

### תרגיל 3

1. חוקר אמד את המודל:

$$y_i = \beta_0 + \beta_1 q_i + \beta_2 r_i + \beta_3 r_i^2 + \beta_4 s_i + u_i$$

כל המשתנים מקבלים ערכים חיוביים בלבד. החוקר מניח ש- $q_i, r_i$  ו- $s_i$  אינם מתואמים עם  $u_i$  ולכן האומדים שהתקבלו הם בלתי מוטים, אך יודע שההפרעה האקראית היא בעלת שונות שונה ולכן אומדי סטיות התקן הם מוטים ולא עקיבים.

א. אם החוקר יודע שמתקיים  $\text{var}(u_i) = \sigma_i^2 r_i^2 s_i^2$ , כתבו את המודל שעליו לאמוד כדי לקבל אומדים עקיבים לסטיות התקן (יש לציין במפורש מהו המשתנה המוסבר ומהם המשתנים המסבירים). אם לאמוד מודל זה, מה יקרה לאומדים לפרמטרים ולאומדי סטיות התקן - האם הם ישתנו או יישארו ללא שינוי? האם האומדים שיתקבלו יהיו יעילים? הסבירו.

ב. אם החוקר יודע רק שמתקיים  $\text{var}(u_i) = \sigma_i^2$ , הסבירו מה עליו לעשות כדי לקבל אומדים עקיבים לסטיות התקן. אם יעשה זאת, מה יקרה לאומדים לפרמטרים ולאומדי סטיות התקן - האם הם ישתנו או יישארו ללא שינוי? האם האומדים שיתקבלו יהיו יעילים? הסבירו.

2. שאלה זו מבוססת על נתוני סקר הכנסות לגברים ב-2011:

- א. הגבילו את המדגם לגברים בגילאי 25-66 שדיווחו שנות לימוד.
- ב. הגדירו את משתנה הסלקציה, המקבל את הערך 1 אם הפרט עובד כשכיר ו-0 אחרת (ניתן להשתמש במשתנה status או במשתנה lwage\_2011).
- ג. אימדו משוואת שכר מסוג Mincer, כפונקציה של הגיל, הגיל בריבוע, מספר שנות הלימוד והאם הגבר הוא ערבי. מהי התשואה להשכלה? האם יש מקום לחשוד כי האומד לתשואה זו הוא מוטה כתוצאה מסלקציה? מדוע?
- ד. כתבו מודל עם תיקון לסלקציה בשיטת Heckman של שני שלבים. **ציינו במפורש** מהי המשוואה הנאמדת בכל אחד מהשלבים, מהם המשתנים המוסברים והמסבירים בכל משוואה ומהם התנאים ההכרחיים לקבלת אומדים עקיבים.
- ה. אימדו את המודל שהצעתם בסעיף ד. מהי התשואה להשכלה כעת? האם מהתוצאות ניתן להסיק שאומד ה-OLS לתשואה זו אמנם היה מוטה, ואם כן - באיזה כיוון?
- ו. חיזרו על סעיפים ג ו-ה רק עבור גברים **נשואים**. האם גם עבור אוכלוסייה זו יש מקום לחשוד כי אומד ה-OLS לתשואה להשכלה הוא מוטה כתוצאה מסלקציה? האם מהתוצאות ניתן להסיק שאומד ה-OLS הוא אמנם מוטה, ואם כן - באיזה כיוון?

3. חוקר מעוניין לאמוד מודל שבו המשתנה התלוי הוא הציון הממוצע בסיום תואר ראשון. בידו מדגם גדול של סטודנטים שסיימו תואר ראשון ב-10 המוסדות האקדמיים הגדולים בישראל בשנה נתונה. עבור כל סטודנט ידוע המין, מוסד הלימודים (אוניברסיטת תל אביב, האוניברסיטה העברית, אוניברסיטת בן-גוריון וכו'), תחום הלימודים (מדעים מדויקים, מדעי החיים, מדעי החברה וכו'), הציון הפסיכומטרי, ממוצע הבגרות, הגיל בעת תחילת הלימודים, השכלת האב, מספר האחים/אחיות, הכנסת האב ומחוז המגורים (מחוז הצפון, חיפה, תל אביב, המרכז, ירושלים, הדרום).

א. כתבו את המשוואה הנאמדת. הסבירו מהם המשתנים המסבירים שכללתם בה, וכיצד ניתן להגדיר משתנים אלו מתוך הנתונים.

ב. מהם כיווני המקדמים שאתם מצפים לקבל באמידת המשוואה בסעיף א?

ג. האם יש מקום לחשוד בהטיית סלקציה כתוצאה מכך שקיימים נתונים רק על סטודנטים שסיימו תואר ראשון? מדוע?

ד. כתבו מודל עם תיקון לסלקציה בשיטת Heckman של שני שלבים. ציינו במפורש מהו המשתנה המוסבר במשוואת השלב הראשון, ומהם המשתנים המסבירים במשוואה זו.

ה. האם ניתן לאמוד את המודל בסעיף ד באמצעות הנתונים הקיימים? אם לא - אילו נתונים נוספים נדרשים? אם כן - מהם התנאים ההכרחיים לקבלת אומדים עקיבים למודל (יש להתייחס לנתונים הנדרשים ולהנחות לגבי המשתנים)?

### מבנה הקובץ ps3

Age	Age, 21 is group of 18-24	
Arab	Population group is Arab?	
children_0_4	Children aged 0-4 in household	
lwage_2011	Log of real hourly wage, 2011 prices	
marital	Marital status	1 Married 2 Divorced 3 Widowed 4 Single 5 Separated
schooling	Years of schooling	
status	Status at work	1 Salaried worker 2 Self-employed and salaried worker 4 Did not work in last 12 months
yeshiva	Last school is Yeshiva?	