתנת למידול מדויק יחסית, כפי שנעשה למשל בסקר הכנסות משולב. לפיכך, יש מקום לבחון את נכונות הזקיפה באמצעות רגרסיה.

כדי להעריך את תוצאות הזקיפה באמצעות רגרסיה אנו נעמיד את הערך החזוי של "bruto" מול

הערך הנגזר מפונקציית מס הפוכה (inverted) שכוללת מס הכנסה, דמי ביטוח לאומי ודמי ביטוח בריאות

במצב הנכון לספטמבר 1995, כפי שתוארו לעיל. משום ש"רגרסיה" אינה מתחשבת במגדר השכיר (נשים מקבלות יותר הטבות מס מגברים) ובמצב משפחתי, החישוב נעשה עבור גבר שמקבל 2.25 נקודות זיכוי ממס הכנסה (יחיד, או נשוי שאישתו עובדת), ולאישה שמקבלת 4.25 נקודות זיכוי ממס הכנסה (נשואה או אם חד-הורית עם שני ילדים עד גיל 18).

מגרף 11 רואים כי ערכים של שכר ספטמבר שנזקפו לפי רגרסיה הולכים ומתרחקים עם הגידול

בהכנסה מהערכים שהיה צריך לזקוף לפי פונקצית מס. הפער (הטעות) עומד על 10 אחוזים ברמת שכר של 3,000 ש"ח, עולה ל-20 אחוזים בסביבה של 5,000 ש"ח, ומגיע ל-30 אחוזים ברמת שכר של 8,000 ש"ח.

לבסוף, בגרף 12 מועמדים ערכים שנזקפו לפי רגרסיה במדגם מול השכר התחשיבי לחודש

ספטמבר מקובץ שכר ביטוח לאומי. התמונה דומה מאוד לזו שראינו בגרף 7, עבור אלה שנתון השכר שלהם

לא תוקן. על פי נתוני לוח 2 ולוח 3 ניתן היה לחשוב שרגרסיה "מדויקת" יותר בזקיפת השכר לעומת

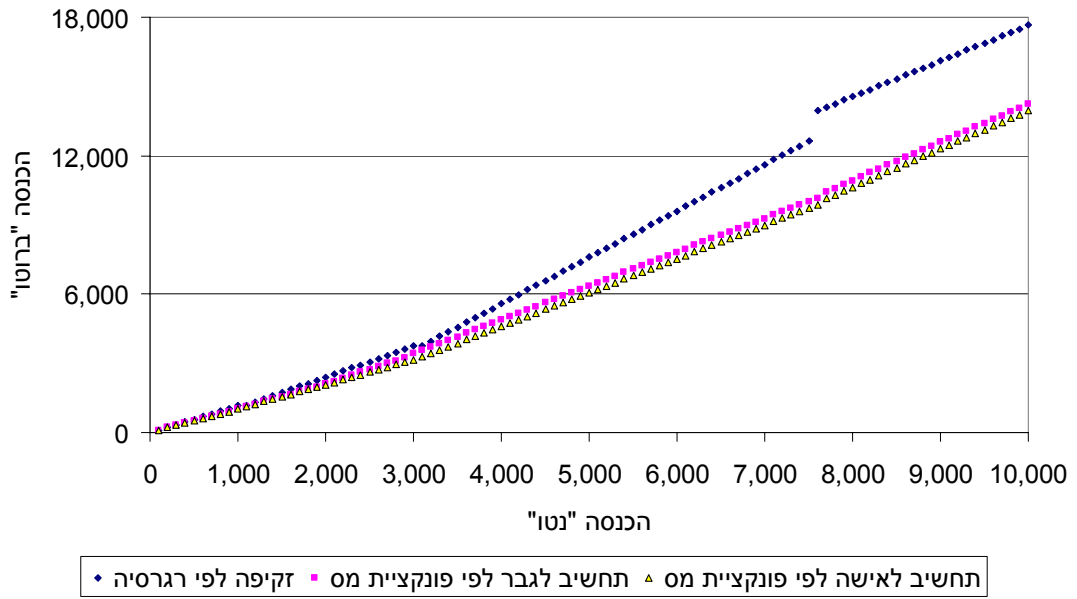
הדיווח הלא מתוקן: הסטייה מהשכר התחשיבי עומדת על 13 אחוזים בקרב שכירים שהועסקו במשרה

אחת ו-2 אחוזים בלבד בקרב בעלי משרות רבות (לעומת סטייה של 19 ו-17 אחוזים בהתאמה בשכר

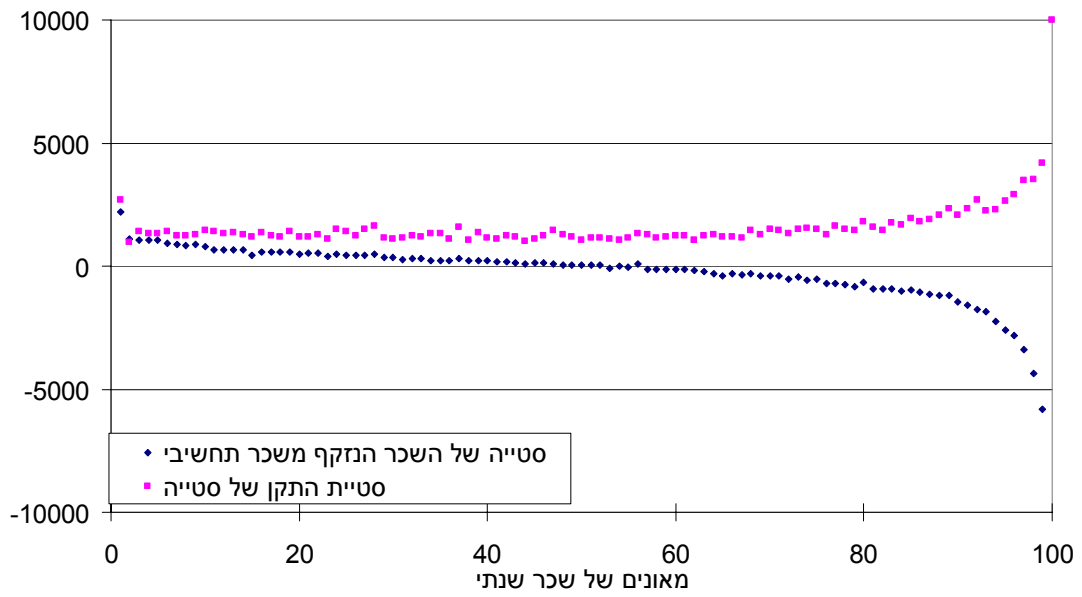
המדווח שלא תוקן). ברם, בגרף 11 נמצא הסבר סוד ה"דיוקן" – הוא תוצאה של ההטיה כלפי מעלה

בחישוב הערך הזקוף, שבמקרה הוביל בכיוון הנכון.

גרף 11. השוואת הערך שנזקף לפי רגרסיה לשכר שנגזר לפי פונקציית מס ישיר



גרף 12. סטיית שכר ספטמבר זקוף לפי רגרסיה משכר חודשי תחשיבי מקובץ ביטוח לאומי, לפי מאונים של השכר השנתי המדווח לביטוח לאומי, ש"ח מחירים שוטפים



הערה: נתונים עבור 59 אלף פרטים שעבורם שכר ספטמבר נזקף לפי רגרסיה.

חוסר ההתאמה בין השכר המדווח ע"י הנפקד בחודש ספטמבר לבין השכר התחשיבי מקובץ שכר ביטוח לאומי יכול לנבוע כאמור ממספר סיבות:

- א. חוסר תפישה אחידה לגבי השכר בחודש מסוים (חודש ספטמבר במפקד). ייתכן והנפקד, כשנשאל בחודש נובמבר על שכרו בספטמבר, התייחס לחודש עבודה בפועל או לחילופין לחודש בו קיבל את השכר, דהיינו לאוקטובר. לעומת זאת, השכר התחשיבי בקובץ ביטוח לאומי יתייחס תמיד לשכר בגין חודש ספטמבר;
 - ב. האפשרות שקיבלה חיזוק בסעיפים קודמים, כי דיווח הנפקד הוא אודות שכר נטו ולא ברוטו, כפי שמופיע בקובץ ביטוח לאומי;
 - ג. השכר הנמצא בקובץ ביטוח לאומי הוא שכר תחשיבי הממוצע על פני חודשי העבודה במהלך השנה, וידוע כי קיימת שונות בשכר לאורך חודשי השנה, שונות הנובעת מגידול טבעי לאורך השנה, תוספות שכר, תשלומים עונתיים וכו'.
- בכדי לנסות ולמקד את הסיבה לשונות נעשתה בדיקה לטיב המתאם בין ההכנסה לחודש ספטמבר המדווחת ע"י הנפקד לעומת השכר התחשיבי בקובץ ביטוח לאומי לחודש ספטמבר. חשוב להדגיש כי בשכר התחשיבי נלקחו בחשבון כל ההכנסות שהיו לפרט בחודש ספטמבר, קרי הכנסה לשכיר (כפי שמופיע בטבלה בעמ' 1). כמו כן נבדקו רק אותם מקרים בהם לא בוצע תיקון כלשהו לדיווח הנפקד. נמצא כי מקדם מתאם פירסון הוא 0.807 (לפי ספירמן 0.81). כאשר הוצאו מהמדגם אותם מקרים שבהם השכר התחשיבי היה 0 (דהיינו עפ"י המוסד לביטוח לאומי הפרט לא עבד בספטמבר), נמצא מתאם של 0.793.
- מקדם המתאם בין שכר ספטמבר המדווח על ידי הנפקד לעומת השכר השנתי בקובץ ביטוח לאומי עומד על 0.825 (מתאם ספירמן 0.843). מתאם חזק יותר, לעומת השכר התחשיבי החודשי, מבטא כנראה עובדה כי הנפקדים נוטים לדווח על רמה "ממוצעת" של שכר על פני השנה (תוך עיגול ערכים כפי שראינו בגרף 5), מה שמתאים יותר לשכר שנתי מאשר שכר תחשיבי חודשי.
- המתאם החזק בין השכר המדווח לבין השכר ממקור מינהלי מסתיר מידה לא מבוטלת של אי-התאמה במדרג ההכנסות של הפרטים, כפי שמוצג בלוח 4.

לוח 4: שכיחות שכירים לפי חמישוני שכר מדווח במפקד ושכר חודשי תחשיבי מקובץ ביטוח לאומי (אחוזים)

סה"כ	חמישוני שכר תחשיבי בקובץ ביטוח לאומי					חמישוני שכר שדווח במפקד
	5	4	3	2	1	
100.0	2.3	2.6	4.7	18.9	71.6	1 (תחתון)
100.0	0.8	4.1	19.5	57.9	17.7	2
100.0	2.6	17.4	53.7	19.4	6.8	3
100.0	12.6	60.2	21.5	2.4	3.3	4
100.0	79.6	17.7	1.0	0.2	1.4	5 (עליון)

הערה: נתונים עבור 197.6 אלף פרטים שעבורם שכר ספטמבר לא תוקן.

מלוח 4 עולה כי בקצה העליון של התפלגות ההכנסות כ-80 אחוזים מהשכירים מדורגים לפי השכר המדווח במפקד באותו החמישון שהיו מקבלים על פי השכר התחשיבי; בקצה התחתון של התפלגות ההכנסות אותו החמישון היה מתקבל עבור כ-72 אחוזים של השכירים, ואילו באמצע ההתפלגות ההתאמה בין שני

המדרגים יורדת לסביבה של 55-60 אחוזים. על גודל הסטייה בין שתי התפלגויות השכר ניתן ללמוד מכך שהשכר התחשיבי הממוצע של אותם שכירים שנמצאים בתאים [3, 4] ו-[2, 3] במטריצה של לוח 4, דהיינו משתייכים לפי המפקד לאותו החמישון השלישי של השכר, עומד על 5,267 ש"ח ו-2,501 ש"ח בהתאמה – הבדל של פי שתיים בערך.

ד. בדיקת דיווח על מעמד בעבודה

כל הרשומות הנמצאות בקובץ שכר מתייחסות למשרות שכיר במובן שדמי ביטוח לאומי בגינן שולמו לפי שיעורים של שכיר (ישנם עוד שני סוגי מבוטחים: עצמאים ו"לא שכירים", לפיהם קיימת הבחנה של מעמד בעבודה במוסד לביטוח לאומי). לכן בדיקה של מהימנות הדיווח על מעמד בעבודה תכלול שתי השוואות: מעמד בעבודה שדווח במדגם לעומת מעמד "שכיר" בענפי ביטוח לאומי בשנת 1995; מעמד בעבודה שדווח במדגם לעומת רישום של עבודה שכירה בחודש ספטמבר 1995. ההשוואה השנייה מהווה, אפוא, בדיקה משולבת של מעמד בעבודה ועצם עבודה כשכיר בחודש ספטמבר. בכדי לבחון את השכירים בקובץ ביטוח לאומי אשר דיווחו על מעמד אחר במדגם המפקד נבחרו כל המקרים שהופיעו בשני המקורות. מתוך כלל המקרים בעלי מעמד מסווג (447.8 אלף) נמצאו 18.8 אלף פרטים שדיווחו על מעמד שלא שכיר (4.2 אחוזים מהמדגם), למרות שהיו על-פי קובץ ביטוח לאומי מבוטחים כשכירים, מתוכם לכ-12.3 אלף היה גם שכר בחודש ספטמבר.

לוח 5: התפלגות מעמד בעבודה המדווח במפקד לשכירים שנמצאו בקובץ ביטוח לאומי

מעמד בעבודה המדווח במפקד	סה"כ במפקד	מזה: נמצאו בקובץ ביטוח לאומי עם שכר חודשי תחשיבי	מזה: דיווחו במפקד על שכר בספטמבר
שכיר	447,635	428,957	402,072
עצמאי שלא מעסיק שכירים	25,023	7,504	4,445
עצמאי המעסיק 1-2 שכירים	11,344	2,861	1,878
עצמאי המעסיק 3-9 שכירים	6,419	2,661	1,906
עצמאי המעסיק 10 שכירים יותר	2,607	1,914	1,466
חבר קואופרטיב	1,363	1,144	1,076
חבר קיבוץ (שאינו שכיר)	11,244	1,812	1,068
בן משפחה ללא תשלום	2,110	701	286
עצמאי שלא ידוע כמה שכירים מעסיק	466	246	211
מעמד לא ידוע	82,158	54,695	7,772
סה"כ	590,369	502,495	422,180

ישנם כ-18.7 אלף פרטים אשר דיווחו על מעמד שכיר במדגם המפקד אולם לא נמצאו כלל בקובץ ביטוח לאומי.

ה. סיכום ומסקנות

חשיבותו של מפקד האוכלוסין והדיוור 1995 לסטטיסטיקה בסיסית של מדינת ישראל עצומה, וכך גם חשיבותו של מידע חברתי כלכלי עשיר שנאסף במסגרת "שאלון ארוך" ממדגם של עשרים אחוזים של הנפקדים – עבור מחקר יישומי במגוון תחומי חברה וכלכלה.

נתוני הכנסה – משכר, עבודה עצאית וחמש קצבאות של המוסד לביטוח לאומי – שנאספו במפקד היוו נתון מפתח לאין ספור מחקרים בזכות גודל המדגם, יכולת קישור רשומות ברמת הפרט למפקד 1983 ומקורות מזוהים אחרים, ושילוב בין נתונים כלכליים לבין מאפיינים דמוגרפיים, תעסוקתיים וחברתיים. לצורך טיוב והשלמת נתוני שכר שנאספו במפקד התקבל מהמוסד לביטוח לאומי קובץ שכר לשנת 1995, הכולל מידע על שכר שנתי של 1.82 מיליון שכירים בכל המשרות שדווחו לביטוח לאומי, ופירוט של חודשי עבודה במהלך השנה. קובץ זה שימש אותנו לבדיקת נתוני שכר שדווחו על ידי הנפקדים ולבחינת נכונות של תיקונים וזקיפות שנעשו לשכר בשלבי עיבוד נתוני המפקד.

השכר של 307 אלף שכירים שרשום בקובץ המפקד נמוך ב-16.5 אחוזים בממוצע משכר של אותם השכירים בקובץ ביטוח לאומי. הסיבה לסטייה כה גדולה אינה נעוצה במדגם שלא מייצג את אוכלוסיית השכירים, כי השכר הממוצע של אותם 307 אלף שכירים שעלו במדגם 20 אחוזים, כפי שמופיע בקובץ ביטוח לאומי, דווקא גבוה ב-1.2 אחוז מהשכר הממוצע של שכר השכירים במשק, לפי אותו המקור. מכאן, שסטייה כלפי מטה בנתון השכר במפקד נובעת מתהליך איסוף ועיבוד נתונים, שכולל בתוכו כמה סיבות אפשריות: דיווח חסר של הנפקדים, טעויות שיטתיות בקליטת נתונים, עריכה שגויה של הנתונים.

עריכת נתוני שכר כללה חמישה סוגי תיקונים, כאשר השכיחים ביניהם היו תיקון על-ידי רגרסיה (21 אחוזים) ושתילת ערכים מתוך קובץ ביטוח לאומי (8.6 אחוזים). רק ל-68 אחוזים מהשכירים, שרשום להם שכר במפקד, הנתון המקורי של שכר לא תוקן – עבור אלה הנתון המדווח היה נמוך ב-15.6 אחוזים בממוצע מהשכר המצוי בקובץ ביטוח לאומי.

מאחר ובמפקד נאספו נתונים של שכר ברוטו ושכר נטו ונתוני שכר נטו אבדו עם הזמן, ניתן רק לשער שנתוני הכנסה לשכירים הם אכן שכר ברוטו. כדי לבחון השערה זו, עבור כ-200 אלף שכירים שנתון השכר שלהם לא תוקן, נערכה השוואה בין שכר מדווח לבין שכר מקובץ ביטוח לאומי מצד אחד ותחשיב שכר נטו על בסיס שכר מקובץ ביטוח לאומי מצד שני. השוואה זו מעלה כי בחצי התחתון של התפלגות ההכנסות נתון של שכר מדווח קרוב יותר לשכר תחשיבי נטו מאשר לשכר תחשיבי ברוטו. לעומת זאת, עבור החצי העליון של התפלגות ההכנסות ניתן לקבוע כי נתון הכנסה שנאסף במפקד הוא לרוב שכר ברוטו שדווח בחסר.

לכאורה, ערכים של שכר שנשתלו מתוך הקובץ של ביטוח לאומי צריכים להתאים לנתוני הקובץ. ברם, לא כך הם פני הדברים. אי התאמות שכיחות נמצאו באותם מקרים כאשר לשכיר היו כמה משרות במהלך השנה – ב-75 אחוזים ממקרים אלה נערך חישוב שגוי של שכר בחודש ספטמבר. כמו כן, ב-15 אחוזים מהערכים השתולים נזקף שכר ספטמבר כאשר הפרט כלל לא עבד בספטמבר.

זקיפת נתונים באמצעות רגרסיה לא הייתה מדויקת גם כן. השימוש ברגרסיה נועד "לחלץ" ערכים של שכר ברוטו מערכים של שכר נטו, הרי בטופס השאלון ניתן היה לציין שני ערכים אלה או אחד מהם. באותם מקרים כאשר בטופס צוין רק ערך של שכר נטו הופעלה רגרסיה של שכר ברוטו על שכר נטו בארבעה קטעים, עם שיפועים מ-1.175 עד 2.013. שימוש ברגרסיה זו היה מיותר משום שניתן היה להיעזר בפונקציות של מס הכנסה, דמי ביטוח לאומי ודמי ביטוח בריאות ולקבל אומדן מדויק הרבה יותר. הניתוח שנערך מלמד כי במפקד הבא, במידה ואיסוף נתוני שכר בפרט והכנסה בכלל יתבסס על דיווח עצמי של הנפקדים ללא וידוא נכונות הנתונים בשטח, יש סבירות גבוהה לריבוי שגיאות ותת-דיווח.

אי לכך, רצוי לאסוף מידע על הכנסות ממקורות מינהליים, ולהגביל חקירה בשטח לבירור של מעמד הנפקד בעבודה ועצם קבלת קצבאות וגמלאות למיניהן.

נסיים בהערה כי אין להסיק/להקיש מעבודה זו מסקנות אודות איכות הנתונים מסקרי הכנסות שוטפים של הלמ"ס. קיים הבדל מהותי ביותר בין האיסוף והטיפול שניתן לנתונים במפקד 95 לבין שיטות האיסוף הטיפול והעיבוד לנתוני הסקרים השוטפים.

WORKING PAPER SERIES

NO.21

**Analysis of Wage Data from the 1995 Census by using
Wage File of the National Insurance Institute**

Dmitri Romanov* Orly Furman*

July 2006

***Central Bureau of Statistics – Chief Scientist Department**

**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: 72-2-6521340
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.

Internet Site: www.cbs.gov.il

E-Mail: info@cbs.gov.il

Abstract

1995 Census of population income statistics have been key data for many research projects because of the large sample, the ability of record linkages with the 1983 Census data as well as other identifiable data sources, and the matching of economic, demographic, business and social variables.

For the purposes of improvement and completion of wage data gathered by the 1995 census the CBS received from the National Insurance Institute a file, that includes information on annual wages on 1.82 million salaried employees, and work-months during the year. This file is useful to us in verifying wage data reported to the Census and in testing the veracity of revisions and imputations that were carried out in the various stages of the development of the Census wage data.

The wages of 307 thousand employees included in the 1995 Census wage file are lower by 16.5% on average than the wages for those same employees in the NII wage file. Editing of the wage data included five types of corrections, the most common of them were correction by regression (21%) and inserting NII values (8.6%). For 68% of the salaried employees, whose wages were reported in the Census, the original value of the wages was not corrected. For this group the value reported in the Census was 15.6% lower on average from the wage value in the NII file.

Correction of wage data through transplanting NII values was not exact in some of the cases. For employees having several positions in 1995, in 75% of cases the wage calculation was incorrect. The use of regression for imputation was unnecessary because it would have been more appropriate to make use of the tax function in order to estimate gross wages on the basis of net wages, in the case where the latter was reported to the Census and not the former.