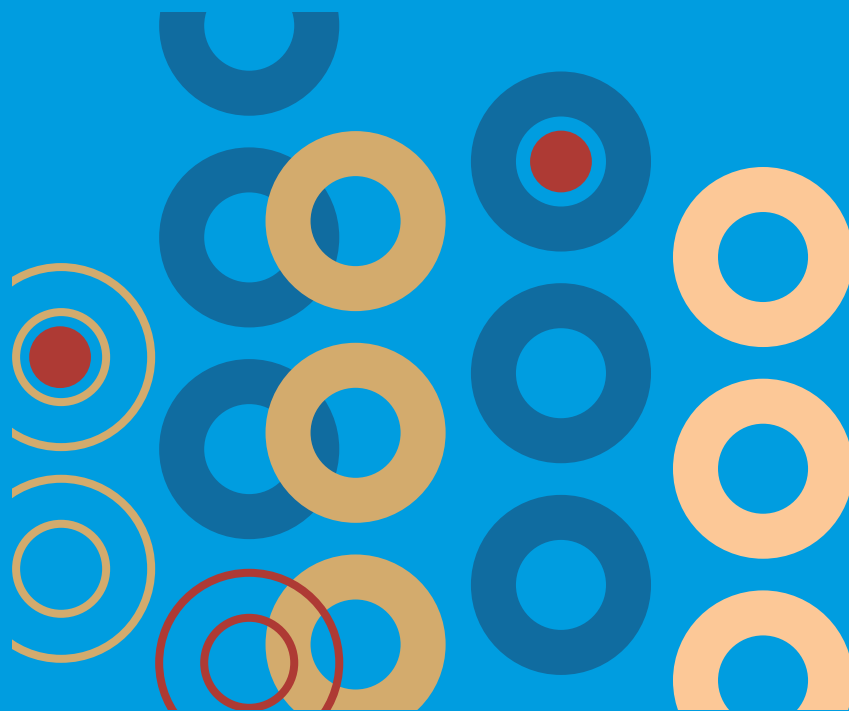




המוסד לביטוח לאומי  
מינהל המחקר והתכנון



## זקיפת רכיבי הכנסה חסרים בנתונים המינהליים

אורן הלה, מירי אנדבלד

**מחקרים לדין 138**  
חשון התשפ"ב, אוקטובר 2021





**המוסד לביטוח לאומי**  
מינהל המחקר והתכנון

## **זקיפת רכיבי הכנסה חסרים בנתונים המינהליים**

**אורן הלר, מירי אנדבלד**

**מחקרים לדיון 138**

ירושלים, חשון התשפ"ב, אוקטובר 2021

[www.btl.gov.il](http://www.btl.gov.il)

## תקציר

בשנים האחרונות הביטוח הלאומי מחשב את מדדי העוני והאי-שוויון לא רק על-פי סקרי משפחות של הלשכה המרכזית לסטטיסטיקה, אלא גם על בסיס הנתונים המנהליים שברשותו. אף שהנתונים המנהליים מקיפים את כלל האוכלוסייה ואין בהם הטיה של מדגמים, חסרים בהם רכיבי הכנסה כמו הכנסה מהון ומתמיכות שמקורן לא בביטוח הלאומי. כדי להשלים את המידע החסר אנו מציגים במחקר הזה זקיפה של הרכיבים החסרים באמצעות מידע מסקרי הלמ"ס על בסיס המאפיינים של משקי הבית שנמצאים בשני מקורות הנתונים ("matching"). המחקר מתאר את השיקולים בבחירת דרך החישוב המתאימה ביותר כדי להגיע לתוצאות מיטביות. הממצאים מראים שלאחר הפעלת הזקיפה של הכנסות חסרות, הממצאים קרובים יותר לאלה המתבססים על סקרי משפחות ומאפשרים לשפר את אומדני העוני והאי-שוויון המופקים מהנתונים המנהליים. אומדנים אלה יעודכנו וישופרו באופן שוטף ככל שיידרש.

## תוכן העניינים

1	.....רקע	.1
2	.....זקיפת הכנסות חסרות בנתונים המנהליים	.2
3	.....א. אמידת רגרסיות באמצעות נתוני הסקרים	.א.
4	.....ב. חישוב משתנים מנובאים בקבצים המנהליים	.ב.
5	.....ממצאים	.3
10	.....נספחים	.4

## 1. רקע

עד 2011 התבססו דוחות העוני של המוסד לביטוח לאומי על סקרי הכנסות של הלשכה המרכזית לסטטיסטיקה, שהכילו מידע על הכנסות משקי הבית מהמקורות השונים ועל המאפיינים הדמוגרפיים והתעסוקתיים של חבריו. ב-2012 בוטל סקר הכנסות ומאז הביטוח הלאומי מבסס על חישובי העוני והאי שוויון (וכן חישובים בנושאים כלכליים חברתיים אחרים) על סקרי הוצאות משק הבית שמכילים בנוסף לנתונים הנ"ל גם נתונים על הוצאות לסוגיהן, אך מאידך מתייחסים למספר משפחות ופרטים קטן בכשליש מהמספר שהיה בסקרי ההכנסות.

בביטוח הלאומי קיים מידע רב על הכנסות הפרטים וכן על המאפיינים הדמוגרפיים והתעסוקתיים שלהם. קיים מידע על הכנסות מעבודה (שכירה ועצמאית), פנסיה וקצבאות הביטוח הלאומי בהיקף מלא.

בשנים האחרונות עסקנו בבניית קבצי נתונים המבוססים על נתונים מינהליים והמאפשרים לנו לחשב מדדי עוני ואי שוויון מקבילים, לצד אלו שמבוססים על סקרי הלמ"ס. למאמץ זה כמה מטרות: ראשית, לנתונים המינהליים יתרונות על פני הסקרים: כיסוי מלא של האוכלוסייה וכן כיסוי מלא ומהימנות גבוהה של רכיבי ההכנסה הנצפים, ולפיכך אפשרות לחשב את המדדים החברתיים לקבוצות אוכלוסייה קטנות, מה שלא ניתן לעשות כאשר מתבססים על מדגם מהאוכלוסייה. לבסוף, מקור כזה עשוי לתת משנה תוקף למקורות הנתונים האחרים כאשר מתבררות מגמות או תוצאות חריגות.

עם זאת לנתונים המינהליים יש גם מגבלות, הנובעים בעיקר מהעדר מידע מלא על הרכב משקי הבית ומהעדר חלק מרכיבי ההכנסה, בהם רווחי הון, העברות כספיות ממשרדי ממשלה שאינם ביטוח לאומי והעברות כספיות ממשקי בית אחרים.<sup>1</sup> על אף ההבדלים, חישובים הראו עקביות רבה במגמות העוני בכלל האוכלוסייה ואף במרבית הקבוצות בתוכה, בעיקר כאשר מדובר בחישובים לנפשות (ופחות לגבי משפחות, שכן, כאמור, אין מידע מספיק בנתונים המינהליים על מנת לזווג פרטים למשפחות, לא כל שכן למשקי בית על פי ההגדרה המקובלת<sup>2</sup>).

המרכיב העיקרי בהכנסות משקי בית הוא הכנסה מעבודה שחלקה מסך ההכנסה ברוטו הוא כ-78% בממוצע, אשר יחד עם רכיבי הכנסה נוספים הזמינים גם בנתונים המינהליים – מפנסיות וקופות גמל ומקצבאות הביטוח הלאומי – מהווים כ-93% מסך ההכנסות של משקי הבית, והשארית – הכנסות מהון וקצבאות ותמיכות ממוסדות אחרים וממשקי בית – חסרה בנתונים המינהליים (לוח 1).

אף על פי שהכנסות מהון וקצבאות ותמיכות ממוסדות אחרים וממשקי בית מהווים נתח לא גדול מסך ההכנסה ברוטו של משקי הבית, ההתפלגות הלא אחידה שלהם באוכלוסייה עשויה לשנות את תמונת האי-שוויון: הכנסות מהון מוטות כלפי בעלי הכנסה גבוהה, כאשר שיעורן מתוך סך ההכנסה ברוטו הוא 1%-1.5% בעשירונים 1-4, 2%-3% בעשירונים 5-9 וכ-9% בעשירון העליון; לעומת זאת, הכנסות מקצבאות ותמיכות ממוסדות אחרים וממשקי בית מוטות כלפי בעלי הכנסה נמוכה עם

---

אנו מודים לניצה קסיר, אלכסנדר גאליה ולז'אנה פרנקין על הערותיהם המועילות, ולרינה פינס על עזרתה בעיבוד הנתונים.

<sup>1</sup> להרחבה אודות שיטת השימוש בנתונים המנהליים, יתרונותיה וחסרונותיה, ראו הלר, אנדבלד וגוטליב "אמידת ממדי העוני והאי שוויון לפי נתונים מנהליים – ישראל, 2003–2014" (2017) באתר המוסד לביטוח לאומי תחת הלשונית פרסומים.

<sup>2</sup> לפי ההגדרה המקובלת משק בית מורכב מאנשים שחולקים דירה או יחידת מגורים משותפת ומנהלים תקציב משותף למזון.

שיעור מתוך סך ההכנסות הולך ופוחת עם ההכנסה הפנויה מ-13.4% בעשירון התחתון עד 1.5% בעשירון העליון (תרשים 1).

## 2. שיטת זקיפת הכנסות חסרות בנתונים המינהליים

### 2א. סוגי ההכנסות ומאפייניהן

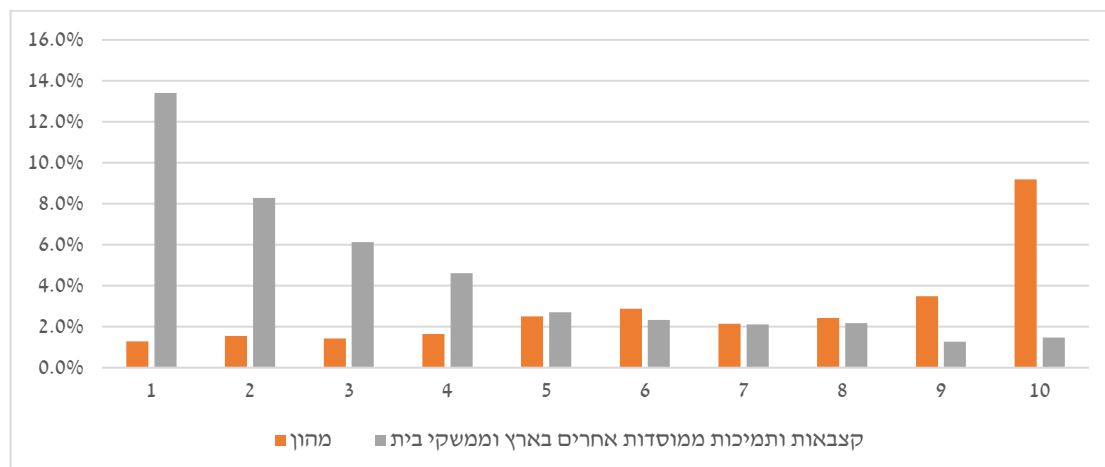
על מנת לדייק את הממצאים העולים מעיבודים שנעשים על הקבצים המינהליים, ובפרט על מנת לצמצם הטיות בהערכות של עוני ואי-שוויון שנובעות מהיעדרם של רכיבי הכנסה, אנו זוקפים את רכיבי ההכנסה החסרים בנתונים המינהליים באמצעות מידע מתוך סקרי הלמ"ס – על בסיס מאפייני משקי הבית שנמצאים בשני מקורות הנתונים ("matching"). מטרת מסמך זה לתאר את השיקולים המנחים בבחירת הדרך המתאימה ביותר לבצע חישוב זה כדי להגיע לתוצאות מיטביות.

לוח 1: רכיבי ההכנסה של משקי בית, 2018

רכיבי הכנסה	ממוצע	שיעור מההכנסה ברוטו	נצפה בנתונים המינהליים
עבודה	15,861	76.3%	כן
הון	937	4.5%	לא
פנסיות וקופות גמל	1,635	7.9%	כן
קצבאות ביטוח לאומי	1,825	8.8%	כן
קצבאות ותמיכות ממוסדות אחרים בארץ	310	1.5%	לא
משקי בית	230	1.1%	לא
<b>סך ההכנסה ברוטו</b>	<b>20,800</b>		
<b>הכנסה פנויה</b>	<b>17,152</b>		

מקור הנתונים: סקר הוצאות 2018

תרשים 1: התפלגות רכיבי ההכנסה החסרים לפי עשירוני ההכנסה, 2018



מקור הנתונים: סקר הוצאות 2018

רכיבי ההכנסה החסרים קובצו לשני משתנים: א. הכנסות מהון; ב. הכנסות ממקורות אחרים – משרדי ממשלה (שאינם ביטוח לאומי) ומשקי בית.<sup>3</sup> כל אחד משני המשתנים הללו נזקף בנפרד ובאופן בלתי תלוי, באמצעות אמידה של שתי רגרסיות הנאמדות בסקרי הוצאות, האחת רגרסיה לוגיסטית המנבאת אם קיימת הכנסה חיובית (שאינה אפס) למשק הבית והשנייה רגרסיה לינארית המנבאת את גודל ההכנסה. אומדי הרגרסיות הללו משמשים לניבוי רכיבי ההכנסה בקבצים המינהליים, כפי שמוסבר בפירוט בהמשך.

יצוין שלצורך הימנעות מאקראיות הקיימת בשנת סקר אחת והישענות על מספר גדול יותר של נדגמים, אומדני הרגרסיות מחושבים על בסיס שלושה סקרי הוצאות (שנים 2016 - 2018).<sup>4</sup> להלן מובאת השיטה בפירוט.

## ב. אמידת רגרסיות באמצעות נתוני הסקרים

התפלגויות ההכנסות מהון וההכנסות האחרות בנויות מריכוז גבוה של משקי בית ללא ההכנסות הללו – רק ל-13.1% יש הכנסה מהון ול-17.9% יש הכנסה ממקורות אחרים (לפי סקרי הוצאות של השנים 2016 עד 2018) – ומהתפלגות לוג נורמאלית בקירוב בקרב בעלי הכנסה (ראו איורים בנספח 1). לכן אמידת ההכנסות החסרות בנויה משני נדבכים: באחד מנובא קיום ההכנסה ובשני מנובא גודלה.

כך, בשלב הראשון אנו מבחינים בין בעלי הכנסה ונעדרי הכנסה באמצעות אמידת הרגרסיה הלוגיסטית הבאה על נתוני הסקר:

$$(1) P_i^S = P(Y_i^S > 0) = \beta X^S + \varepsilon$$

כאשר  $Y$  הוא ההכנסה; הסימון  $i$  הוא לציון הכנסות מהון / הכנסות אחרות; הסימון העילי  $S$  מציינ שמקור הנתונים הוא סקר (Survey); ו- $X$  הוא וקטור של משתנים מסבירים הנצפים גם בסקר וגם בקובץ המינהלי וכולל את הנתונים הבאים עם אינטראקציות שונות ביניהם: מצב תעסוקתי; לוג הכנסות מעבודה ביחס למוצע ההכנסות מעבודה; קבוצות גיל; נפת מגורים; מספר נפשות במשק הבית; מספר ילדים; מגדר; מצב משפחתי; שיוך לקבוצת אוכלוסייה – יהודים / ערבים / חרדים; עולה; קבלת פנסיה וגובה הפנסיה ביחס למוצע ההכנסות מעבודה; יבשת לידה.<sup>5</sup>

בשלב השני נאמדת באמצעות נתוני הסקר הרגרסיה הלינארית הבאה:

$$(2) L_i^S = \log\left(\frac{Y_i^S}{\bar{w}^S}\right) = \gamma X^S + u \mid Y_i^S > 0$$

כאשר המשתנה המנובא הוא הלוגריתם הטבעי של היחס בין ההכנסה (מהון או אחרת) ושל  $\bar{w}$  – ממוצע ההכנסות מעבודה. הפיכת משתני ההכנסה – גם המנובאת וגם מבין המשתנים המסבירים

<sup>3</sup> מטעמי נוחות, הכנסות ממשרדי ממשלה אחרים והכנסות ממשקי בית אחרים אוחדו בעבודה זו לכדי משתנה אחד. שני המשתנים הללו מתפלגים בצורה דומה בין עשירי הכנסה.

<sup>4</sup> סקרי הוצאות של השנים 2016 עד 2018 נבחרו כבסיס הנתונים עליו מבוסס המודל הואיל ובעת כתיבת מאמר זה אלו שלושת סקרי הוצאות העדכניים ביותר הזמינים. יחד עם זאת, תוצאות המודל אינן רגישות לשינוי בבחירת שנות הסקרים – ראו מבחן רגישות בנספח 6.

<sup>5</sup> ראו רשימה של המשתנים המסבירים ואת תוצאות הרגרסיות 1 ו-2 (מטעמי נוחות – ללא משתני האינטראקציה) בנספחים 2 ו-3 בהתאמה.

– מאפשרת שימוש במקדמי הרגרסיה שנאמדים בסקר לניבוי ההכנסה בקבצים המינהליים גם עבור שנים שונות מאלה שהסקרים מתייחסים אליהן, ללא צורך בקידום מחירים.<sup>6</sup>

## ג. חישוב משתנים מנובאים בקבצים המינהליים

באמצעות הערכים שנאמדו על בסיס נתוני הסקרים בשלבים הקודמים, מחושבת בנתונים המינהליים ההסתברות להכנסה חיובית:

$$P_i^A = \hat{\beta} X^A$$

כאשר  $\hat{\beta}$  נאמדה לעיל והסימון העילי  $A$  מצייין שמקור הנתונים הוא מינהלי (Administrative).  $P_i^A$  הוא הסיכוי המנובא לכך שמשק בית  $i$  הוא בעל הכנסה חיובית (שאינה אפס), והוא משמש אותנו להלן על מנת לסמן את משק הבית כבעל הכנסה חיובית או כנעדר הכנסה. באופן דומה מחושבת ההכנסה המנובאת:

$$L_i^A = \hat{\gamma} X^A$$

באמצעות המשתנים מחושבים  $P_i^A$  ו- $L_i^A$  אנו יוצרים בקבצים המינהליים את המשתנים המנובאים עבור רכיבי ההכנסה החסרים, באופן הבא:

$$Y_i^A = \begin{cases} \exp(L_i^A + v) \cdot \bar{w}^A, & P_i^A \leq U \\ 0, & \text{else} \end{cases}$$

כאשר  $U$  הוא משתנה אקראי בעל התפלגות אחידה בין 0 ו-1, שבאמצעותו משק הבית  $i$  מסומן כבעל הכנסה חיובית, בהתאם להסתברות המנובאת  $P_i^A$  שיש לו הכנסה; ו- $v$  הוא משתנה אקראי בעל התפלגות נורמלית ושונות זהה לזו של השארית  $\hat{u}$  שנאמדה במשוואה (2).

עבור מי שסומן כבעל הכנסה חיובית, נזקפת הכנסה הבנויה מסכום של הגודל המנובא  $L_i^A$  ושל הפרמטר  $v$  אשר מוסיף "רעש" כנגד השונות הבלתי מוסברת על ידי  $X$  (במשוואה 2) וזאת על מנת לשמר את השונות המקורית של רכיב ההכנסה כפי שהייתה בסקר. הסכום  $L_i^A + v$  הוא הערך המנובא של לוג המנה של ההכנסה החסרה מחולקת בממוצע ההכנסות מעבודה. לכן, מכפלה של ממוצע ההכנסות מעבודה עם האקספוננט של הסכום  $L_i^A + v$  מניב את הערך החזוי של רכיב ההכנסה ביחידות של שקלים.

יצוין שהשכר, כמשתנה מנבא חשוב, הוכנס לרגרסיה בשתי צורות. פעם כמשתנה רציף, ופעם כמשתנה קטגוריאל. מקדמי הרגרסיה היו דומים, וכך גם תוצאות אומדני העוני (ראו נספח 8 על שני חלקיו).

<sup>6</sup> השימוש במודל הוא כאמור על בסיס שלושה סקרי הוצאות מאוחדים, כלומר על בסיס מידע שמייצג שלוש שנים. בהתאם לכך, בחישוב המנה של המשתנה המנובא מחולק בשכר הממוצע, אנו עושים שימוש בשכר הממוצע של השנה הרלוונטית.



### 3. ממצאים ומסקנות

לוח 2 להלן מראה את קווי העוני ב-2018, לפי הסקר ולפי הנתונים המינהליים – לפני ואחרי זקיפת רכיבי ההכנסה החסרים. קו העוני לפי נתוני סקר הוצאות 2018, הכולל את כלל ההכנסות, גבוה בכ-400 ₪ לחודש לעומת הקו לפי הנתונים המינהליים כאשר אין כוללים את ההכנסות הזקופות. הכללתן מעלה את קו העוני בכ-200 ₪ - והוא נותר רחוק בסכום דומה לעומת קו העוני לפי סקר ההוצאות. המשמעות היא שהזקיפות של ההכנסות החסרות תיקנו חלקית את קו העוני אך ההבדלים הבסיסיים בין שני מקורות הנתונים אינם מאפשרים להביא לזהות של קווי העוני בשני מקורות הנתונים.

לוח 2 : קווי העוני לשנת 2018

במונחים שנתיים	במונחים חודשיים	סקר הוצאות נתונים מינהליים	לפני זקיפות כולל זקיפות
2,875	2,435	29,220	31,959
2,663			

לוח 3 להלן מראה כי לאחר אמידת ההכנסה מהון וההכנסה ממוסדות אחרים ומשקי בית (להלן: "תמיכות אחרות"), שיעור בעלי ההכנסה החיובית כפי שמתקבל בבסיס הנתונים המינהלי מגיע ל-14.8% בהשוואה ל-13.1% ול-13.8% בנתוני הסקרים – של השנים 2016-2018 אשר על בסיסם התבצעה הסימולציה ו-2018 לבדה, בהתאמה – לגבי ההכנסה מהון. במקרה של ההכנסות האחרות, ל-19.6% מהמשפחות בנתונים המינהליים יש הכנסה חיובית, בהשוואה ל-17.9% ו-17.1% בסקרים – בשנים 2016-2018 ובשנת 2018 לבדה, בהתאמה. גם ממוצעי ההכנסה בכל אחד משני הסעיפים קרובים לאלה המתקבלים על פי סקרי ההוצאות של הלמ"ס.<sup>7</sup>

לוח 3 : הכנסה נזקפת בנתונים המינהליים בהשוואה לסקר, ממוצעים בכלל האוכלוסייה

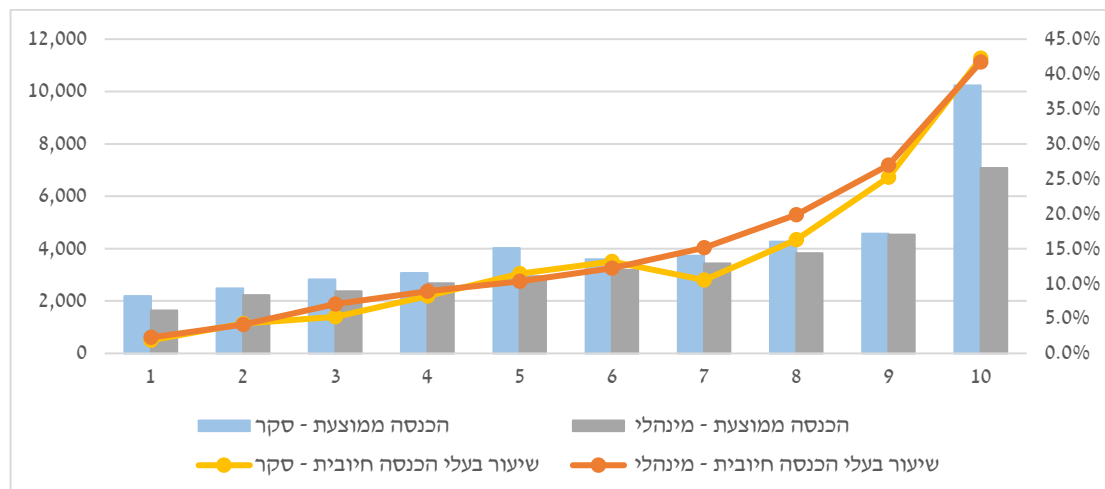
הכנסה ממוסדות אחרים ומשקי בית			הכנסה מהון			
מינהלי 2018	סקר 2018	סקר 2016-2018	מינהלי 2018	סקר 2018	סקר 2016-2018	
19.6%	17.1%	17.9%	14.8%	13.8%	13.1%	שיעור משקי בית בעלי הכנסה חיובית ממוצע למשפחה בעלת הכנסה*
2,376	2,587	2,635	4,528	5,815	5,117	

\* במחירי 2018

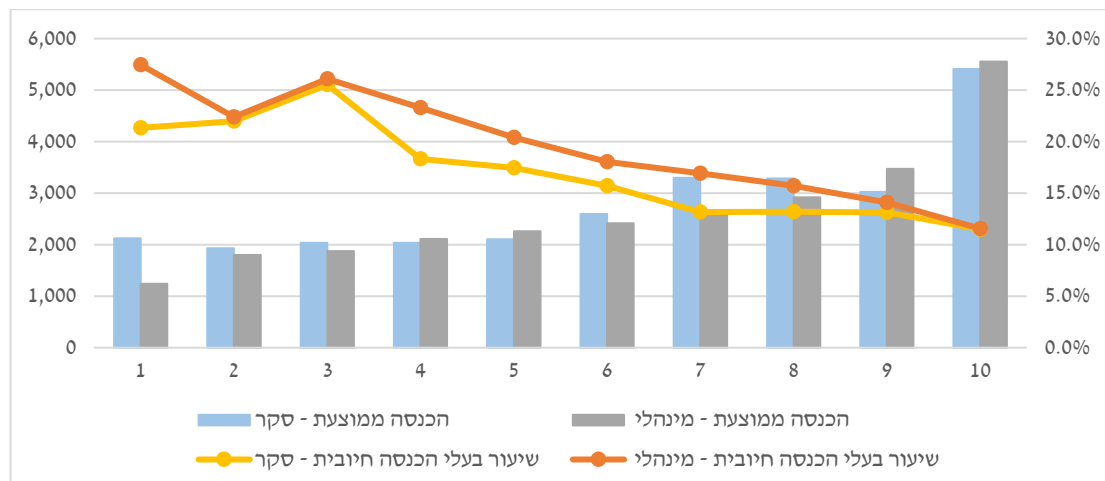
<sup>7</sup> הפער בשיעור בעלי הכנסה – מהון וממקורות אחרים – בין שני מקורות הנתונים נובע מהבדלים בהרכב משקי הבית ובמאפייניהם. על מנת לאמוד את ההשפעה של חלק זה של המודל על מדדי העוני, ערכנו בדיקה בה הגדלנו את ההסתברות של כל משפחה לקבל הכנסה חיובית כך ששיעור המשפחות בעלות הכנסות אלה יהיה זהה לזה שבסקר. בבדיקה זו השתנתה תחולת העוני אך במעט: תחולת העוני של משפחות עלתה בכרבע נקודות אחוז ושל נפשות ב-0.13 נקודות אחוז בהשוואה לשימוש במודל המתואר במאמר ללא שינוי זה.

תרשים 2: רכיבי ההכנסה הנזקפת – סכום ממוצע ושיעור בעלי ההכנסה, לפי עשירוני הכנסה, סקרים בהשוואה לנתונים המינהליים

תרשים א2: הכנסה מהון



תרשים א2: הכנסה ממוסדות אחרים ומשקי בית



גם חישוב לפי עשירוני הכנסה (תרשים א2 לגבי הכנסה מהון ותרשים ב2 לגבי תמיכות אחרות), מצביע על דמיון בשיעור המקבלים לאורך העשירוניים. שיעור בעלי הכנסה מהון עולה עם העשירוניים, לעומת שיעור בעלי הכנסה הולך ופוחת כשמדובר בהכנסה מתמיכות אחרות. עם זאת בשני המקרים, ואף בשני בסיסי הנתונים, מתקבל כי גובה ההכנסה למקבל עולה עם העלייה בעשירוניים.

לוח 4 מראה את מדדי העוני לשנת 2018 לפי הסקר ולפי הנתונים המינהליים, עם וללא רכיבי ההכנסה הנזקפים. תחולת העוני של נפשות לפי הסקר נמוכה ב-0.9 נקודות האחוז מזו המתקבלת מבסיס הנתונים המינהלי, וזקיפת ההכנסות החסרות מצמצמת את הפער למחצית נקודת האחוז (תחולת העוני של משפחות לא משתנה משמעותית לאחר הזקיפות ב-2018). עובדה זו מלמדת כי עיקר ההבדלים בתחולת העוני המוחלטות (להבדיל מהמגמות לאורך זמן), נעוצים בהגדרות

השונות של "משק בית" ולא בהכנסות החסרות בנתונים המינהליים. גם שאר המדדים המוצגים בלוח 3 מתקרבים עם הזקיפות לשיעורים שמתקבלים לפי סקר הוצאות, אך בולטת תחולת העוני של קשישים, שבניגוד למדדי העוני האחרים הגבוהים יותר בנתונים המינהליים, היא נמוכה יותר בנתונים המינהליים בהשוואה לסקר - בכ-7 נקודות אחוז, ולאחר הזקיפות הפער מצטמצם לכ-4 נקודות אחוז. ממצא זה עשוי להפתיע מאחר שהכנסות הקשישים מהון גבוהות יותר מאשר בכלל האוכלוסייה. עם זאת בדיקה מעלה שהכנסות אלה, שהן אכן גבוהות יותר אצל הקשישים, מתחלקות באופן לא שיוויוני כך שהעניים נהנים מהן הרבה פחות מהקשישים המבוססים יותר, ולכן העלאת קו העוני הנובעת מתוספת ההכנסה הזקופה גוברת בהשפעתה על תוספת ההכנסה לקשישים חסרי האמצעים.<sup>8</sup>

נספח 5 מציג את התפלגות ההכנסה הפנויה לנפש תקנית של הסקר בהשוואה לנתונים המינהליים ללא הזקיפות ועם הזקיפות (נספח 5א ו-5ב בהתאמה), ומדגים את התקרבות התפלגות ההכנסות לפי הנתונים המינהליים לזה שמתקבל לפי הסקר, עם הוספת רכיבי ההכנסה הנזקפים. תורמת לכך ההפחתה של מספר משקי הבית בעלי הכנסה אפס בנתונים המינהליים כתוצאה מזקיפת רכיבי ההכנסה למשקי הבית הללו.

**לוח 4: מדדי העוני, לפי סקר הוצאות ולפי הנתונים המינהליים, לשנת 2018 – עם וללא רכיבי ההכנסה הנזקפים**

מינהלי	סקר		
	אחרי זקיפות	לפני זקיפות	
21.4	21.5	18.0	תחולת העוני משפחות
21.7	22.1	21.2	תחולת העוני נפשות
29.4	30.6	30.0	תחולת העוני ילדים
19.2	15.7	23.4	תחולת העוני קשישים
39.5	45.5	33.3	יחס פער ההכנסות
0.3771	0.3894	0.3559	מדד גיני

לוח 5 מציג את תחולת העוני לפי קבוצות אוכלוסייה, לפני ואחרי זקיפת הכנסות אחרות בבסיס הנתונים המינהלי, ובהשוואה לסקרי הוצאות. הנתונים מראים כי למעט אוכלוסיית הקשישים, שאצלה העוני גדל עם הוספת הזקיפות מהכנסות חסרות, במרבית הקבוצות תחולת העוני קטנה לאחר הוספת רכיבי הכנסה אלו. בולטים במיוחד בהקשר זה החרדים, החד הוריות (שיחסית רבות מהן מקבלות סיוע ממשרד השיכון וחלקן אף ממשרד הביטחון כאלמנות צה"ל), מי שאינם עובדים בגיל העבודה (שסביר להניח שמקבלים אף הם תמיכה כלשהי, ממשרדי ממשלה או מקרובים, בייחוד כאשר אין תמיכה מהמוסד לביטוח לאומי) וכן אוכלוסיית הצעירים מאוד.

נספח 7 מציג את תחולת העוני לפי הנתונים המינהליים, ללא ההכנסות הנזקפות ולאחר שהן מובאות בחשבון, מ-2003 עד 2018. בכל השנים הללו זקיפת משתני ההכנסה החסרים משפיעים על תחולת העוני בכיוונים דומים לאלה שהוצגו לעיל עבור שנת 2018: תחולת העוני של משפחות,

<sup>8</sup> על מנת לוודא שתוצאה זו תקפה ערכנו בדיקה נוספת בה הוספנו לרגרסיות שבמודל משתנה אינטראקציה של קשיש עם לוג ההכנסות מעבודה, זאת בנוסף לאינטראקציה של קבוצת גיל (ובכלל זה קשישים) עם מצב תעסוקתי ויבשת לידה וכן גובה ההכנסות מפנסיה שרלוונטי בעיקר לקשישים – כפי שמפורט בנספח 2. לשינוי זה במודל השפעה זניחה על מדדי העוני: תחולת עוני משפחות יורדת ב-0.06 נקודות אחוז ותחולת עוני נפשות יורדת ב-0.03 נקודות אחוז.

נפשות וילדים יורדת – אם כי תחולת העוני של משפחות יורדת ברמה משמעותית יותר בשנים עברו בהשוואה ל-2018 – בעוד תחולת העוני של קשישים עולה עם זקיפת ההכנסות רק משנת 2015. נספח 8 מראה את התוצאות המתקבלות כאשר משתמשים בשכר כמשתנה קטגוריאלי במקום כלוג רציף. ההבדל בתוצאות הוא מזערי ולכן הושאר החישוב על כנו (עם זאת התווסף משתנה של "שכר מעל 20 אלף ₪ כמשתנה דמי).

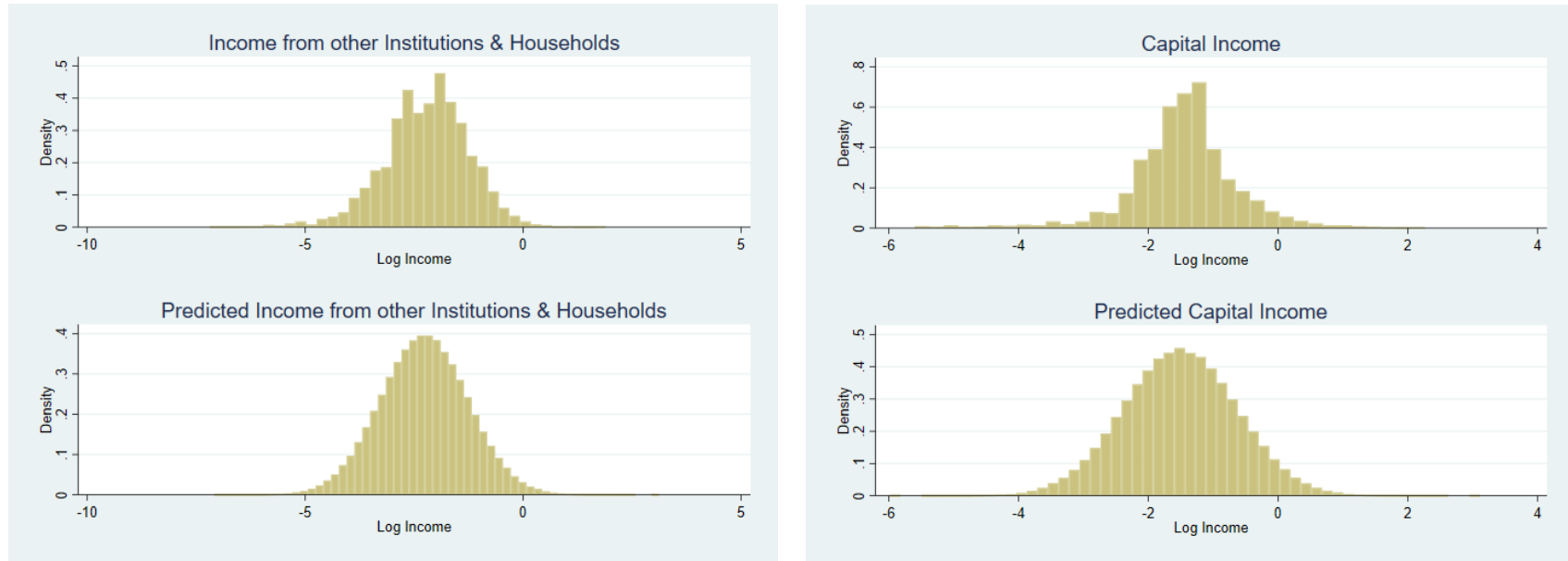
לסיכום, על אף המגבלות שנסקרו, עבודה זו מקרבת את הנתונים המינהליים למקור הנתונים המקובל לצורך חישובי עוני ומדדים חברתיים (סקרי משפחות) בכך שהיא מאפשרת הסתכלות על מושג הכנסה רחב יותר מאשר זה הקיים בנתונים המינהליים ללא הזקיפה של הכנסות חסרות, ובכך היא מאפשרת לטייב את אומדני העוני והאי שוויון המופקים בהתבסס על הנתונים המינהליים של הביטוח הלאומי. לבסוף, האומדנים יעודכנו וישופרו גם בהמשך ככל שיימצא צורך.

לוח 5: תחולות העוני, לפי סקר הוצאות ולפי הנתונים המינהליים, לשנת 2018 – עם וללא רכיבי ההכנסה הנזקפים, לפי קבוצות אוכלוסייה

	מינהלי				סקר							
	אחרי זקיפות		לפני זקיפות		אחרי זקיפות		לפני זקיפות					
	ילדים	נפשות	משפחות	קשישים	ילדים	נפשות	משפחות	קשישים				
סה"כ	17.1	29.4	21.7	21.4	13.6	30.6	22.1	21.5	17.8	30.0	21.2	18.0
יהודים	15.7	22.4	16.9	18.0	12.8	25.0	18.2	18.7	14.1	21.2	14.5	13.4
חרדי	13.3	51.7	43.5	36.9	13.7	60.1	51.3	44.2	19.4	60.4	51.8	42.3
עולים מ-1990	35.2	24.1	21.0	24.5	25.9	25.0	19.8	22.2	21.6	21.8	15.3	16.0
ערבים	31.8	50.5	39.8	38.5	21.3	47.6	36.8	35.6	55.0	57.8	47.4	45.3
משפחות עם ילדים	14.2	29.4	25.2	22.4	13.9	30.7	26.1	23.5	24.3	30.0	25.5	21.4
מספר הילדים 1-3	11.9	18.5	17.0	17.5	11.7	19.3	17.5	18.5	21.0	16.8	15.6	15.6
מספר הילדים 4+	45.2	48.8	47.5	45.6	43.0	51.0	49.5	47.2	50.0	54.1	53.0	49.9
מספר הילדים 5+	52.5	58.1	57.1	56.2	51.4	61.6	60.6	59.3	71.4	67.9	66.7	66.2
חד-הוריות	16.8	29.1	25.8	23.6	19.5	38.3	34.2	31.8	22.7	32.9	28.0	25.1
עובד	3.3	26.5	18.3	15.4	3.2	27.4	19.0	16.5	5.1	26.3	17.4	12.3
שכיר	3.1	25.5	17.7	15.1	3.1	26.3	18.3	16.0	5.1	26.2	17.4	12.4
עצמאי	2.8	22.7	15.2	12.5	2.4	22.8	15.2	12.8	5.2	27.1	17.4	11.8
לא עובד בגיל עבודה	32.2	85.2	74.6	68.3	28.2	95.1	81.1	73.0	31.3	92.0	82.3	71.0
מפרנס אחד	4.6	57.0	36.7	24.0	4.5	60.6	39.3	26.4	6.7	60.9	41.2	24.1
שני מפרנסים ויותר	1.2	15.1	10.4	7.9	1.0	15.0	10.2	7.7	2.2	11.4	7.8	5.3
רמ"ב עד גיל 29	61.2	51.0	43.6	42.4	57.1	55.3	48.2	47.6	9.1	41.3	25.1	22.2
רמ"ב בגילאים 30-44	24.4	30.9	26.3	21.1	22.2	31.9	27.0	21.9	6.3	30.1	23.2	16.7
רמ"ב בגילאים 45-גיל הפנסיה	5.5	19.8	13.3	14.0	5.1	20.7	13.6	14.5	8.2	21.7	14.7	12.5
קשישים לפי הגדרה ישנה	17.4	22.0	15.5	19.2	13.8	22.6	12.8	15.6	20.4	70.9	23.3	23.4
בגיל פנסיוני לפי חוק	17.7	23.5	16.5	20.0	14.0	23.2	13.3	15.9	20.8	78.1	24.9	24.6

## 4. נספחים

נספח 1 : התפלגויות רכיבי ההכנסה בסקרים



התפלגויות הלוגריתם הטבעי של ההכנסה מהון (איורים ימניים) וההכנסה ממוסדות אחרים וממשקי בית (איורים שמאליים), בניכוי השכר הממוצע, בקרב משקי בית שיש להם הכנסות אלה.

האיורים העליונים מראים את ההתפלגויות בסקרי הוצאות 2016-2018 עליהם מבוסס המודל.

האיורים התחתונים מראים את תוצאות המודל – ההתפלגויות של המשתנים המחושבים בקבצים המינהליים של שנת 2018.

## נספח 2: רשימת המשתנים המנבאים

להלן רשימת המשתנים המנבאים, לרבות משנתי האינטראקציה, באמידת הרגרסיות הלוגיסטיות והלינאריות (משוואת 1 ו-2, בהתאמה) לניבוי הן ההכנסות מהון והן ההכנסות ממקורות אחרים.

- האם עובד
- קבוצת גיל; האם עובד X קבוצת גיל
- לוג הכנסות מעבודה (פחות לוג השכר הממוצע); ריבוע לוג הכנסות מעבודה (פחות לוג השכר הממוצע)
- נפת מגורים; האם עובד X נפת מגורים
- גודל משפחה
- משפחה עצמאית (חד הורית)
- האם יש ילדים; מספר ילדים; מספר ילדים עד גיל 3
- האם יש ילדים X משפחה עצמאית (חד הורית)
- מספר ילדים X משפחה עצמאית (חד הורית)
- מגדר
- מצב משפחתי
- חרדי
- ערבי
- מגדר X מצב משפחתי
- מגדר X מצב משפחתי X חרדי
- מגדר X מצב משפחתי X ערבי
- עולה
- מקבל פנסיה
- לוג גובה הפנסיה (פחות לוג השכר הממוצע)
- יבשת לידה; יבשת עלייה X קבוצת גיל
- האם השכר מעל 20 אי"ש

נספח 3 : תוצאות רגרסיות המודל, על בסיס סקרי הוצאות 2016 – 2018

הכנסות אחרות		הכנסות מהון		
רגרסיה לינארית לניבוי גודל ההכנסה	רגרסיה לוגיסטית לניבוי קיום ההכנסה	רגרסיה לינארית לניבוי גודל ההכנסה	רגרסיה לוגיסטית לניבוי קיום ההכנסה	
0.681 (0.690)	-1.036 (0.971)	1.330* (0.765)	2.936*** (1.031)	האם עובד
-0.0634* (0.0366)	-0.743*** (0.0495)	0.127*** (0.0307)	0.336*** (0.0431)	לוג הכנסות מעבודה
0.0975*** (0.0181)	0.0742*** (0.0228)	0.0678*** (0.0204)	0.0274 (0.0248)	מספר נפשות במשפחה
0.0310 (0.222)	1.339*** (0.329)	1.053*** (0.302)	0.0874 (0.429)	משפחה עצמאית (חד הורית)
-0.281** (0.111)	-0.328** (0.164)	-0.0452 (0.132)	0.0927 (0.176)	האם יש ילדים
-0.0350 (0.0261)	0.0873** (0.0372)	-0.0514* (0.0309)	0.0564 (0.0399)	מספר הילדים
0.196*** (0.0752)	0.138 (0.111)	-0.0247 (0.0916)	-0.158 (0.122)	מספר ילדים עד גיל 3
0.0778 (0.0629)	0.0666 (0.0804)	-0.0761 (0.0758)	-0.177* (0.0972)	נשים
0.200 (0.126)	0.349** (0.163)	0.114 (0.142)	-0.0645 (0.186)	אלמן
-0.276 (0.193)	0.826*** (0.284)	0.103 (0.380)	-0.613 (0.481)	חרדי
0.0335 (0.0949)	0.233** (0.118)	0.0471 (0.0942)	0.00642 (0.125)	גרוש
-0.0159 (0.0654)	0.202** (0.0825)	0.0545 (0.0698)	-0.162* (0.0898)	נשוי
0.0948 (0.148)	-1.321*** (0.174)	0.180 (0.282)	-2.248*** (0.337)	ערבי
-0.372*** (0.0546)	0.383*** (0.0694)	-0.150** (0.0712)	-0.841*** (0.0830)	עולה
-0.141* (0.0801)	-0.183* (0.0969)	-0.111* (0.0644)	-0.245*** (0.0890)	מקבל פנסיה
0.0129* (0.0012)	-0.0221** (0.0091)	0.00385 (0.0012)	0.0704*** (0.0012)	לוג גובה הפנסיה
-0.372*** (0.0012)	0.383*** (0.0012)	-0.150** (0.0012)	-0.841*** (0.0012)	האם השכר מעל 20 א"ש
0.00244 (0.0639)	-0.0725 (0.0770)	-0.0495 (0.0543)	0.153** (0.0721)	יבשת לידה [בהשוואה לאירופה וצפון אמריקה] אפריקה או אסיה
-0.195 (1.038)	-0.0966 (1.175)	-0.421 (0.290)	-1.017 (1.147)	ישראל
-1.028 (1.163)	0.0132 (0.789)	-0.105 (0.196)	-1.498 (1.112)	לא ידוע
0.0260 (0.436)	1.382*** (0.507)	0.413 (0.289)	-0.392 (0.442)	פיקוח על קבוצת גיל
V	V	V	V	פיקוח על נפת מגורים
-3.125*** (0.590)	-5.028*** (1.250)	-0.833*** (0.216)	2.110*** (0.336)	חותך
5,135 0.123	26,703 0.204	3,505 0.118	26,553 0.109	מספר תצפיות R <sup>2</sup> Pseudo R <sup>2</sup>

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

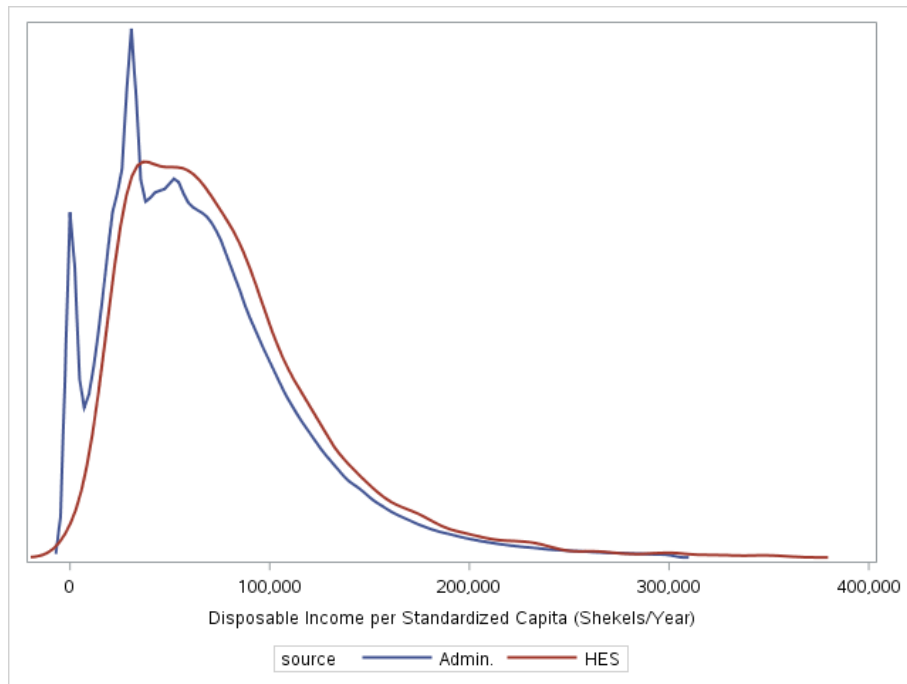


נספח 4 : הכנסה נזקפת בנתונים המינהליים בהשוואה לסקר, לפי קבוצות אוכלוסייה

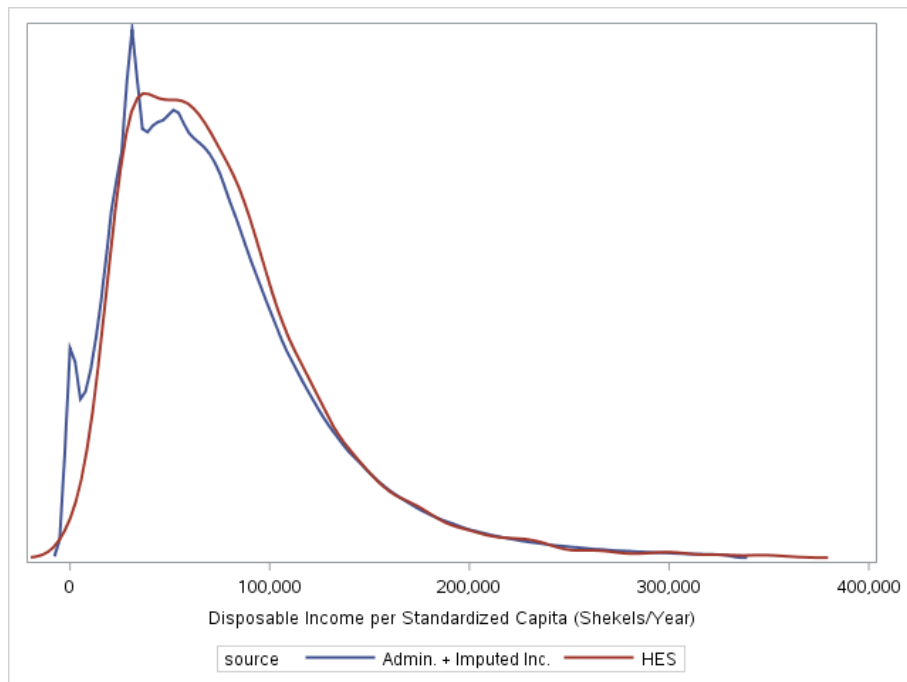
הכנסות ממוסדות אחרים וממשקי בית מינהלי				הכנסות מהון				סקר	סקר
מינהלי		סקר		מינהלי		סקר			
גובה ההכנסה החודשית	שיעור משקי בית בעלי הכנסה	גובה ההכנסה החודשית	שיעור משקי בית בעלי הכנסה	גובה ההכנסה החודשית	שיעור משקי בית בעלי הכנסה	גובה ההכנסה החודשית	שיעור משקי בית בעלי הכנסה		
2,376	19.6%	2,587	17.1%	4,528	14.8%	5,815	13.8%	סה"כ	
2,403	22.0%	2,615	19.2%	4,584	16.1%	5,863	15.7%	יהודים	
2,806	58.4%	2,575	52.2%	4,574	17.7%	4,689	14.9%	חרדים סובייקטיבי	
1,732	25.8%	2,316	22.5%	3,901	7.8%	5,629	8.8%	עולים מ-1990	
1,965	7.4%	1,996	5.2%	3,981	8.3%	4,328	3.0%	ערבים	
2,675	21.2%	2,571	18.0%	4,564	12.7%	5,796	12.4%	משפחות עם ילדים	
2,600	19.4%	2,566	15.8%	4,513	12.4%	5,976	12.0%	מספר הילדים 1-3	
2,912	29.3%	2,584	28.4%	4,775	14.0%	5,063	14.3%	מספר הילדים 4+	
3,180	38.1%	2,748	38.4%	4,864	15.0%	6,231	13.4%	מספר הילדים 5+	
3,104	48.4%	2,698	43.1%	5,011	14.3%	14,252	11.1%	חד-הוריות	
2,302	18.1%	2,495	15.4%	4,552	14.9%	5,801	13.8%	עובד	
2,294	18.1%	2,454	15.5%	4,491	14.7%	6,021	13.2%	שכיר	
2,383	14.5%	2,765	14.9%	5,118	17.8%	4,741	17.5%	עצמאי	
2,872	41.4%	3,086	37.2%	5,069	10.8%	4,629	8.9%	לא עובד בגיל עבודה	
2,287	24.0%	2,727	22.3%	4,220	13.1%	5,052	12.1%	מפרנס אחד	
2,326	13.0%	2,219	11.2%	4,786	16.4%	6,166	14.8%	שני מפרנסים ויותר	
2,092	36.0%	2,218	23.4%	3,591	6.8%	5,531	10.0%	רמ"ב עד גיל 29	
2,539	19.8%	2,717	17.4%	4,255	10.0%	5,220	10.9%	רמ"ב בגילאים 30-44	
2,593	15.6%	2,817	12.6%	4,722	18.7%	5,751	16.4%	רמ"ב בגילאים 45-גיל הפנסיה	
2,309	15.3%	2,556	17.8%	4,577	20.6%	6,483	18.3%	קשישים לפי הגדרה ישנה	
2,198	15.5%	2,528	18.3%	4,622	19.9%	6,599	18.0%	בגיל פנסיוני לפי חוק	

נספח 5 : התפלגות ההכנסה הפנויה לנפש תקנית

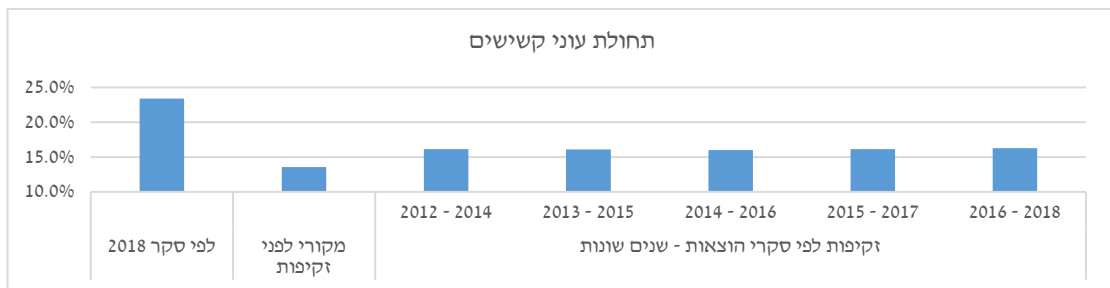
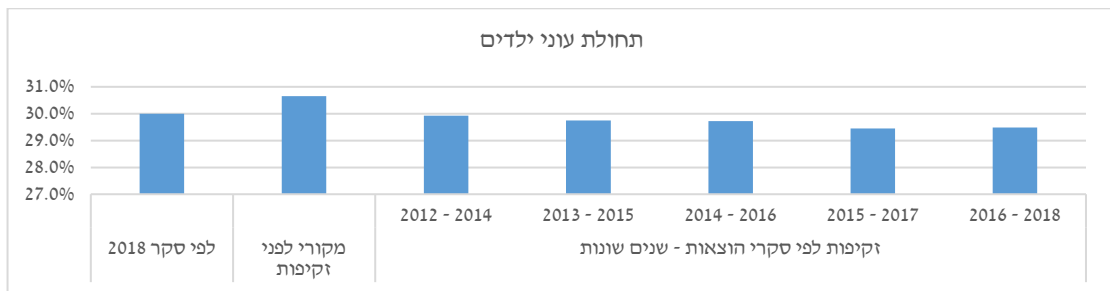
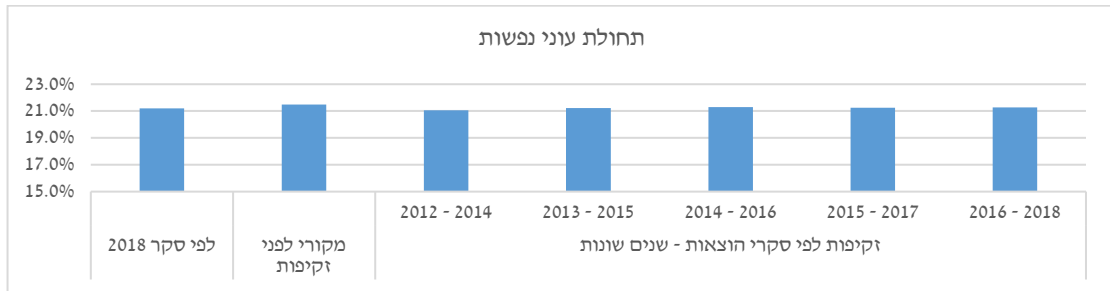
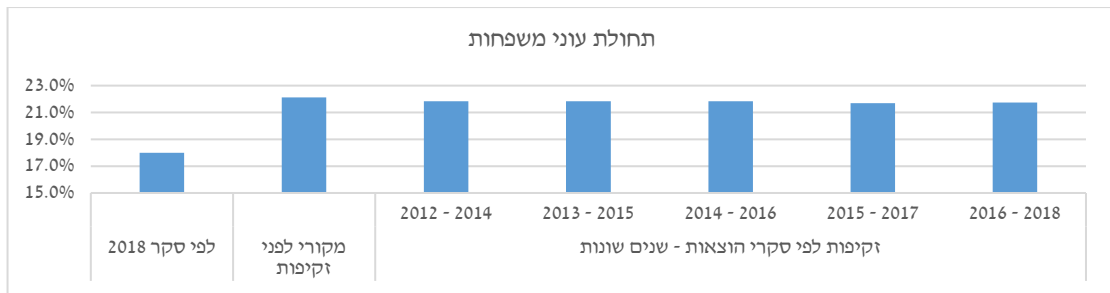
נספח 5א : התפלגות ההכנסה הפנויה לנפש תקנית – נתוני סקר בהשוואה לנתונים מינהליים לפני זקיפת הכנסות חסרות, 2018

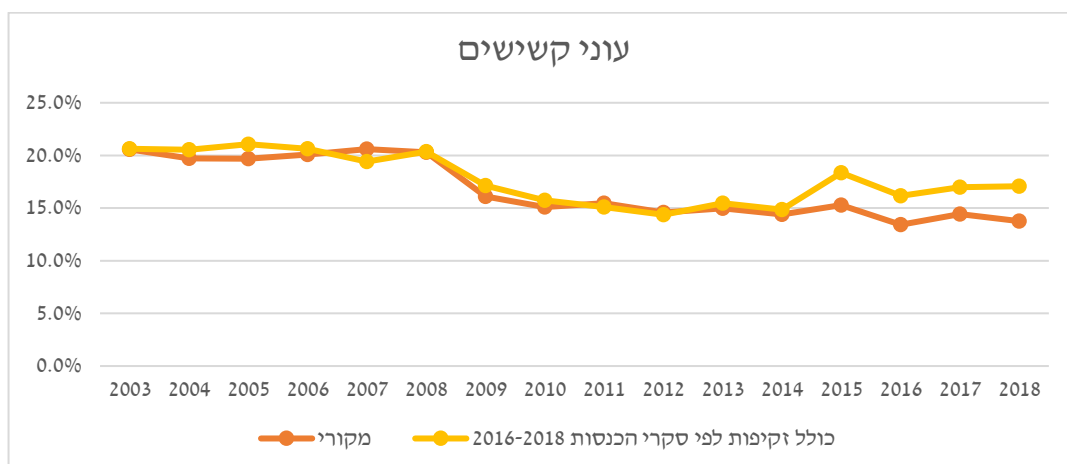
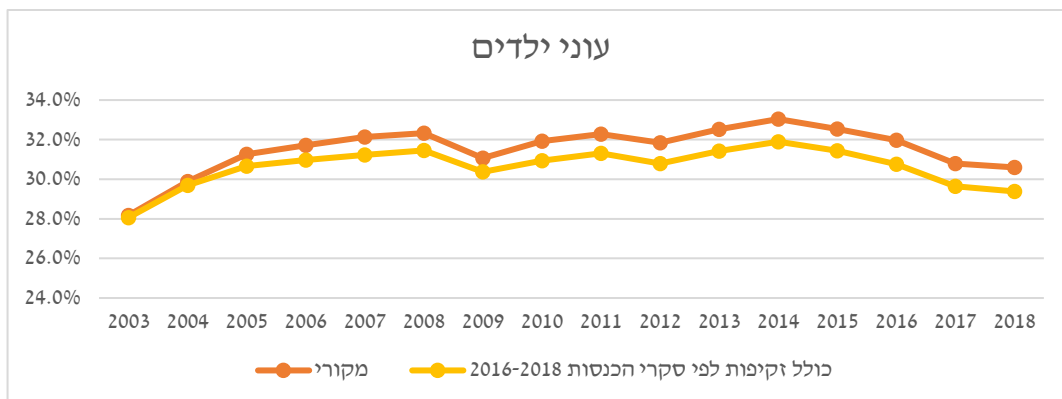
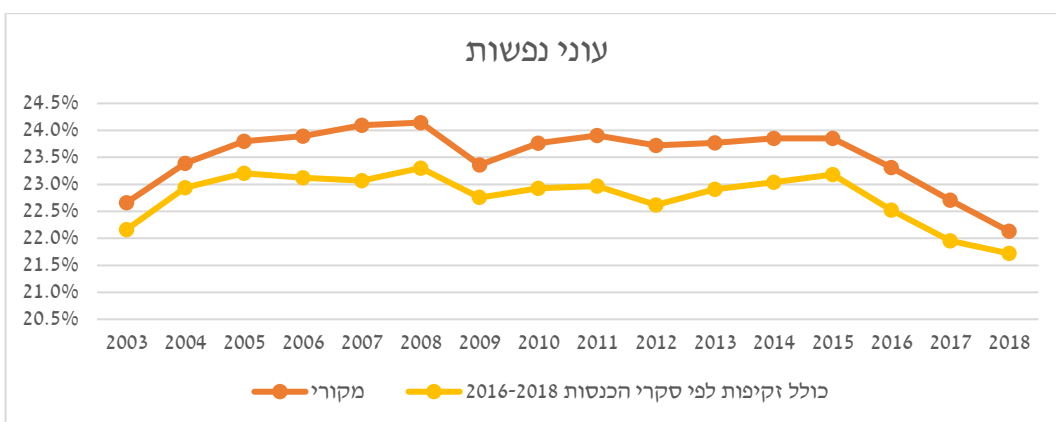
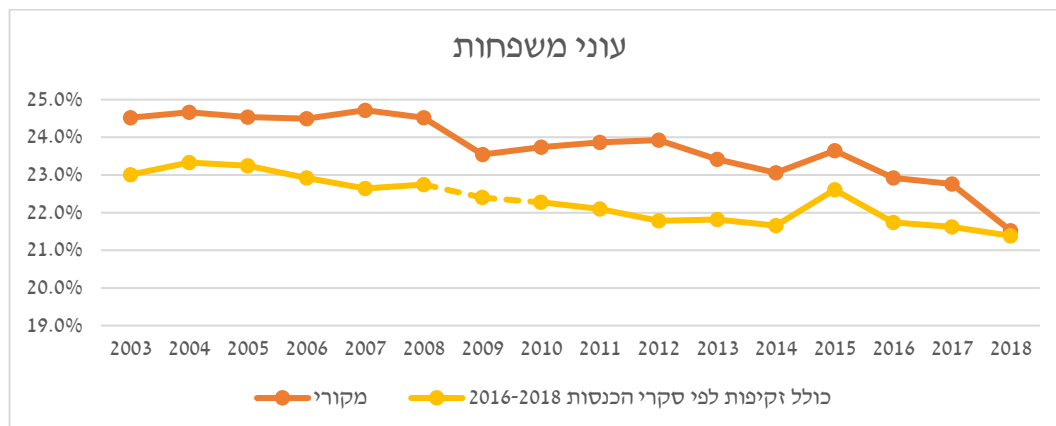


נספח 5ב : התפלגות ההכנסה הפנויה לנפש תקנית – נתוני סקר בהשוואה לנתונים מינהליים אחרי זקיפת הכנסות חסרות, 2018



נספח 6 : רגישות מדדי העוני לסקרים עליהם מבוסס המודל





### נספח 8 א: תוצאות רגרסיות המודל עם שכר כמשתנה קטגורי

בוצעה בדיקה של הכנסת השכר קטגורי. במודל כזה מתקבלים נתונים קרובים למודל עם משתנה שכה רציף. ההבדל בעוני משפחות 21.3 במשתנה קטגורי למול 21.4. בעוני נפשות 21.7 מול 21.7 בהתאמה.

הכנסות אחרות		הכנסות מהון		
רגרסיה לינארית לניבוי גודל ההכנסה	רגרסיה לוגיסטית לניבוי קיום ההכנסה	רגרסיה לינארית לניבוי גודל ההכנסה	רגרסיה לוגיסטית לניבוי קיום ההכנסה	
-0.0637 (0.381)	0.798 (0.524)	-0.365 (0.510)	0.0505 (0.676)	האם עבד
				שכר כמשתנה קטגורי בקבוצות של 1000
				ביחס לשכר קטן מ1000
-0.0941 (0.0985)	-0.242 (0.152)	0.254 (0.171)	-0.120 (0.224)	שכר 2000-2999
-0.240** (0.0971)	-0.486*** (0.152)	0.0466 (0.167)	-0.0238 (0.219)	שכר 3000-3999
-0.0987 (0.0977)	-0.607*** (0.155)	0.126 (0.167)	-0.0140 (0.218)	שכר 4000-4999
-0.0803 (0.0932)	-0.589*** (0.145)	0.215 (0.158)	0.0487 (0.206)	שכר 5000-5999
-0.0750 (0.0926)	-0.670*** (0.143)	0.0850 (0.153)	0.191 (0.200)	שכר 6000-6999
-0.175* (0.0944)	-0.796*** (0.145)	0.259* (0.157)	0.140 (0.205)	שכר 7000-7999
-0.177* (0.0962)	-0.978*** (0.146)	0.0489 (0.155)	0.230 (0.203)	שכר 8000-8999
-0.190* (0.103)	-1.223*** (0.153)	0.0258 (0.164)	-0.000771 (0.212)	שכר 9000-9999
-0.212** (0.103)	-1.152*** (0.154)	0.136 (0.166)	0.0383 (0.216)	שכר 10000-10999
-0.168 (0.104)	-1.138*** (0.154)	0.0998 (0.161)	0.162 (0.210)	שכר 11000-11999
-0.354*** (0.110)	-1.353*** (0.158)	0.133 (0.162)	0.198 (0.212)	שכר 12000-12999
-0.248** (0.111)	-1.364*** (0.159)	0.166 (0.161)	0.298 (0.211)	שכר 13000-13999
-0.382*** (0.114)	-1.459*** (0.162)	-0.00638 (0.170)	-0.0368 (0.221)	שכר 14000-14999
-0.622*** (0.119)	-1.530*** (0.165)	0.284* (0.163)	0.273 (0.214)	שכר 15000-15999
				שכר 16000-16999
				שכר 17000-17999
				שכר 18000-18999
				שכר 19000-19999
				שכר +20000
				מספר נפשות במשפחה
				משפחה עצמאית (חד הורית)
				האם יש ילדים
				מספר ילדים
				מספר ילדים עד 3
				נשים

(0.0628)	(0.0805)	(0.0767)	(0.0973)	
0.200	0.351**	0.141	-0.0602	אלמן
(0.126)	(0.163)	(0.142)	(0.186)	
-0.254	0.824***	0.0564	-0.648	חרדי
(0.193)	(0.285)	(0.383)	(0.481)	
0.0206	0.237**	0.0364	-0.00522	גרוש
(0.0947)	(0.118)	(0.0949)	(0.125)	
-0.00936	0.204**	0.0665	-0.146	נשוי
(0.0653)	(0.0826)	(0.0703)	(0.0897)	
0.0529	-1.322***	0.154	-2.274***	ערבי
(0.148)	(0.174)	(0.285)	(0.337)	
-0.376***	0.387***	-0.154**	-0.848***	עולה
(0.0545)	(0.0694)	(0.0717)	(0.0830)	
-0.127	-0.165*	-0.0968	-0.231***	מקבל פנסיה
(0.0800)	(0.0970)	(0.0649)	(0.0890)	
0.00936	-0.0241***	0.000947	0.0682***	לוג גובה פנסיה
(0.00729)	(0.00883)	(0.00585)	(0.00809)	
				יבשת לידה [בהשוואה לאירופה וצפון אמריקה]
				אפריקה או אסיה
-0.165	-0.0891	-0.385	-1.006	
(1.034)	(1.174)	(0.292)	(1.147)	
-0.979	0.0311	-0.0614	-1.482	ישראל
(1.158)	(0.789)	(0.198)	(1.112)	
0.0442	1.382***	0.450	-0.380	לא ידוע
(0.434)	(0.506)	(0.292)	(0.442)	
7.520***	0.230	8.315***	-0.710	חותך
(1.015)	(1.595)	(0.510)	(0.455)	
5,135	26,703	3,505	26,553	מספר תצפיות
0.131		0.117		R2
	0.204		0.107	Pseudo R2

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

### נספח 8 ב : השוואת מדדי עוני לאחר שימוש בשכר כמשתנה קטגוריאל

מינהלי			סקר		
אחרי זקיפות קטגוריאל	אחרי זקיפות	לפני זקיפות			
21.3	21.4	21.5	18.0		תחולת העוני משפחות
21.7	21.7	22.1	21.2		תחולת העוני נפשות
29.2	29.4	30.7	30.0		תחולת העוני ילדים
19.4	19.2	15.6	23.4		תחולת העוני קשישים
39.3	39.4	45.4	33.3		יחס פער ההכנסות
23.5	23.4	30.1	16.4		FGT
0.3756	0.3766	0.3888	0.3559		מדד גיני



## **Abstract**

In the recent years, the National Insurance (NII) compiled indices of poverty and inequality not only based on households surveys (of CBS), but also based on administrative data of the NII. Some of the income components such as other (non-NII) support and capital income, do not exist in administrative data. The purpose of this study is to describe the considerations to choose the most appropriate calculation in order to estimate these incomes, using the characteristics of households in the two data sources ("matching"). The findings show that after the imputation of missing income components, the indices are closer to those who are based on household surveys and enabling to improve the estimates of poverty and inequality produced from administrative data. These estimates will be updated and will improve as it will be required.





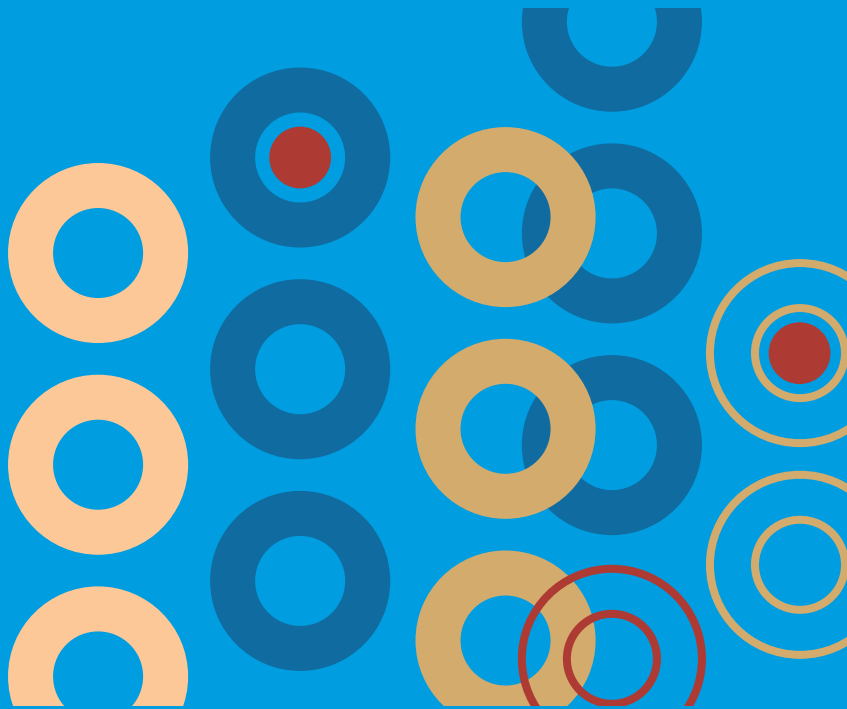
# Imputation of Missing Income Components in the Registered Data

Oren Heller and Miri Endeweld

Working Papers **138**

Jerusalem, October 2021

[www.btl.gov.il](http://www.btl.gov.il)



## Imputation of Missing Income Components in the Registered Data

Oren Heller, Miri Endeweld

