מבוא לאקונומטריקה ב

תוכן העניינים

1. מבחן ML 1
2. משתנה דמי 5
3. הטרוסקדסטיות 22
4. מתאם סדרתי 30
5. משוואות סימולטניות 41
6. מודלים דינאמיים 57
7. שאלות חזרה למבחן מבוססות על תוכנת ATATS 63

מבוא לאקונומטריקה ב

פרק 1 - מבחן LM

תוכן העניינים

1. כללי 1

מבחן LM:

רקע:

במבחן כופלי לגרנג' (LM) אנו בודקים האם משתנה או משתנים מסבירים מסוימים רלוונטיים למודל.

לדוגמא:
נניח שיש לנו מודל הכולל 4 משתנים מסבירים (UNRESTRICTED):
$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t}$.
לגבי השניים הראשונים אנו בטוחים כי הם רלוונטיים וחייבים להופיע במודל. לגבי השניים האחרונים אנחנו לא בטוחים.
השערות:
$$\begin{matrix} H_0 : \beta_3 = \beta_4 = 0 \\ H_1 : OTHERWISE \end{matrix}$$.

המודל המוגבל (RESTRICTED): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t}$.
במבחן LM אומדים את המודל המוגבל ומקבלים עבור כל תצפית את הסטייה מקו הרגרסיה: $Y_t - \hat{Y}_t = \hat{u}_t$.
כעת אומדים את רגרסיית העזר שבה מנסים לנבא את הסטייה מקו הרגרסיה עבור כל תצפית: $\hat{u}_t = \delta_0 + \delta_1 x_{1t} + \delta_2 x_{2t} + \delta_3 x_{3t} + \delta_4 x_{4t} + \omega_t$.
חישוב הסטטיסטי: (R^2 של רגרסיית העזר * מספר התצפיות) $LM_{stat} = R^2 \cdot T$.
כלל הכרעה: אם $LM_{stat} > \chi^2_m$ נדחה את H_0 (m = מס' ההגבלות ב-H_0).

- שימו לב כי:
  עבור המשתנים הנוספים למודל – כל המדדים (הבטות, ערכי t וה-$Pvalu$) ברגרסיית העזר שווים לאלו של הרגרסיה הלא מוגבלת.
  עבור המשתנים הקיימים במודל – המדדים אינם שווים בין שתי הרגרסיות.

שאלות:

1) נניח מודל הכולל 4 משתנים מסבירים (UNRESTRICTED):

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \beta_4 x_{4t} .$$

UNRESTRICTED

Dependent variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	------	646169.84	----------	----------	0.0000
Error	------	620.17			
C Total	203	646790.01			

Root MSE	------	R-square	-----
Dep Mean	----	Adj R-sq	-----
C.V.	-----		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.067731	0.456604	11.09874	0.0000
X1	1	0.975726	0.042711	22.84485	0.0000
X2	1	3.005385	0.008679	346.2721	0.0000
X3	1	-5.029101	0.073149	-68.75146	0.0000
X4	1	8.974106	0.029075	308.6485	0.0000

RESTRICTED

Dependent variable: Y

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	------	646001.81			
Error	------	788.2			
C Total	203	646790.01			

Root MSE	---------	R-square	-----
Dep Mean	---------	Adj R-sq	-----
C.V.	---------		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	7.067731	0.656604	10.76406	0.0000
X1	1	26.36455	0.756627	34.84485	0.0000
X2	1	29.58626	0.076993	384.2721	0.0000

רגרסיית עזר

Dependent variable :RES

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	------	646169.84	--------	-------	0.0000
Error	------	620.17			
C Total	203	646790.01			

Root MSE	---------	R-square	0.213
Dep Mean	---------	Adj R-sq	-------
C.V.	---------		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.9608892	0.776604	7.675584	0.0000
X1	1	1.2077723	0.978845	1.233875	0.8455
X2	1	0.4840697	0.886754	0.545889	0.9976
X3	1	-5.029101	0.073149	-68.75146	0.0000
X4	1	8.974106	0.029075	308.6485	0.0000

א. בדוק את הטענה כי לפחות אחד מן המשתנים הנוספים רלוונטי למודל בשתי דרכים.

ב. איזה מן המשתנים הנוספים רלוונטי למודל?

ג. הסבירו את הקשרים בין שלוש המשוואות: R, U ועזר ואת הקשר בין מבחן WALD ומבחן LM.

U: $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 x_{1t} + \hat{\beta}_2 x_{2t} + \hat{\beta}_3 x_{3t} + \hat{\beta}_4 x_{4t} + \hat{v}_t$.

R: $\hat{Y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_{1t} + \hat{\alpha}_2 x_{2t} + \hat{u}_t$.

עזר: $\hat{u}_t = \hat{\delta}_0 + \hat{\delta}_1 x_{1t} + \hat{\delta}_2 x_{2t} + \hat{\delta}_3 x_{3t} + \hat{\delta}_4 x_{4t} + \hat{\omega}_t$.

ד. שחזרו בעזרת שתי המשוואות הראשונות (U ו-R) את LM_{stat}.

ה. שחזרו בעזרת המשוואה האחרונה (רגרסיית העזר) את $WALD_{stat}$.

תשובות סופיות:

1) א. מבחן LM ומבחן WALD, יש עדות לכך.

ב. $pt_{\hat{\beta}_3} = pt_{\hat{\beta}_4} = 0.00$.

ג. i. עזר+R=U.

ii. $ESS_U = ESS_y$.

iii. $ESS_R = TSS_y$.

iv. $R^2 = 1 - \frac{ESS}{TSS} = 1 - \frac{ESS_U}{ESS_R} = \frac{ESS_R - ESS_U}{ESS_R}$.

ד. $LM_{stat} = 43.489$.

ה. $WALD_{stat} = 26.962$.

מבוא לאקונומטריקה ב

פרק 2 - משתנה דמי

תוכן העניינים

1. כללי 5

משתנה דמי:

רקע:

הכנסת משתנים ב"ת איכותיים למודל הרגרסיה.
למשל, נתונה משוואת הרגרסיה: $W_t = \alpha + \beta \cdot S_t$.
W_t = השכר (התלוי).
S_t = שנות לימוד (הב"ת) שניהם כמותיים.
נניח שאנו סבורים שגם משתנה המגדר (משתנה איכותי) משפיע על השכר.
כדי להכניסו למשוואת הרגרסיה יש להגדיר משתני דמי (dummy variable):
נגדיר משתנה D שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר".
ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

1. משתנה דמי לחותך – המגדר משפיע על השכר ההתחלתי בלבד.
2. משתנה דמי לשיפוע – המגדר משפיע על התוספת לשכר בגין שנות הלימוד.
3. משתנה דמי לכל הפונקציה – המגדר משפיע גם על החותך וגם על השיפוע.

משתנה דמי לחותך:

המין משפיע על השכר ההתחלתי בלבד.
המודל: $W_t = \alpha_0 + \alpha_1 D + \beta \cdot S_t + u_t$ החותך מייצג כאן את השכר ההתחלתי.
שכר ההתחלתי של אישה: α_0.
שכר התחלתי של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר בין נשים וגברים: α_1 (הפרש בין החותכים).
בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש החותכים: $H_0 : \alpha_1 = 0$.

- השיפוע מייצג את התוספת בשכר כפונקציה של מס' שנות הלימוד והוא זהה עבור נשים וגברים.

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

המגדר הוא המשתנה היחיד במשוואה: $W_t = \alpha_0 + \alpha_1 D + u_t$.
החותך מייצג כאן את השכר הממוצע עבור כל קטגוריה:
שכר הממוצע של אישה: α_0.
שכר הממוצע של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר הממוצע בין נשים וגברים: α_1 (הפרש בין החותכים).
בדיקת השערות על משתנה הדמי: מבחן t: $H_0 : \alpha_1 = 0$ (מבחן זהה למבחן t להבדל בין ממוצעים).

משתנה דמי לשיפוע:

המגדר משפיע על התוספת לשכר בגין שנות הלימוד: $W_t = \alpha + \beta_0 S_t + \beta_1 DS_t + u_t$.
השיפוע מייצג כאן את התוספת לשכר בגין שנות לימוד.
אצל אישה: התוספת לשכר בגין שנות לימוד: β_0.
אצל גבר: התוספת לשכר בגין שנות לימוד: $\beta_0 + \beta_1$.
הבדל בין גברים לנשים בתוספת לשכר בגין שנות הלימוד: β_1 (הפרש השיפועים).
בדיקת השערות על משתנה הדמי: מבחן t למובהקות הפרש השיפועים: $H_0: \beta_1 = 0$.

- החותך, המייצג את השכר ההתחלתי, יהיה זהה עבור גברים ונשים.

משתנה דמי לכל הפונקציה:

המין משפיע גם על החותך וגם על השיפוע – גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.
המודל: $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 DS_t + u_t$.
השכר ההתחלתי של אישה: α_0.
השכר ההתחלתי של גבר: $\alpha_0 + \alpha_1$.
הבדל בשכר ההתחלתי בין המינים: α_1 (הבדל בחותכים).
אצל אישה: התוספת לשכר בגין שנות הלימוד: β_0.
אצל גבר: התוספת לשכר בגין שנות הלימוד: $\beta_0 + \beta_1$.
הבדל בין המינים בתוספת לשכר בגין שנות הלימוד: β_1 (הבדל בשיפועים).

2 דרכים לבדיקה האם יש השפעה למשתנה האיכותי:

1. בדיקת השערות למשתני הדמי:
באמצעות מבחן WALD יש לבדוק: $H_0: \alpha_1 = \beta_1 = 0$.
לפחות אחד הפרמטרים שונה מ-0: H_1.
אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים בנפרד: $H_0: \alpha_1 = 0$ ו- $H_0: \beta_1 = 0$.

2. מבחן CHOW:
דרך נוספת לבדיקת ההבדל בין הקטגוריות בלא יצירת משתני דמי:
חלוקת המדגם לפי הקטגוריות של המשתנה האיכותי.
מדגם של גברים (T_m) ושל נשים (T_f).
עבור כל קבוצה לאמוד משוואות רגרסיה לניבוי שכר על ידי שנות לימוד:
נשים: $W_t = \alpha_f + \beta_f X_t + u_t$.
גברים: $W_t = \alpha_m + \beta_m X_t + u_t$.
השערות: $H_0: \alpha_f = \alpha_m; \beta_f = \beta_m$.

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALD):
המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן יכלול את המדגם המאוחד: $W_t = \alpha + \beta X_t + u_t$.

המודל הלא מוגבל (U) כולל את שני חלקי המדגם: $\begin{matrix} ESS_U = ESS_f + ESS_m \\ DF_U = DF_f + DF_m \end{matrix}$.

סטטיסטי המבחן: $CHOW_{stat} = \dfrac{\dfrac{ESS_R - \left(ESS_f + ESS_m\right)}{DF_R - \left(DF_f + DF_m\right)}}{\dfrac{ESS_f + ESS_m}{DF_f + DF_m}} = WALD_{stat}$.

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

1. אם דחינו את H_0 במבחן CHOW נתקשה לברר את מקור ההבדל שנמצא.
2. בהרצת שתי רגרסיות נפרדות אנו בודקים הבדל בכל הפונקציה ואילו שיטת משתני הדמי מאפשרת לבדוק הבדל רק בחותך או רק בשיפוע.

סיכום ביניים:

	משתנה דמי לחותך	משתנה דמי לשיפוע	משתנה דמי לכל הפונקציה
המודל	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$
ההשערה במילים	קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותך).	קיים הבדל בין הקטגוריות בתוספת ל- Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותך ובשיפוע).
בדיקת ההשערה	מבחן t להפרש החותכים: $H_0 : \alpha_1 = 0$	מבחן t להפרש השיפועים: $H_0 : \beta_1 = 0$	מבחן WALD להפרש בין הפונקציות (החותכים והשיפועים): $H_0 : \alpha_1 = \beta_1 = 0$ **ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW. אם דוחים את H_0 יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALD): $H_0 : \alpha_1 = 0$ $H_0 : \beta_1 = 0$

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים:

כאשר המשתנה האיכותי כולל יותר משני ערכים/קטגוריות נגדיר מס' משתני דמי כמספר הקטגוריות פחות אחד.

למשל, את המשתנה האיכותי של עונות השנה הכולל 4 ערכים: אביב, קיץ, סתיו, חורף נייצג באמצעות 3 משתני דמי:

D_1 יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.

D_2 יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.

D_3 יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת הייחוס.

נניח שאנו רוצים לבדוק עונתיות במחירי הירקות:

V_t = מדד מחירי הירקות.

p_t = מדד המחירים לצרכן.

1. משתני דמי לחותך:

הטענה: יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות.

המודל: $V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t$.

כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב- β. למחיר זה יתווסף α_0 בחורף, $\alpha_0 + \alpha_1$ באביב, $\alpha_0 + \alpha_2$ בקיץ ו- $\alpha_0 + \alpha_3$ בסתיו.

ניתן לראות כי: α_0 - החותך בקטגוריה שהושמטה, $\alpha_0 + \alpha_1$ - החותך בקטגוריה i.

בדיקת השערות:

השערות: $\begin{matrix} H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0 \\ H_1 : OTHERWISE \end{matrix}$.

המבחן הסטטיסטי – מבחן WALD:

(U): $V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \beta \cdot P_t + u_t$.

(R): $V_t = \alpha + \beta \cdot P_t + u_t$.

- שימו לב שהחותך במשוואה המוגבלת איננו α_0 שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את H_0 במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין החותכים על ידי מבחני t:

1. האם יש הבדל במחיר ההתחלתי של הירקות בין האביב לחורף:
$H_0:\alpha_1=0$.
2. האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:
$H_0:\alpha_2=0$.
3. האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:
$H_0:\alpha_3=0$.

2. משתני דמי לשיפוע:

הטענה: יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן.

המודל: $V_t=\alpha+\beta_0P_t+\beta_1\left(D_{1i}P_t\right)+\beta_2\left(D_{2i}P_t\right)+\beta_3\left(D_{3i}P_t\right)+u_t$.

המחיר ההתחלתי של הירקות שווה בין עונות השנה (α) אולם כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב: β_0 בחורף, $\beta_0+\beta_1$ באביב, $\beta_0+\beta_2$ בקיץ ו- $\beta_0+\beta_3$ בסתיו.

ניתן לראות כי- β_0 : השיפוע בקטגוריה שהושמטה $\beta_0+\beta_i$: השיפוע בקטגוריה i.

בדיקת השערות:

השערות: $\begin{matrix}H_0:\beta_1=\beta_2=\beta_3=0\\H_1:OTHERWISE\end{matrix}$.

המבחן הסטטיסטי – מבחן WALD:

(U): $V_t=\alpha+\beta_0P_t+\beta_1\left(D_{1i}P_t\right)+\beta_2\left(D_{2i}P_t\right)+\beta_3\left(D_{3i}P_t\right)+u_t$.

(R): $V_t=\alpha+\beta\cdot P_t+u_t$.

- שימו לב שהשיפוע במשוואה המוגבלת איננו β_0 שכן המשתנה המסביר של עונות השנה ירד.

אם נדחה את H_0 במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני t.

3. משתני דמי לכל הפונקציה:

הטענה: יש הבדל בין עונות השנה בפונקציית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.

המודל:

$$V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 \left(D_{1i} P_t\right) + \beta_2 \left(D_{2i} P_t\right) + \beta_3 \left(D_{3i} P_t\right) + u_t .$$

בדיקת השערות:

השערות: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$.

המבחן הסטטיסטי - מבחן WALD:

(U):

$$V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 \left(D_{1i} P_t\right) + \beta_2 \left(D_{2i} P_t\right) + \beta_3 \left(D_{3i} P_t\right) + u_t .$$

(R): $V_t = \alpha + \beta \cdot P_t + u_t$.

אם דוחים את H_0, יש לבדוק במבחני WALD האם ההבדל הוא בין החותכים או בין השיפועים: $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$, $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$.

אם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחני t:

$H_0 : \alpha_j = 0$, $H_0 : \beta_j = 0$.

משתני דמי עבור שני משתנים איכותיים:

לדוגמא – שני משתנים איכותיים המשפיעים על פונקציית השכר:
מגדר (אישה, גבר) וגזע (לבן, שחור).
נגדיר משתנה דמי G שיקבל 1 אם מדובר בגבר ו-0 אחרת (אישה).
נגדיר משתנה דמי R שיקבל 1 אם מדובר בלבן ו-0 אחרת (שחור).
נבדוק כיצד מגדר וגזע משפיעים על השכר ההתחלתי (החותך), כאשר השכר תלוי גם בשנות לימוד (S_t).

1. הבדל בחותך ללא אינטראקציה:

המודל: $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$.

במודל זה – אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.
ניתן לבדוק השערות על כל אחד מהמשתנים הב"ת האיכותיים בנפרד:

1. הבדל בשכר ההתחלתי בין גברים לנשים: $H_0 : \alpha_1 = 0$.

2. הבדל בשכר ההתחלתי בין שחורים ללבנים: $H_0 : \alpha_2 = 0$.

2. הבדל בחותך עם אינטראקציה:
המודל: $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$.
במודל זה הטענה היא כי קיימת, בנוסף להשפעה של מגדר וגזע בנפרד על השכר, גם השפעה משולבת (אינטראקציה) של מגדר וגזע על השכר ההתחלתי.
במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי:

3. $H_0 : \alpha_3 = 0$.

3. דרך נוספת ליצירת מודל עם אינטראקציה:
הגדרת משתני דמי המייצגים שילוב בין המשתנים האיכותיים גזע ומגדר באופן הבא:
D_1 יקבל 1 אם מדובר בגבר לבן ו-0 אחרת.
D_2 יקבל 1 אם מדובר בגבר שחור ו-0 אחרת.
D_3 יקבל 1 אם מדובר באשה לבנה ו-0 אחרת.
הנשים השחורות מהוות כאן את קבוצת הייחוס.
המודל: $W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$.
נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה:

	גבר	אישה	הפרש
לבן	$\gamma_0 + \gamma_1$	$\gamma_0 + \gamma_3$	$\gamma_1 - \gamma_3$
שחור	$\gamma_0 + \gamma_2$	γ_0	γ_2
הפרש	$\gamma_1 - \gamma_2$	γ_3	

ההשערות לבדיקת קיום האינטראקציה: $H_0 : \gamma_1 - \gamma_3 = \gamma_2$ או $H_0 : \gamma_1 - \gamma_2 = \gamma_3$
התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך הקודמת: $WALD = t^2$ / $PF = Pt$.

שאלות:

משתנה דמי לחותך:

1) על בסיס מדגם של 50 איש העובדים בחברה מסוימת התקבלו התוצאות הבאות:

$W_t = 5500 + 1043 \cdot D + 119 \cdot S_t$.

(S.E) (134) (56) (24)

המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.

א. מהו השכר ההתחלתי של גבר בעל 12 שנות לימוד?

ב. מה ההבדל בשכר ההתחלתי בין גברים לנשים?

ג. האם הבדל זה מובהק באוכלוסייה?

ד. בדקו את הטענה כי השכר ההתחלתי של גברים גבוה ביותר מ-500 ₪ מזה של נשים.

ה. בדקו את הטענה שהשכר ההתחלתי של נשים נמוך ב-600 ₪ מזה של גברים.

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

2) על אותו המדגם של 50 איש העובדים בחברה מסוימת ביקש החוקר לבדוק האם יש הבדל בשכר הממוצע בין גברים לנשים.

תוצאות האמידה: $W_t = 5200 + 1120 \cdot D$.

נתון: $S_{\hat{\alpha}_1} = 63$.

בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

משתנה דמי לשיפוע:

3) על בסיס אותו מדגם, ביקש החוקר לדעת האם קיים הבדל מובהק בין גברים לנשים בתוספת לשכר בגין שנות הלימוד.

תוצאות האמידה נתונות להלן:

$W_t = 5000 + 110 \cdot S_t + 120 \cdot D \cdot S_t + u_t$

(68) (23) (25)

בדוק את ההשערה.

משתנה דמי לכל פונקציה:

4) חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.
החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

1. $NUM_t = \gamma_1 + \delta_1 \cdot AVGD_t + \varepsilon_{1t}$ כבישים מהירים בלבד.
2. $NUM_t = \gamma_2 + \delta_2 \cdot AVGD_t + \varepsilon_{2t}$ כבישים לא מהירים בלבד.
3. $NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t}$ שני סוגי הכביש (כל המדגם).
4. $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t$

כאשר:

NUM_t - מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה.

$AVGD_t$ - נפח התנועה בקטע כביש t ליום באלפים.

$TYPE_t$ - משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר, ו-0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

א. בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.

ב. חשבו את הערכים המספריים עבור אומדני משוואה (4).

ג. מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על ארבעת מכוניות ליום בקטע הכביש האמור?

הועלתה הטענה כי המקדם להשפעה של נפח התנועה בדרכים מהירות הינו כפול מזה שבדרכים לא-מהירות.

ד. מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4))?

ה. מהי הרגרסיה "תחת H_0" למבחן WALD?

משוואה (1) - כבישים מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	344
Number of Observations Used	344

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	4700.81174	4700.81174	89.12	<.0001
Error	342	18039	52.74684		
Corrected Total	343	22740			

Root MSE	7.26270	R-Square	0.2067
Dependent Mean	5.10465	Adj R-Sq	0.2044
Coeff Var	142.27617		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.55289	0.54303	2.86	0.0045
avgd	1	0.02098	0.00222	9.44	<.0001

משוואה (2) - כבישים לא מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	410
Number of Observations Used	410

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	971.99073	971.99073	145.83	<.0001
Error	408	2719.34830	6.66507		
Corrected Total	409	3691.33902			

Root MSE	2.58168	R-Square	0.2633
Dependent Mean	1.38780	Adj R-Sq	0.2615
Coeff Var	186.02612		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.14978	0.16360	0.92	0.3605
avgd	1	0.02877	0.00238	12.08	<.0001

משוואה (3) - שני סוגי הכביש (כל המדגם):

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	8052.00804	8052.00804	288.84	<.0001
Error	752	20964	27.87730		
Corrected Total	753	29016			

Root MSE	5.27990	R-Square	0.2775
Dependent Mean	3.08355	Adj R-Sq	0.2765
Coeff Var	171.22758		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.73903	0.23665	3.12	0.0019
avgd	1	0.02330	0.00137	17.00	<.0001

משוואה (4):

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	8256.966	2752.322	99.44	<.0001
Error	750	20759	27.678		
Corrected Total	753	29016			

Root MSE	5.26102	R-Square	0.2846
Dependent Mean	3.08355	Adj R-Sq	0.2817
Coeff Var	170.61553		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.14978	0.33340	0.45	0.6534
type	1				0.0067
avgd	1				<.0001
avgdtype	1				0.1283

משתנה איכותי עם יותר משתי קטגוריות:

5) ענה על הסעיפים הבאים:

א. הועלתה הטענה כי יש הבדל במחיר ההתחלתי בין האביב לקיץ.

i. מהי השערת האפס לבדיקת הטענה?

ii. פרטו שני מבחנים סטטיסטיים בעזרתם ניתן לבדוק את הטענה.

ב. הועלתה הטענה כי יש רק שתי עונות המשפיעות על מחיר הירקות ההתחלתי: קיץ + אביב, חורף + סתיו.

i. מהי השערת האפס לבדיקת הטענה?

ii. מהו המבחן הסטטיסטי המתאים? פרטו.

משתנה דמי עבור שני משתנים איכותיים:

6) חוקר בדק השפעות של השכלה, גזע (שחור, לבן) וניסיון (EXP) על לוג השכר (ln(Y)) במדגם בן 306 תצפיות:

$$\ln(Y)_t = \alpha_0 + \alpha_1 D_1 + \alpha_2 D_2 + \alpha_3 D_3 + \beta_1 EXP_t + \beta_2 EXP_t^2 + u_t .$$

$\ln(Y)$ - לוג השכר.

EXP - שנות ניסיון.

D_1 - מקבל את הערך 1 עבור שחורים בעלי השכלה גבוהה (ו-0 אחרת).

D_2 - מקבל את הערך 1 עבור שחורים בעלי השכלה נמוכה (ו-0 אחרת).

D_3 - מקבל את הערך 1 עבור לבנים בעלי השכלה גבוהה (ו-0 אחרת).

תוצאות אמידת משוואת הרגרסיה מוצגות בפלט להלן:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	--------	------	------	------
Error	300	140	------		
Corrected Total	305	210			

Root MSE	-------	R-Square	-------
Dependent Mean	-------	Adj R-Sq	-------
Coeff Var	-------		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	-------	--------	60.84	0.00
D1	1	-------	--------	-3.20	0.00
D2	1	-------	--------	-5.56	0.00
D3	1	-------	--------	7.23	0.00
EXP	1	-------	--------	8.11	0.00
EXP^2	1	--------	--------	-7.45	0.00

א. לפי המשוואה הניסיון זהה עבור שחורים ולבנים: נכון/לא נכון/ לא ניתן לדעת

ב. בדוק את הטענה כי בקרב אנשים בעלי השכלה נמוכה אין השפעה לגזע.

ג. בדוק את הטענה כי אין השפעות השכלה בקרב לבנים.

ד. מהי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?

ה. לבדיקת ההשערה של הסעיף הקודם בוצע מבחן WALD.
הרגרסיה המוגבלת תחת השערת האפס הינה:
$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$.
מהם ה-Zים?

ו. בדוק את ההשערה אם ידוע שבמודל המוגבל $R^2 = 0.33$.

ז. החוקר החליט לאמוד במקום את המשוואה המקורית את המשוואה:
$$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \lambda_3 (S \cdot E) + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$
כאשר:
S מקבל את הערך 1 עבור שחורים ו-0 אחרת (לבנים).
E מקבל את הערך 1 עבור השכלה גבוהה ו-0 אחרת (השכלה נמוכה).
מה הקשר בין המקדמים של שני המודלים?

ח. אם יאמוד החוקר את המשוואה:
$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$, האם תהיה טעות ספציפיקציה של השמטת משתנה רלוונטי (היעזר בסעיפים ד', ו' ו-ז').

7) חוקרת בדקה השפעות השכלה, מגדר וניסיון על הכנסה מעבודה לפי המשוואה הבאה:

$$\ln(MWAGE)=\alpha_0+\alpha_1\cdot S+\alpha_2\cdot E+\alpha_3\cdot(S\cdot E)+\beta_0\cdot EXP+\beta_1\cdot(EXP\cdot S)+\beta_2\cdot(EXP\cdot S)+\beta_3\cdot(EXP\cdot S\cdot E)+U,$$

כאשר:

S משתנה דמי: 1= עבור נשים, 0=גברים.

E משתנה דמי: 1= עבור השכלה גבוהה $(scl>12)$, 0=השכלה נמוכה.

א. רשמו את הפונקציה לחישוב:

i. תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.

ii. תחזית לוג השכר ההתחלתי עבור נשים משכילות.

iii. לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:

i. אין השפעה של מגדר והשכלה על השכר.

ii. השפעת ההשכלה אינה תלויה במגדר.

iii. אין השפעות השכלה אצל גברים.

iv. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

תשובות סופיות:

1) א. $W_t = 7971$. ב. 1,043 ₪. ג. כן. ד. יש עדות לכך.
ה. יש עדות לכך.

2) יש עדות לכך.

3) יש עדות לכך.

4) א. יש עדות לכך, מבחן CHOW: 1 , 2 ו-3, משתנה דמי: 3 ו-4.

ב. $\hat{\alpha} = 0.14978$, $\hat{\beta}_1 = 1.40311$, $\hat{\beta}_2 = 0.002877$, $\hat{\beta}_3 = -0.008$.

ג. $NUM_t = 1.532398$. ד. $\begin{matrix} H_0 : \beta_2 + \beta_3 = 2 \cdot \beta_2 \\ H_0 : \beta_3 = \beta_2 \end{matrix}$.

ה. $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_3 \cdot \left(AVGD_t + AVGD \cdot TYPE \right)_t + U_t$.

5) א.i. $H_0 : \alpha_1 = \alpha_2$. ii. WALD ו-t. ב.i. $H_0 : \alpha_1 = \alpha_2, \alpha_3 = 0$.
ii. WALD.

6) א. נכון. ב. יש עדות לכך. ג. יש עדות לכך.
ד. $H_0 : \alpha_2 = \alpha_1 - \alpha_3$ או $H_0 : \alpha_3 = \alpha_1 - \alpha_2$.

ה. $Z_4 = EXP_t^2$, $Z_3 = EXP_t$, $Z_2 = D_2 - D_3$, $Z_1 = D_1 + D_3$, $Z_0 = \ln\left(Y\right)_t$

ו. אין עדות לכך. ז. $\lambda_0 = \alpha_0$, $\lambda_1 = \alpha_2$, $\lambda_2 = \alpha_3$, $\lambda_3 = \alpha_1 - \alpha_2 - \alpha_3$.

ח. לא.

7) א.i. $\hat{\ln}\left(MWAGE\right) = \hat{\alpha}_0 + \hat{\beta}_0 \cdot 10$.

ii. $\hat{\ln}\left(MWAGE\right) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$.

iii. $EXP_t = \dfrac{-\left(\alpha_1 + \alpha_3\right)}{\beta_1 + \beta_3}$.

ב.i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$.

ii. $H_0 : \alpha_3 = \beta_3 = 0$.

iii. $H_0 : \alpha_2 = \beta_2 = 0$.

iv. $H_0 : \beta_2 + \beta_3 = 0$.

מבוא לאקונומטריקה ב

פרק 3 - הטרוסקדסטיות

תוכן העניינים

1. כללי 22

הטרוסקדסטיות:

רקע:

הטרוסקדסטיות הוא מצב שבו מופרת הנחת ההומוסקדסטיות, הגורסת כי שונות הטעויות היא אותה שונות עבור כל תצפית ותצפית: $V(u_t)=\sigma^2$ לכל t, כלומר התצפיות מפוזרות באופן אחיד סביב קו הרגרסיה.
במצב של הטרוסקדסטיות שונות הטעויות של כל תצפית היא שונה: $V(u_t)=\sigma_t^2$.

ההשלכות של הטרוסקדסטיות על אומדי OLS:

בהינתן הטרוסקדסטיות מופרת תכונת היעילות של אומדי הריבועים הפחותים שכן בכדי לחשב שונות יעילה של האומדים השתמשנו בהנחה של שונות קבועה.

מבחנים לזיהוי הטרוסקדסטיות:

החשד לקיומה של בעיית הטרוסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות – באיזה אופן השונות של הטעויות משתנה בין תצפית לתצפית.
שיטות לזיהוי הטרוסקדסטיות: מבחן GQ (Goldfeld-Qauandt) ומבחן White.
מבחן GQ מניח כי במקום שונות אחת אחידה של הטעויות לכל התצפיות, קיימות שתי שונויות שונות בלבד. ואילו מבחן White מניח כי לכל תצפית ותצפית שונות שונה של טעויות.

1. מבחן GQ:
   ההנחה העומדת בבסיס מבחן זה היא כי קיימות שתי שונויות שונות של טעויות.
   ביצוע המבחן:
   - מחלקים את המדגם לשני חלקים:
     1. החלק שבו אנו חושדים שיש שונות גבוהה יותר.
     2. החלק שבו אנו חושדים שיש שונות נמוכה יותר.

     מקובל להשמיט מס' תצפיות (בין 1/6 ל-1/3) במרכז המדגם.
   - אומדים כל אחד מהחלקים בנפרד ומקבלים את ה-ESS של כל חלק.
   - מחשבים את הסטטיסטי: $F_{stat}=\frac{ESS_1/T_1-K-1}{ESS_2/T_2-K-1}$ (תמיד השונות הגבוהה חלקי הקטנה).

- סטטיסטי זה מתפלג: $F_{(\alpha;T_1-K-1,T_2-K-1)}$.
- כלל ההכרעה: אם $F_{stat} > F_C$ אז דוחים את H_0.

- ההשערות: $\begin{matrix} H_0:\sigma_1^2=\sigma_2^2 \\ H_1:\sigma_1^2>\sigma_2^2 \end{matrix}$.

2. מבחן White:

ההנחה העומדת בבסיס מבחן זה כי לכל תצפית ותצפית שונות שונה של טעויות. הביטוי המתמטי של הנחה זו היא היותה של השונות פונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים:

$$\sigma_t^2 = f\left(x_j, x_j^2, x_j x_j\right)$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 x_1 + \ldots + \alpha_k x_k + \beta_1 x_k^2 + \ldots + \beta_k x_k^2 + \gamma_{12} x_1 x_2 + \gamma_{13} x_1 x_3 \ldots$$

האומד ל- σ_t^2 הוא $\hat{u}_t^2$.

המבחן הוא מבחן LM:

- אומדים את המודל המקורי ומקבלים את הסטיות מקו הרגרסיה $\hat{u}_t$ (המכונה בתוכנה הסטטיסטית RES-SAS).
- אומדים את $\hat{u}_t^2$ כפונקציה ליניארית של כל המשתנים המסבירים, ריבועיהם והאיברים הצולבים: $\hat{u}_t^2 / x_j, x_j^2, x_j x_j$ זוהי רגרסיית העזר.
- נחשב את סטטיסטי LM: $LM_{stat} = T_y \cdot R_y^2$.
- אם $LM_{stat} > \chi_m^2$ כאשר, m = מס' המשתנים ברגרסיית העזר.

השערות: $\begin{matrix} H_0:\alpha_j=\beta_j=\gamma_{jj}=0 \\ H_1:OTHERWISE \end{matrix}$.

פיתרון בעיית ההטרוסקדסטיות – ריבועים פחותים משוקללים (WLS):

נניח שאנו רוצים לאמוד את המודל: $Y_t = \alpha + \beta X_t + u_t$ וידוע כי לכל קריזה שונות אחרת. ההנחה שעומדת בבסיס שיטת ה-WLS היא כי השונות המשתנה כוללת בתוכה מרכיב קבוע ומרכיב משתנה: $\sigma_t^2 = Z_t \cdot \sigma^2$ את המרכיב המשתנה בשונות $\left(Z_t\right)$ יש לנטרל. לשם כך ניצור משתנה חדש W_t – שיהווה השורש ההופכי לאותו מרכיב משתנה: $W_t = \frac{1}{\sqrt{Z_t}}$.

נכפיל כל תצפית במשתנה החדש W_t וניצור משוואה שהיא קומבינציה ליניארית של

המשוואה המקורית:

$$Y_t = \alpha + \beta X_t + u_t$$
$$\cdot Y_t W_t = \alpha \cdot W_t + \beta\left(X_t W_t\right) + u_t W_t$$

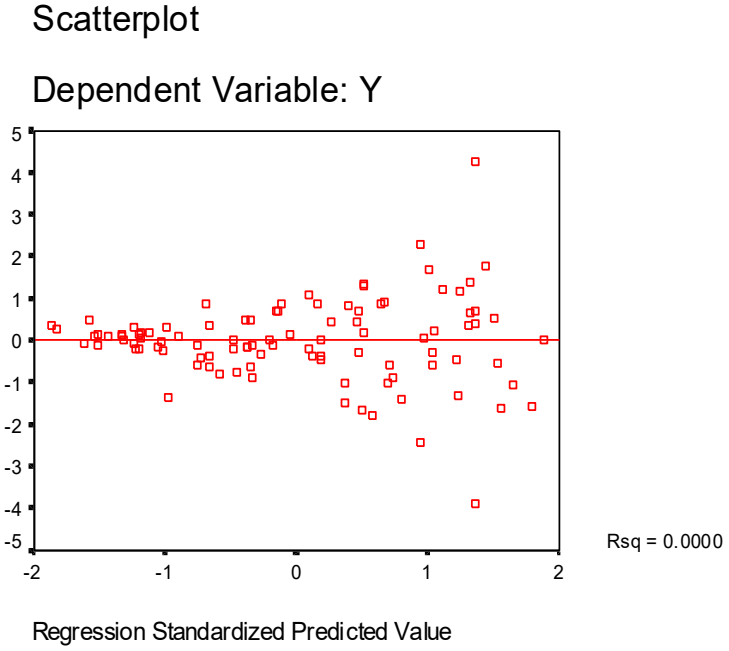
בצורתה המפורשת המשוואה החדשה נראית כך:

$$\cdot \frac{Y_t}{\sqrt{Z_t}} = \alpha \cdot \frac{1}{\sqrt{Z_t}} + \beta \cdot \frac{X_t}{\sqrt{Z_t}} + \frac{u_t}{\sqrt{Z_t}}$$

שאלות:

מבחן GQ:

1) נאמד הקשר שבין הכנסה לתצרוכת: $Y_t = \alpha + \beta X_t + u_t$.
גרף השאריות של הרגרסיה הנ"ל נתון להלן:



מגרף זה אנו למדים כי השונות איננה אחידה סביב קו הרגרסיה אלא תלויה ברמת ההכנסה – זהו מצב של הטרוסקדסטיות.

בכדי לבצע מבחן GQ:

- התצפיות של משתנה ההכנסה סודרו מהגדול לקטן והמדגם חולק לשלוש קבוצות שוות.
- רגרסיה נפרדת הורצה על השליש הראשון ועל השליש האחרון.

התוצאות של אמידת הקשר בין הכנסה לתצרוכת מוצג בפלטים 1 ו-2 בהתאמה:

משוואה (1)

```
The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read          16
Number of Observations Used          16

Analysis of Variance

                           Sum of            Mean
Source           DF        Squares          Square     F Value     Pr > F
  Model
Error            14          166452.9
Corrected Total
```

משוואה (2)

```
The REG Procedure
Model: MODEL1
Dependent Variable: y

Number of Observations Read          16
Number of Observations Used          16

Analysis of Variance

                           Sum of            Mean
Source           DF         Squares        Square     F Value     Pr > F
Model
Error            14        2934638
Corrected Total
```

האם יש עדות לקיום הטרוסקדסטיות בנתונים? (בצעו את המבחן המתאים: רשמו השערות, חשבו סטטיסטי מבחן, רשמו כלל הכרעה והגיעו למסקנה).

2) על מנת לבחון את פונקציית הייצור בענף מסוים נאספו נתונים על 150 פירמות. נסמן:

Q - תפוקה שנתית באלפי שקלים.

L - מספר עובדים.

המודל הנאמד: $\ln(Q)=\alpha+\beta\cdot\ln(L)$.

החוקר חשש שההפרעה המקרית איננה הומוסקדסטית. לשם כך הוא מיין את התצפיות בסדר עולה של מספר העובדים, השמיט $\frac{1}{3}$ מהתצפיות האמצעיות והריץ שתי רגרסיות נפרדות עם מספר שווה של תצפיות:
ברגרסיה הכוללת את הערכים הנמוכים יחסית של תשומת העבודה הוא

קיבל: $R^2 = 0.403$, $ESS = 279.3$.

ברגרסיה הכוללת את הערכים הגבוהים יחסית של תשומת העבודה הוא

קיבל: $R^2 = 0.238$, $ESS = 493.8$.

האם יש עדות לקיום הטרוסקדסטיות בנתונים? (בצעו את המבחן המתאים: רשמו השערות, חשבו סטטיסטי מבחן, רשמו כלל הכרעה והגיעו למסקנה).

מבחן WHITE:

3) על אותו הקשר שבין הכנסה לתצרוכת: $Y_t = \alpha + \beta X_t + u_t$.

שאנו חושדים על פי גרף השאריות כי קיים בו מצב של הטרוסקדסטיות.

בכדי לבצע את מבחן WHITE:

- נחשב את השאריות של הרגרסיה: $\hat{u}_t = Y_t - \hat{Y}_t$.
- נעלה את השאריות בריבוע: $\hat{u}_t^2$.
- נאמוד את המשוואה: $\hat{u}_t^2 = \alpha_0 + \alpha_1 X_t + \beta_1 X_t^2 + \nu_t$.

תוצאות האמידה מוצגות להלן:

The REG Procedure
Model: MODEL1
Dependent Variable: RES^2

Number of Observations Read	48
Number of Observations Used	48

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model					
Error					
Corrected Total					

Root MSE		R-Square	0.390763
Dependent Mean		Adj R-Sq	

בצעו מבחן White (השערות, סטטיסטי המבחן, כלל הכרעה ומסקנה).

4) חוקר מניח כי מכירות של חנות הן פונקציה של שיטחה, דמי שכירות והאפשרות של מכירת עיתונים.
נסמן:
Y_SALES - מכירות חודשיות (₪).
$X1_SQUARES$ - שטח החנות (מ״ר).
$X2_RENT$ - דמי שכירות ($).
$PAPERS$ - משתנה איכותי המקבל 1 אם החנות מוכרת גם עיתונים ו-0 אם לא.
החוקר חשד כי קיימת בעיה של הטרוסקדסטיות בנתונים.
החוקר ביצע מבחן לזיהוי הטרוסקדסטיות שתוצאותיו נתונות להלן:

```
                    Dependent Variable:
                Number of Observations Read          20
                Number of Observations Used          20

                   Analysis of Variance

                              Sum of          Mean
Source              DF        Squares         Square     F Value     Pr > F
Model
Error
Corrected Total
                    Root MSE                  R-Square     0.086942
                    Dependent Mean            Adj R-Sq

                    Parameter Estimates
                                 Parameter      Standard
Variable             DF          Estimate       Error     t Value     Pr > |t|
Intercept
X1_SQUARE
X1_SQUARE^2
X1_SQUARE*X2_RENT
X2_RENT
X2_RENT^2
X2_RENT*PAPERS
 PAPERS
```

הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: _____.
על ידי מבחן: _____.
המשתנה התלוי הינו: _____.
המשתנים הב״ת: _____.
ההשערות הינן: _____.
גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): _____.
המסקנה המתקבלת היא: _____.

שיטת WLS:

5) נתון המודל:

1. $Y_t = \alpha + \beta \cdot X_t + U_t$.

ונתון כי: $VAR(U_t) = \frac{\sigma^2}{Z_t^2}$ (Z_t משתנה ידוע).

א. מהי הבעיה שנוצרת באמידת משוואה (1)?

ב. מהן תכונות אומדי הריבועים הפחותים של משוואה (1)?

כדי לפתור את הבעיה שנוצרה, נאמדה המשוואה הבאה:

2. $Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + U_t \cdot W_t$.

ג. מהו W_t שבעזרתו ניתן לאמוד את α ו- β בצורה יעילה?

ד. מהו האומד היעיל של σ^2?

ה. האם ניתן להשוות בין המודלים על בסיס R^2? אם לא, האם ניתן להחליט בכל זאת איזה מודל טוב יותר?

ו. חוו דעתכם על הטענות הבאות, ונמקו:

i. אם נתון כי: $Z_t = a + b \cdot \bar{X}$, התשובות לסעיפים א' ו-ב' נשארות ללא שינוי.

ii. המשוואה הנורמאלית: $\sum \hat{\varepsilon}_t = 0$ (כאשר: $\varepsilon_t = U_t \cdot W_t$) היא אחת המשוואות הנורמאליות לאמידת משוואה (2).

6) ענו על השאלה הקודמת, כאשר נתון כי: $VAR(U_t) = \sigma^2 \cdot X_t^2$.

7) נתון המודל: $Y_t = \alpha + \beta X_t + u_t$. וקיים מדגם של 100 תצפיות כאשר נתון כי: $V(u_t) = \sigma_t^2 = \begin{cases} X_t \sigma^2 \Leftrightarrow t \geq 50 \\ X_t^2 \sigma^2 \Leftrightarrow t \leq 50 \end{cases}$. (שאר ההנחות הקלאסיות מתקיימות).

א. במשוואה מס' 1 יש בעיה של: _____.

ב. אמידת משוואה (1) תניב אומדים בלתי מוטים ועקיבים: נכון/ לא נכון/ לא ניתן לדעת

ג. פתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה הבאה: $Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + \omega_t$.

כאשר: $W_t =$ _____.

ד. אם נתון כי: $V(u_t) = \sigma_t^2 = \begin{cases} 3\sigma^2 \Leftrightarrow t \geq 50 \\ \sigma^2 \Leftrightarrow t \leq 50 \end{cases}$

האם ישתנו תשובותיכם לסעיפים א' ו-ב': כן/לא/לא ניתן לדעת

תשובות סופיות:

1) יש עדות לכך, השערות: $\begin{matrix} H_0:\sigma_1^2=\sigma_2^2 \\ H_1:\sigma_1^2>\sigma_2^2 \end{matrix}$, חישוב סטטיסטי: 17.62, כלל הכרעה: 2.48.

2) יש עדות לכך, השערות: $\begin{matrix} H_0:\sigma_1^2=\sigma_2^2 \\ H_1:\sigma_1^2>\sigma_2^2 \end{matrix}$, חישוב סטטיסטי: 1.77, כלל הכרעה: 1.69.

3) יש עדות לכך, השערות: $\begin{matrix} H_0:\alpha_1=\beta_1=0 \\ H_1:OTHERWISE \end{matrix}$, חישוב סטטיסטי: 18.75,

כלל הכרעה: 5.991.

4) קיום הטרוסקדסטיות בנתונים.

$WHITE\left(LM\right)$.

RES^2.

$X1_SQUARE$, $X1_SQUARE^2$, $X1_SQUARE*X2_RENT$,

$PAPERS$, $X2_RENT*PAPERS$, $X2_RENT^2$, $X2_RENT$.

$LM_{stat}=1.73$.

אין עדות לכך.

5) א. הטרוסקדסטיות. ב. ראו סרטון. ג. $W_t=\frac{1}{\sqrt{\frac{1}{Z_t^2}}}=Z_t$.

ד. $\sigma^2=\frac{ESS}{T-K}$. ה. לא, המודל השני. ו.i לא נכון. ii. לא נכון.

6) א. הטרוסקדסטיות. ב. ראו סרטון. ג. $W_t=\frac{1}{X_t}$.

ד. $S^2=\frac{ESS}{T-k-1}$. ה. ראו סרטון. ו. ראו סרטון.

7) א. הטרוסקדסטיות. ב. נכון.

ג. $t\ge 50$ $W_t=\frac{1}{\sqrt{X_t}}$, $t\le 50$ $W_t=\frac{1}{\sqrt{X_t^2}}$. ד. לא.

מבוא לאקונומטריקה ב

פרק 4 - מתאם סדרתי

תוכן העניינים

1. כללי 30

מתאם סדרתי:

רקע:

מתאם סדרתי עוסק במצב שבו מופרת ההנחה (מס' 6) של אי תלות בין הטעויות: $\text{cov}(u_t,u_s)=0$ ונוצרת תלות סטטיסטית בין הטעויות במודל: $\text{cov}(u_t,u_s)\neq 0$.
תלות כזו בין הטעויות קיימת בדרך כלל כאשר הנתונים הנאספים הם נתוני סדרות עיתיות ולא נתוני חתך בהם עסקנו עד כה. בנתוני סדרות עיתיות, מאחר ומדובר באותו הפרט הנמדד בזמנים שונים סביר שהטעויות בניבוי שלו תהיינה תלויות אחת בשנייה.

השלכות על אומדי הריבועים הפחותים (OLS):

מבין התכונות של אר"פ (ליניאריות, חוסר הטיה, עקיבות ויעילות) היחידה שמופרת כאשר קיים מתאם סידרתי היא: תכונת היעילות.
משום שתכונת היעילות היא היחידה מבין תכונות אר"פ התלויה להוכחתה בקיומה של הנחת אי התלות בין הטעויות. משום הפגיעה בתכונת היעילות, בדיקת ההשערות לא תהיה תקפה.

- שימו לב: כי במידה וקיים מתאם סידרתי חיובי בין הטעויות ולמשתנים יש מגמת זמן (X עולה או יורד עם הזמן) אומד השונות (ESS) יהיה מוטה כלפי מטה ואז נקבל: R^2 , F ו-t מוטים כלפי מעלה.

מבנה המתאם הסדרתי:

מתאם סדרתי מסדר ראשון:
ההנחה היא כי יש מתאם בין הטעויות במרחק אחד, כלומר u_t תלוי ישירות רק ב-u_{t-1}: $\text{cov}(u_t,u_{t-1})\neq 0$.
את המתאם בין הטעויות מסדר ראשון ניתן לנסח באופן הבא: $u_t=\rho\cdot u_{t-1}+\varepsilon_t$

כך ש:

1. $\rho\neq 0$ (כי אם $\rho=0$ אין מתאם סידרתי).
2. $-1<\rho<1$ (כי אם חורג מ-1 הטעות הולכת וגדלה עם הזמן).
3. ρ חיובי פירושו מתאם סדרתי חיובי ואילו ρ שלילי פירושו מתאם סדרתי שלילי (לא נפוץ).

4. ε_t מקיים את ההנחות הקלאסיות מאחר ומהווה סטייה מקרית לחלוטין

(בניגוד ל-u_t) כך ש:

$$E(\varepsilon_t)=0$$
$$V(\varepsilon_t)=\sigma_\varepsilon^2$$
$$\operatorname{cov}(\varepsilon_t,\varepsilon_{t-s})=0$$

.

המודל יכיל שתי משוואות-המשוואה העיקרית והגדרת המתאם הסדרתי (מסדר ראשון):

$$\begin{aligned} Y_t &= \alpha+\beta X_t+u_t \\ u_t &= \rho\cdot u_{t-1}+\varepsilon_t \end{aligned}$$

.

מלבד α ו-β נרצה לאמוד גם את ρ.

מתאם סדרתי מסדר שני:

$$u_t=\rho_1\cdot u_{t-1}+\rho_2\cdot u_{t-2}+\varepsilon_t$$

.

הן u_{t-1} והן u_{t-2} משפיעים ישירות על u_t.

מתאם סדרתי מסדר P:

$$u_t=\rho_1\cdot u_{t-1}+\rho_2\cdot u_{t-2}+\ldots+\rho_P\cdot u_{t-P}+\varepsilon_t$$

.

u_t מושפע מתקופות שונות בעבר.

תכונות המתאם הסדרתי:

מכיוון שכל טעות בזמן מסוים מתואמת עם הטעות הסמוכה לה בזמן: $r_{(u_t,u_{t-s})}=\rho^s$

המתאם של u_t הולך ופוחת עם הזמן: $\rho_{u_tu_{t-1}}>\rho^2_{u_tu_{t-2}}>\rho^3_{u_tu_{t-3}}>\ldots>\rho^s_{u_tu_{t-s}}$.

בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

$$E(u_t)=0$$
$$V(u_t)=\sigma_u^2=\frac{\sigma_\varepsilon^2}{1-\rho^2}$$
$$COV(u_t,u_{t-s})=\rho^s\sigma_{u_t}^2=\rho^s\frac{\sigma_s^2}{1-\rho^2}$$

מבחנים לזיהוי מתאם סדרתי:

מבחן DW (דרבין ווטסון) לקיום מתאם סדרתי מסדר ראשון:

- נניח תחילה כי אין מתאם סדרתי ונאמוד את המשוואה הראשית בשיטת OLS.
- כחלק מתוצאות האמידה נקבל ציון DW (יכול לקבל ערכים בין 0 ל-4 בלבד).
- נתבונן בטבלת DW ולפי K = מס' המשתנים הב"ת במודל ו-T = מס' התצפיות במדגם נשלוף שני ערכים: d_U ו- d_L.
- נחלק את הטווח שבין 0 ל-4 באופן הבא:

$$0 \text{—} \rho>0 \text{—} d_L \text{——} d_U \text{——} \rho=0 \text{——} 4-d_U \text{——} 4-d_L \text{——} \rho<0 \text{—} 4$$

- נראה היכן נופל ציון ה-DW שהתקבל כחלק מתוצאות האמידה.
ניתן לדעת אם יש מתאם ואיזה סוג של מתאם רק אם ציון ה-DW ייפול בחלקים המודגשים באדום.

השערות: $\begin{matrix} H_0:\rho=0 \\ H_1:\rho>0,\rho<0 \end{matrix}$.

חישוב הסטטיסטי: $DW_{stat}\cong 2\cdot\left(1-\hat{\rho}\right)$.

אם אנו מקבלים ציון $\hat{\rho}$ ניתן להציב בנוסחה ולקבל DW_{stat}.

למבחן DW יש שתי בעיות עיקריות:

1. מתאים רק למתאם סדרתי מסדר ראשון.
2. יש אזורים "מתים" בטווח בהם לא ניתן לדעת האם יש מתאם סדרתי.

בנוסף לכך על מספר תנאים להתקיים כדי שאפשר יהיה להשתמש במבחן DW:

1. הרגרסיה כוללת חותך.
2. ה-Xים קבועים ולא משתנים.
3. אין משתנים מסבירים שהם פיגור של המשתנה המוסבר.
4. אין תצפיות חסרות באמצע.
5. אם קיים מתאם סדרתי מסדר ראשון אז הוא מהצורה: $u_t=\rho\cdot u_{t-1}+\varepsilon_t$.

מבחן LM:

לעומת מבחן DW מבחן LM מתאים גם לבחינת קיומו של מתאם סדרתי מסדרים גבוהים יותר מסדר ראשון: $\begin{matrix} Y_t = \alpha + \beta \cdot X_t + u_t \\ u_t = \rho \cdot u_{t-1} + \varepsilon_t \end{matrix}$.

המודל הלא מוגבל:

(U): $Y_t = \alpha + \beta \cdot X_t + \rho \cdot u_{t-1} + \varepsilon_t$.

ניתן להתייחס לבחינת קיומו של מתאם סדרתי כהוספת משתנה מסביר: $\hat{u}_{t-1}$.

השלבים לביצוע המבחן:

- נאמוד את המודל המקורי ונחשב $\hat{u}_t$ ו- $\hat{u}_{t-1}$.
- נאמוד את רגרסיית העזר: $\hat{u}_t / x_1, x_2, ..., x_k; \hat{u}_{t-1}$.
- נחשב סטטיסטי LM: $LM_{stat} = T \cdot R^2$.
- נדחה את H_0 כאשר: $LM_{stat} > \chi^2_m$ כאשר m = סדר המתאם הסדרתי.

אם נדחה את H_0 נדע את סימנו של המתאם הסדרתי לפי המקדם של $\hat{u}_{t-1}$ ברגרסיית העזר ששווה ל- $\hat{\rho}$.

שימו לב כי אם נרצה לבדוק מתאם סדרתי מסדרים גבוהים יותר:

ההשערות: $\begin{matrix} H_0 : \rho_1 = \rho_2 = ... = \rho_s = 0 \\ H_1 : OTHERWISE \end{matrix}$.

רגרסיית העזר: $\hat{u}_t / x_1, x_2, ..., x_k; \hat{u}_{t-1}, \hat{u}_{t-2}, ..., \hat{u}_{t-s}$.

פתרון בעיית המתאם הסדרתי – רגרסיית הפרשים (שיטת קוקרן-אורקט):

ניצור משוואה שהיא קומבינציה ליניארית של המשוואה המקורית שבה לא יהיה מתאם סדרתי ולכן ניתן יהיה לאמוד אותה בשיטת הריבועים הפחותים, האומדים יהיו יעילים וניתן יהיה לבצע בדיקת השערות.

משוואה (1): המודל בזמן t: $Y_t = \alpha + \beta X_t + u_t$.

משוואה (2): המודל בזמן $t-1$ מוכפל ב- ρ: $\rho \cdot Y_{t-1} = \rho \cdot \alpha + \rho \cdot X_{t-1} + \rho \cdot u_t$.

החסרת משוואה (2) ממשוואה (1):

$Y_t - \rho Y_{t-1} = \alpha(1-\rho) + \beta(X_t - \rho X_{t-1}) + (u_t - \rho u_{t-1})$.

כדי לאמוד את הפרמטרים של רגרסיית ההפרשים נגדיר:

$$Y_t^* = Y_t - \rho Y_{t-1}$$
$$\alpha^* = \alpha(1-\rho)$$
$$\beta^* = \beta$$
$$X_t^* = X_t - \rho X_{t-1}$$
$$\varepsilon_t = u_t - \rho u_{t-1}$$

כך "נטרלנו" את המתאם הסדרתי: $\Leftarrow u_t = \rho \cdot u_{t-1} + \varepsilon_t$, $\varepsilon_t = u_t - \rho \cdot u_{t-1}$.

ε_t מקיים את כל ההנחות הקלאסיות ולכן שונות הרגרסיה וכן שונות הפרמטרים הנאמדים לא תהיה תלויה במקדם המתאם הסדרתי.

המשוואה "המתוקנת" אותה נאמוד: $Y^* = \alpha^* + \beta^* X_t^* + \varepsilon_t$.

לאחר אמידת משוואה זו ניתן לחלץ את האומדים של הפרמטרים המקוריים: $\hat{\alpha}$, $\hat{\beta}$.

מאחר ש- ρ איננו ידוע יש צורך לאמוד אותו.

אמידת ρ בשיטת קוקרן אורקוט:

שיטת קוקרן אורקוט לאמידת ρ היא שיטה איטראטיבית – מבוססת על חזרות של תהליך מסוים עד להתכנסות.

התהליך הממוחשב נקרא אוטורגרסיה (AUTOREGRESION) מסדר ראשון, שני, שלישי וכו' (תלוי בסדר המתאם הסדרתי). התיקון למתאם הסדרתי יתבצע על ידי הרצת רגרסיה עם משתנה AR(1) (אוטו רגרסיה מסדר ראשון), AR(1) ו-AR(2) (אם מניחים קיום אוטורגרסיה מסדר שני) וכו'. אם משתנה AR מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

שאלות:

תכונות המתאם הסדרתי:

1) נתון מתאם סדרתי מסדר ראשון: $u_t = \rho \cdot u_{t-1} + \varepsilon_t$.

נתון כי: $\rho = 0.9$ וכי: $V(\varepsilon_t) = \sigma_\varepsilon^2 = 1$.

מצאו את:

א. המתאם בין u_t ל-u_{t-1}.

ב. המתאם בין u_t ל-u_{t-4}. הסבר את ההבדל בין המתאמים (סעיף א' ו-ב').

ג. השונות σ_u^2.

ד. חזרו על סעיפים א' עד ג' עבור $\rho = 0.4$. הסבירו את ההבדל בין התוצאות.

מבחן DW:

2) חוקר רצה לאמוד את מחיר סגירה של מניה כפונקציה של הזמן שעובר:

$CLOSE_t = \alpha + \beta \cdot TIME_t + u_t$.

כאשר:

$CLOSE_t$ = מחיר סגירה של מניה ב-$ ביום t.

$TIME_t$ = משתנה זמן שמקבל את הערכים: 1,2,3...

תוצאות האמידה שהתקבלו:

Dependent variable: CLOSE

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	-------			55.78
Error	151	-------			
C Total	152	-------			
	---- Root MSE	R-square	0.181		
	---- Dep Mean	Adj R-sq	------		
	---- C.V.				

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	1.3474	0.0148	91.047	0.0000
TIME	1	-0.00075	0.0001	0.0000	-7.468
Durbin-Watson D	0.150				

האם קיים מתאם סדרתי?

TABLE 12 Cutoff Points for the Distribution of the Durbin-Watson Test Statistic

Let d_α be the number such that $P(d < d_\alpha) = \alpha$, where the random variable d has the distribution of the Durbin-Watson statistic under the null hypothesis of no autocorrelation in the regression errors. For probabilities $\alpha = .05$ and $\alpha = .01$, the tables show, for numbers of independent variables, K, values d_L and d_U such that $d_L \le d_\alpha \le d_U$, for numbers n of observations.

$\alpha = .05$

n	K									
	1		2		3		4		5	
	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
15	1.08	1.36	0.95	1.54	0.82	1.75	0.69	1.97	0.56	2.21
16	1.10	1.37	0.98	1.54	0.86	1.73	0.74	1.93	0.62	2.15
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.67	2.10
18	1.16	1.39	1.05	1.53	0.93	1.69	1.82	1.87	0.71	2.06
19	1.18	1.40	1.08	1.53	0.97	1.68	0.86	1.85	0.75	2.02
20	1.20	1.41	1.10	1.54	1.00	1.68	0.90	1.83	0.79	1.99
21	1.22	1.42	1.13	1.54	1.03	1.67	0.93	1.81	0.83	1.96
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.90	1.92
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.93	1.90
25	1.29	1.45	1.21	1.55	1.12	1.66	1.04	1.77	0.95	1.89
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.88
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.76	1.01	1.86
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.15	1.81
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80
36	1.41	1.52	1.35	1.59	1.29	1.65	1.24	1.73	1.18	1.80
37	1.42	1.53	1.36	1.59	1.31	1.66	1.25	1.72	1.19	1.80
38	1.43	1.54	1.37	1.59	1.32	1.66	1.26	1.72	1.21	1.79
39	1.43	1.54	1.38	1.60	1.33	1.66	1.27	1.72	1.22	1.79
40	1.44	1.54	1.39	1.60	1.34	1.66	1.29	1.72	1.23	1.79
45	1.48	1.57	1.43	1.62	1.38	1.67	1.34	1.72	1.29	1.78
50	1.50	1.59	1.46	1.63	1.42	1.67	1.38	1.72	1.34	1.77
55	1.53	1.60	1.49	1.64	1.45	1.68	1.41	1.72	1.38	1.77
60	1.55	1.62	1.51	1.65	1.48	1.69	1.44	1.73	1.41	1.77
65	1.57	1.63	1.54	1.66	1.50	1.70	1.47	1.73	1.44	1.77
70	1.58	1.64	1.55	1.67	1.52	1.70	1.49	1.74	1.46	1.77
75	1.60	1.65	1.57	1.68	1.54	1.71	1.51	1.74	1.49	1.77
80	1.61	1.66	1.59	1.69	1.56	1.72	1.53	1.74	1.51	1.77
85	1.62	1.67	1.60	1.70	1.57	1.72	1.55	1.75	1.52	1.77
90	1.63	1.68	1.61	1.70	1.59	1.73	1.57	1.75	1.54	1.78
95	1.64	1.69	1.62	1.71	1.60	1.73	1.58	1.75	1.56	1.78
100	1.65	1.69	1.63	1.72	1.61	1.74	1.59	1.76	1.57	1.78

3) המודל הינו: $\begin{aligned} Y_t &= \alpha + \beta X_t + \varepsilon_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{aligned}$.

u_t - הפרעה מקרית קלאסית.

וידוע כי: $T = 100$

$$u_t = 0.9u_{t-1}$$
$$d_L = 1.57$$
$$d_U = 1.65$$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

מבחן LM:

4) עבור הדוגמא הקודמת – ניבוי מחיר סגירה של מניה כפונקציה של הזמן:

$CLOSE_t = \alpha + \beta \cdot TIME + u_t$.

נבחן את קיומו של מתאם סדרתי מסדר ראשון באמצעות מבחן LM.

נאמוד את רגרסיית העזר: $u_t = \gamma_0 + \gamma_1 \cdot TIME_t + \rho \cdot u_{t-1} + \varepsilon_t$.

תוצאות האמידה שהתקבלו:

Dependent variable: RES

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	-------			
Error	150	--------			
C Total	152	-------			

Root MSE	------	R-square	0.855
Dep Mean	----	Adj R-sq	------
C.V.	-----		

Parameter Estimates

Variable	DF	Parameter Estimate
INTERCEP	1	-.000096
TIME	1	1.16331E-05
RES1	1	0.927172

א. האם קיים מתאם סדרתי?

ב. מהו ערכו של המתאם הסדרתי הנאמד?

ג. מהו כיוונו של המתאם הסדרתי באוכלוסייה?

תיקון המתאם הסדרתי:

5) סטודנט הניח כי במודל: $Y_t = \alpha + \beta X_t + u_t$ קיים מתאם סדרתי מסדר ראשון בשאריות כך שמתקיים: $u_t = 0.7u_{t-1} + \varepsilon_t$ ולכן במקום לאמוד את המודל המקורי אמד את המודל: $Y_t - 0.7Y_{t-1} = \alpha(1-0.7) + \beta(X_t - 0.7X_{t-1}) + u_t$.
הסטודנט טען כי במודל החדש לא קיים מתאם סדרתי.
טענת הסטודנט: נכונה /לא נכונה /לא ניתן לדעת

6) נמשיך עם הדוגמא של ניבוי מחיר סגירה של מניה כפונקציה של הזמן:
$CLOSE_t = \alpha + \beta \cdot TIME + u_t$.
נניח כי קיים מתאם סדרתי מסדר ראשון בנתונים.
תוצאות האמידה בשיטת אוטורגרסיה מסדר ראשון מוצגות להלן:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	1.333			
TIME	1	-0.0006			
(1)AR	1	0.927			0.000
Durbin-Watson D		2.235			

א. בדקו האם נפתרה בעיית המתאם הסדרתי.
ב. מהי המשוואה לאמידת מחיר הסגירה הצפוי ביום המסחר הבא?

תרגול מסכם:

7) נאמד הקשר שבין הכנסה לתצרוכת לתקופה ינואר 1994 עד דצמבר 1997 (T=48).
המודל הינו: $C_t = \alpha + \beta Y_t + u_t$.
בניסיון לבדוק האם מתקיים קשר מהסוג הבא: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t$.
נאמדה המשוואה הבאה: $\hat{u}_t = \gamma_1 \hat{u}_{t-1} + \gamma_2 \hat{u}_{t-2} + \gamma_3 \hat{u}_{t-3} + \gamma_4 Y_t + \omega_t$.
תוצאות האמידה מוצגות להלן:

Depended Variable: RES

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-54.709	710.85	-0.076	0.939
RESID1	1	0.705	0.152	4.631	0.000
RESID2	1	-0.0066	0.188	-0.035	0.972
RESID3	1	-0.337	0.167	-2.012	0.051
Y	1	0.0027	0.032	0.085	0.932
Durbin-Watson D		1.954			

נתון בנוסף כי: $R^2 = 0.479$.

א. הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: _____.
על ידי מבחן: _____.
ההשערות הינן: _____.
גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית): _____.
המסקנה המתקבלת היא: _____.

בהנחה כי קיים מתאם סדרתי מסדר שלישי בנתונים נאמד מחדש הקשר שבין ההכנסה לתצרוכת בהתאם לשיטתם של קוקרן ואורקוט.
תוצאות האמידה מוצגות להלן:

Depended Variable:C

Parameter Estimates

		Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|	DF	Variable
		-2103.53	1128.38	-1.864	0.069	1	INTERCEP
		0.71944	0.0510	14.086	0.000	1	Y
AR(1)	1	0.7070	0.1525	4.634	0.000		
AR(2)	1	-0.0064	0.1889	-0.034	0.972		
AR(3)	1	-0.3282	0.1669	-1.966	0.0562		
Durbin-Watson D		1.954					

ב. רשמו את המשוואה המתוקנת המשמשת לעריכת תחזיות.

8) חוקר רצה לאמוד את עקומת הביקוש לטיסות לאירופה.
לרשותו נתונים שבועיים לאורך 3 שנים (52 שבועות).
נסמן:
Y_t - מספר כרטיסי הטיסה לאירופה שנמכרו בשבוע t.
p_t - מחיר ממוצע ב-$ של הכרטיסים שנמכרו בשבוע t.
החוקר אמד את המודל: $Y_t = e^{\alpha} \cdot P_t^{\beta_1} \cdot P_{t-1}^{\beta_2} \cdot e^{u_t}$.
וקיבל לאחר הטרנספורמציה הלוגריתמית: $R^2 = 0.81$.
לבדיקת ההשערה כי קיים מתאם סדרתי בנתונים מסדר ראשון הוא חישב את ערכי $\hat{u}_t$ ולאחר מכן חישב את הרגרסיה: $u_t = \gamma_0 + \gamma_1 \ln P_t + \gamma_2 \ln P_{t-1} + \gamma_3 u_{t-1} + v_t$.
מקדם ההסבר המרובה ברגרסיה זו הוא 0.282.

א. נסח את ההשערה ובחן אותה בר"מ של 0.05.
החוקר מניח שיש מתאם סדרתי מסדר ראשון.
לאחר תיקון Cochrane-Orcutt התקבל: $\hat{\ln Y_t} = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$, $\hat{\rho} = 0.2$.
הניחו שהשבוע ובשבוע שעבר מחיר ממוצע של כרטיס היה 500$.
השבוע נמכרו 6,185 כרטיסים. בשבוע הבא צפוי מחיר של 400$.

ב. כמה כרטיסים יימכרו?
החוקר גם מנסה לקבוע האם בנתונים אלה קיים מתאם מסדר שני.

ג. רשמו את המשוואה הנוספת שעליו לאמוד.
במשוואה הנוספת התקבל מתאם מרובה השווה ל-0.12.

ד. מהי המסקנה בר"מ של 0.05?

תשובות סופיות:

1) א. $r_{(u_t,u_{t-1})}=0.9$. ב. $r_{(u_t,u_{t-4})}=0.6561$. ג. $\sigma_u^2=5.263$.

ד. $\sigma_u^2=1.19$, $r_{(u_t,u_{t-4})}=0.0256$, $r_{(u_t,u_{t-1})}=0.4$.

2) יש עדות לכך.

3) יש עדות לכך.

4) א. יש עדות לכך. ב. $\hat{\rho}=0.927$. ג. חיובי.

5) לא נכונה.

6) ראו סרטון.

7) א. קיומו של מתאם סדרתי מסדר שלישי בנתונים.

LM.

$$H_0:\rho_1=\rho_2=\rho_3=0$$
$$H_1:OTHERWISE$$

$LM_{stat}=22.99$.

יש עדות לכך.

ב. $\hat{C}_t=-2103.53+0.719\cdot Y_t+0.707\cdot\hat{u}_{t-1}-0.0064\cdot\hat{u}_{t-2}-0.328\cdot\hat{u}_{t-3}$.

8) א. $H_0:\rho=0$, $H_1:\rho\neq 0$, יש עדות לכך. ב. $Y_{t+1}=5,568$.

ג. $u_t=\gamma_0+\gamma_1\ln P_t+\gamma_2\ln P_{t-1}+\gamma_3 u_{t-1}+\gamma_4 u_{t-2}+\omega_t$. ד. אין עדות לכך.

מבוא לאקונומטריקה ב

פרק 5 - משוואות סימולטניות

תוכן העניינים

1. כללי 41

משוואות סימולטניות:

רקע:

עוסקות בהפרת ההנחה של אי תלות בין המב"ת לטעויות בניבוי: $\cov(x,u)=0$.
ה- X ים במשוואה נחשבו משתנים אקסוגניים – משפיעים על Y אך לא מושפעים ממנו בחזרה לעומת זאת משתנים אנדוגניים – משפיעים על Y אך גם מושפעים ממנו בחזרה. מאחר ומשתנים אלו הם גם מסבירים וגם מוסברים, הם נחשבים כמשתנים מקריים, המתואמים עם הטעויות במודל: $\cov(x,u)\neq 0$.

משוואות המבנה (משוואות סימולטניות):

מערכת משוואות הכוללות משתנים מסבירים אנדוגניים ואקסוגניים.
בד"כ מדובר בשתי משוואות אשר המשתנה המוסבר בראשונה הוא משתנה מסביר בשנייה והמשתנה המוסבר בשנייה הוא משתנה מסביר בראשונה.
משתנים המופיעים באחת המשוואות כמוסברים ובאחרת כמסבירים הם משתנים אנדוגניים. יתר המשתנים במשוואות הם אקסוגניים.
המטרה היא לאמוד בצורה יעילה את הפרמטרים (אלפות ובטות) ולבצע בדיקת השערות.

השלכות על אר"פ:

הנחת אי תלות בין המשתנה הב"ת והטעויות שימשה אותנו להוכחת ליניאריות, חוסר הטיה ועקיבות.
לכן הפרתה משמעה פגיעה בכל תכונות אר"פ.
האומדים לא ליניאריים, מוטים לא עקיבים ולכן גם לא יעילים (לפי גאוס מרקוב).
אומד השונות מוטה גם הוא ובדיקת ההשערות לא תקפה (ללא תלות בגודל המדגם).

הצורה המצומצמת של מודל עם משוואות סימולטניות:

משוואות הצורה המצומצמת הן פתרון עבור המשתנים האנדוגניים במערכת:
הגדרת המשתנים האנדוגניים כפונקציה של המשתנים האקסוגניים במערכת בלבד.
מספר המשוואות המצומצמות הוא כמספר המשתנים האנדוגניים במערכת (במקרה זה שניים).

תכונות המשוואות מהצורה המצומצמת:

- מס' המשוואות הוא כמספר המשתנים האנדוגניים במערכת (X ו- Y).
- המשתנה המוסבר הוא אנדוגני וכל המסבירים אקסוגניים.
- המשתנים המסבירים הם זהים בכל המשוואות (ה- Z ים).
- מכיוון שכל המשתנים המסבירים הם אקסוגניים ניתן לאמוד את הפרמטרים (ה- λ ות וה- μ ים) ב-OLS ולקבל אומדים ליניאריים, חסרי הטיה, יעילים ועקיבים עם יכולת לבצע בדיקת השערות.

אמידת הפרמטרים של משוואות המבנה באמצעות משוואות הצורה המצומצמת:
משוואות הצורה המצומצמת מאפשרות לאמוד את הפרמטרים בשיטת OLS אבל אנחנו מעוניינים למעשה לאמוד את הפרמטרים של המשוואות המקוריות – משוואות המבנה. מתוך הפרמטרים של הצורה המצומצמת נחלץ את הפרמטרים של משוואות המבנה.

בתהליך החילוץ של הפרמטרים המבניים ייתכנו 3 מצבים:

1. אין זיהוי: לא ניתן לחלץ את הפרמטרים המבניים מתוך הפרמטרים של הצורה המצומצמת.
2. זיהוי מדויק: יש רק דרך אחת לחלץ את הפרמטרים המבניים מהפרמטרים של הצורה המצומצמת.
3. זיהוי יתר: יש יותר מדרך אחת לחלץ את הפרמטרים המבניים מתוך הפרמטרים של הצורה המצומצמת.

בכדי להקל על בעיית הזיהוי מומלץ לאמץ את הכלל הבא:
עבור כל אחת מהמשוואות המבניות יש לחשב:

1. $g-1$: מס' אנדוגניים במשוואה הספציפית פחות 1 ולהשוות עם:
2. $K-k$: מספר אקסוגניים סה"כ בשתי המשוואות כולל חותך (K) פחות מספר אקסוגניים במשוואה הספציפית כולל חותך (k).

אם 1=2 זיהוי מדויק ; 1<2 זיהוי יתר ; 1>2 אין זיהוי.

שיטות לפתרון משוואות סימולטניות:

1. שיטת ריבועים פחותים עקיפה (ILS):
   א. יש להציג את מערכת משוואות המבנה בצורתה המצומצמת.
   ב. יש לאמוד בשיטת OLS את הפרמטרים של המשוואות בצורה המצומצמת.
   ג. יש לחלץ מן הפרמטרים של המערכת המצומצמת את הפרמטרים של הצורה המבנית.

משום שתהליך החילוץ איננו ליניארי האומדים המבניים המתקבלים הם מוטים אך עקיבים.
כאשר הזיהוי מדויק: האומדים יהיו גם אסימפטוטית יעילים (במדגמים גדולים).
כאשר הזיהוי הוא יתר: האומדים לא יהיו יעילים.

2. שיטת ריבועים פחותים בשני שלבים (2SL2):
   א. אמידת משוואות הצורה המצומצמת בשיטת OLS ושימוש בתוצאות האמידה כדי לחשב את המשתנים האנדוגניים (המסבירים).
   ב. הצבת המשתנים האנדוגניים שהתקבלו במשוואות המבנה ואומדתם ב-OLS.

אם משוואות המבנה מזוהות בדיוק או ביתר – האומדים שיתקבלו יהיו אמנם מוטים אבל עקיבים ויעילים אסימפטוטית. האומדים שיתקבלו יהיו זהים לאומדים שהתקבלו בשיטת הריבועים הפחותים העקיפה.
כאשר אין זיהוי: אין אקסוגניים ולכן אין משתנים מסבירים בצורה המצומצמת או שכל האקסוגניים בצורה המצומצמת כבר קיימים במשוואה המקורית ולכן החלפת x ב-$\hat{x}$ תיצור בעיה של מולטיקוליניאריות מלאה.

3. שיטת משתני העזר (IV):
   משתנה עזר הוא משתנה שיחליף את המשתנה המסביר האנדוגני במשוואת המבנה ויעזור לאמוד את הקשר בינו לבין התלוי.
   משתנה העזר צריך להיות:
   א. משתנה אקזוגני או פונקציה ליניארית של משתנים אקזוגניים: $\mathrm{cov}(Z,u)=0$.
   ב. מתואם עם המשתנה האנדוגני אותו הוא מחליף: $\mathrm{cov}(Z,X)\neq 0$.

ככל שהמתאם גבוה יותר, האומד שיתקבל באמצעותו יהיה טוב יותר.
הבעיה: אומדני OLS שיתקבלו יהיו מוטים, לא עקיבים ולא יעילים.
הפתרון בשיטת IV: אמידת ההשפעה של Y על X עם משתנה אקסוגני שלא קיים במערכת שמתואם עם Y (אותו הוא מחליף) אך לא עם u.

אם יש יותר ממשתנה עזר אחד המקיימים את התנאים הנ"ל, האומדים שיתקבלו יהיו כולם מוטים אך עקיבים (ניתן להשתמש בהם במדגמים גדולים).
משתנה העזר היחיד שיניב אומד יעיל יהיה בעל המתאם הגבוה ביותר עם המשתנה האנדוגני אותו הוא בא להחליף. משתנה עזר זה יהיה אומדן לאנדוגני שהתקבל מאמידת משוואת הצורה המצומצמת בשלב הראשון של 2SLS.

משתנה לא יוכל לשמש כמשתנה עזר:
אם נוסחתו מכילה רק משתנים אקזוגניים המצויים במשוואת המבנה בה הוא משמש כמשתנה עזר, שכן אז תיווצר בעיית מולטיקוליניאריות מלאה.
במילים אחרות, נוסחת משתנה העזר צריכה להיות מורכבת מלפחות משתנה אקזוגני אחד שלא מופיע במשוואה כדי שהמשתנה יוכל לשמש כמשתנה עזר.

משתני עזר שונים יכולים להניב את אותם האומדים לפרמטרים:
נבדוק זאת בצורה הבאה: נמחק מהנוסחאות של משתני העזר את המשתנים האקסוגניים המופיעים במשוואה. אם נשארנו עם שני ביטויים שהם מכפלה אחד של השני, יתקבלו אותם האומדים.

סיכום תוצאות אמידה של משוואות סימולטניות:

מס' האומדים שיתקבלו בשלושת השיטות ותכונותיהם תלויים בזיהוי של המשוואה:
אם המשוואה לא מזוהה: לא ניתן להשתמש באף אחת מהשיטות.
כאשר המשוואה מזוהה (בדיוק או ביתר): האומדים שיתקבלו בשלושת השיטות יהיו תמיד מוטים אך עקיבים.

תכונת היעילות ומס' האומדים האפשרי מסוכמים בטבלה הבאה:

	מזוהה בדיוק	**מזוהה ביתר**
שיטת ILS	אומד אחד לפרמטר **יעיל**	יתכן יותר מאומד אחד לפרמטר **לא יעילים**
שיטת 2SLS	אומדן אחד למשתנה האנדוגני **יעיל**	
שיטת IV	אינסוף משתני עזר אם משתנה העזר זהה לאומדן לאנדוגני המתקבל בשלב הראשון בשיטת 2SLS – הוא יהיה גם **יעיל**	

כאשר הזיהוי מדויק יתקבל אותו אומד מוטה אך עקיב ויעיל בשלושת השיטות:
ILS, 2SLS ו-IV (במידה ומשתנה העזר הוא $\hat{X}_t$ מהשלב הראשון של 2SLS).

משתנים בפיגור ומשוואות סימולטניות:

אם X_t אקסוגני אז גם המשתנים בפיגור X_{t-p} בוודאות אקסוגניים.

אם Y_t אנדוגני אז מעמדם של המשתנים בפיגור תלוי בקיומו של מתאם סדרתי:

אם יש מתאם סדרתי: $\text{cov}(Y_{t-1}, u_t \neq 0)$, אז Y_{t-1} אנדוגני.

אם אין מתאם סדרתי: $\text{cov}(Y_{t-1}, u_t = 0)$, אז Y_{t-1} אקסוגני.

מבחנים סטטיסטיים לבחינת אנדוגניות ולחוזק משתנה עזר:

מבחן האוזמן (Hausman Test):
מבחן המשמש אותנו לבחינת אנדוגניות של משתנה מסוים.

- השלב הראשון לביצוע מבחן האוזמן הוא הרצת המשוואה המצומצמת – כלומר, המשתנה שחושדים שהוא אנדוגני כתלוי על כל האקסוגניים.
- מאמידה זו נשמור את סידרת השאריות הנאמדות (vhat).
- כעת נאמוד את המודל המקורי (משוואת המבנה) ונוסיף לו את (vhat) כמשתנה מסביר חדש.
- לפי תוצאות האמידה – אם המקדם של vhat מובהק נסיק כי המשתנה הוא אכן משתנה אנדוגני במודל.

מבחן לחוזק IV:
מבחן שמתבצע על המשוואה המצומצמת שבה נעשה שימוש במשתני העזר.
בודקים:

א. האם משתנה העזר לניבוי המשתנה התלוי מובהק באוכ' באמצעות מבחן t למובהקות מקדם הרגרסיה. אם כן- ניתן להסיק כי המשתנה האקסוגני, המשמש כמשתנה עזר, מתואם עם האנדוגני אותו הוא אמור להחליף.

ב. אולם בכדי לבדוק האם משתני העזר חזקים מספיק נבצע מבחן F למובהקות כל משתני העזר המוצעים במשוואה המצומצמת. כלל אצבע-רק אם: $F_{stat} > 10$ נוכל להסיק כי משתני העזר חזקים מספיק בכדי שנוכל לקבל תוצאות אמינות כאשר אנו משתמשים בהם.

שאלות:

זיהוי משוואות המבנה:

1) חוקר רצה לאמוד את פונקציית הביקוש ואת פונקציית ההיצע לתות שדה.
הוא אסף נתונים עבור 30 תקופות:

Pt - מחיר קופסא בש"ח בתקופה t.

Qt - כמות נקנית בק"ג בתקופה t.

Zt - מחיר פרי תחליפי ב-₪ בתקופה t.

$INCOMEt$ - הכנסת הצרכנים באלפי ₪ בתקופה t.

Lt - מחיר שעת עבודה ב-₪ בתקופה t.

א. החוקר מניח שהכמות המבוקשת היא פונקציה של מחיר התות שדה, של מחיר הפרי התחלפי ושל הכנסת הצרכנים, והכמות המוצעת היא פונקציה של מחיר התות שדה ושל מחיר העבודה.
נסחו את המודל הסימולטני, תחת ההנחה שהגמישויות קבועות.
הציגו גם את תנאי הסדר וקבעו עבור כל משוואה אם היא מזוהה במדויק, ביתר או בחסר.

ב. עיינו במודל 1 שבדפי הפלט (ראו סרטון) והשיבו: איזו פונקציה נאמדה, והאם תוצאות האמידה שהתקבלו מתיישבות עם התיאוריה הכלכלית? נמקו.

ג. עיינו בדפי הפלט המתאימים (ראו סרטון) והשיבו: אם העלות של שעת עבודה תעלה באחוז אחד, מהם השינויים הצפויים בכמות ובמחיר של שווי משקל?

ד. בתקופה מסוימת אנו צופים שמחיר המוצר התחלפי יהיה 10 ₪, ההכנסה תהיה 50 אלף ₪, מחיר שעת עבודה 25 ₪.
מה יהיה מחיר שווי המשקל של תות השדה?
האם ניתן גם לאמוד את כמות שווי המשקל?

להלן הפלטים:

Model 1: TSLS estimates using the 30 observations 1-30

Dependent variable: l_Q

Instruments: l_L

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-1.83485	1.14385	-1.604	0.10869
l_P	-1.34898	0.645690	-2.089	0.03669 **
l_Z	1.72145	0.467875	3.679	0.00023 ***
l_income	0.984145	0.483543	2.035	0.04182 **

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 2.67757

Standard error of residuals = 0.32091

Unadjusted R-squared = 0.222881

Adjusted R-squared = 0.133214

F-statistic (3, 26) = 2.48564 (p-value = 0.0829)

Model 3: OLS estimates using the 30 observations 1-30

Dependent variable: l_Q

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	0.499595	0.630065	0.793	0.43500
l_L	-0.611731	0.163450	-3.743	0.00091 ***
l_income	0.395076	0.142590	2.771	0.01019 **
l_Z	0.937441	0.197381	4.749	0.00007 ***

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 0.834357

Standard error of residuals = 0.179139

Unadjusted R-squared = 0.681008

Adjusted R-squared = 0.644201

F-statistic (3, 26) = 18.5023 (p-value < 0.00001)

Log-likelihood = 11.1662

(Log-likelihood for Q = -75.1618)

Model 4: OLS estimates using the 30 observations 1-30

Dependent variable: l_P

VARIABLE	COEFFICIENT	STDERROR	T STAT	P-VALUE
const	-1.73053	0.419481	-4.125	0.00034 ***
l_L	0.453478	0.108821	4.167	0.00030 ***
l_income	0.436678	0.0949326	4.600	0.00010 ***
l_Z	0.581185	0.131411	4.423	0.00015 ***

Mean of dependent variable = 2.99427

Standard deviation of dep. var. = 0.314761

Sum of squared residuals = 0.369832

Standard error of residuals = 0.119266

Unadjusted R-squared = 0.871281

Adjusted R-squared = 0.856428

F-statistic (3, 26) = 58.6632 (p-value < 0.00001)

Log-likelihood = 23.3704

(Log-likelihood for P = -66.4576)

שיטת ILS:

2) נניח שאנו מתכוונים לאמוד את המשוואות:

$$C_t = \alpha + \beta Y_t + u_t$$
$$Y_t = C_t + Z_t$$

כאשר:

C_t - הוצאות לתצרוכת פרטית.

Y_t - הכנסה לאומית.

u_t - הפרעה אקראית.

א. מהי הבעיה באמידת המשוואות בשיטת הריבועים הפחותים? מהן תכונות אר"פ?

ב. האם המשוואות מזוהות?

ג. אמדו את מערכת המשוואות בצורתה המצומצמת באופן ידני.

ד. מהו הפתרון של המשוואות המצומצמות בשיטת ILS?

להלן תוצאות אמידת מערכת המשוואות בצורה המצומצמת:

Dependent Variable: C

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	14848.38	2568.8	5.78027	0.000
Z	1	-0.087066	0.3036	-0.2867	0.776

Dependent Variable: Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	14848.38	2568.8	5.78027	0.0000
Z	1	0.912934	0.3036	3.00699	0.0049

ה. חשבו את האומדים המבניים.

שיטת 2SLS:

3) תאר את תהליך האמידה בשני שלבים (2SLS) של משוואות המבנה:

$$C_t = \alpha + \beta Y_t + u_t$$
$$Y_t = C_t + Z_t$$

כאשר:

C_t - הוצאות לתצרוכת פרטית.

Y_t - הכנסה לאומית.

u_t - הפרעה אקראית.

א. מה ניתן יהיה לומר על האומדים שהתקבלו בשיטה זו?

ב. מה יהיה ערכם של האומדים $\hat{\beta}$ ו- $\hat{\alpha}$?

להלן תוצאות האמידה בשיטת 2 השלבים:

Dependent variable: C

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	16264.47	8221.233	1.978349	0.0520
y	1	-0.095370	0.364274	-0.261808	0.7943

Dependent variable: Y

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-9.95E-09	3.52E-09	-2.828212	0.0062
C	1	1.00000	2.08E-13	4.80E+12	0.0000
Z	1	1.00000	1.99E-13	5.04E+12	0.0000

4) לפניך המודל הסימולטני הבא:

משוואת הביקוש: $Q_t^D = \alpha_0 + \alpha_1 P_t + \alpha_2 Z_t + u_t$.

משוואת ההיצע: $Q_t^S = \beta_0 + \beta_1 P_t + v_t$.

P_t - מחיר המוצר בתקופה t.

Q_t^D - כמות מבוקשת בתקופה t.

Q_t^S - כמות מוצעת בתקופה t.

Z_t - מחיר המוצר התחלפי בתקופה t.

Z_t הוא משתנה אקסוגני.

א. רשום את המשוואות המצומצמות וקבע את התכונות של אומדי OLS למשוואות אלה.

ב. היעזר בשיטת ILS לאמידת הפרמטרים של המשוואה שניתן לזהות, אם התקבלו המשוואות המצומצמות הבאות:

$$\hat{Q}_t = 2 + 3Z_t$$

$$\hat{P}_t = 1 + 4Z_t$$

ג. באם ננסה לאמוד את משוואת הביקוש בשיטת TSLS:
האם ניתן יהיה לאמוד מספרית את המשוואה של השלב הראשון? נמק.
האם ניתן יהיה לאמוד מספרית את המשוואה של השלב השני? נמק.

ד. החוקר מנסה לאמוד את משוואת ההיצע בשיטת TSLS.
למה שווה האומדן שיתקבל ל-β_1?

שיטת IV:

5) נתונות המשוואות הבאות:

1. $Y_t = \alpha_0 + \alpha_1 \cdot X_t + \alpha_2 \cdot Z_{1t} + \alpha_3 \cdot Z_{2t} + \varepsilon_t$.
2. $X_t = \beta_0 + \beta_1 \cdot Y_t + \beta_2 \cdot Z_{1t} + \omega_t$.

נתון כי: T_t , X_t משתנים אנדוגניים ו- Z_{1t} , Z_{2t} משתנים אקסגוניים.
חוו דעתכם על כל אחת מהטענות הבאות, והסבירו:

א. ניתן להשתמש ב-Z_{1t} כמשתנה עזר לאמידת משוואה מס' 1.

ב. ניתן להשתמש ב-$\frac{Z_{1t} + Z_{2t}}{2}$ כמשתנה עזר לאמידת משוואה מס' 2.

ג. יתכנו מספר אומדים עקיבים שונים זה מזה ל-β_2 במשוואה מס' 2.

ד. שימוש ב-Z_2 כמשתנה עזר לאמידת משוואה מס' 2 יניב אומדים עקיבים וגם יעילים.

ה. משתנה העזר $Z_{1t} + Z_{2t}$ יניב אומדים זהים לאלו שהתקבלו בסעיף ב'.

ו. משתנה העזר $3Z_{1t} + 5Z_{2t}$ יניב אותם אומדים כמו משתנה העזר בסעיף ד'.

ז. משתנה העזר $7Z_{1t} + 5Z_{2t}$ יניב אומדים זהים לאלו שהתקבלו בסעיף ב'.

מבחן האוזמן:

6) נניח שאתם מעוניינים לאמוד את הקשר בין כלכלה לפיתוח פוליטי.
לכל מדינה i נסמן ב- Y_i – את הרמה הנאמדת של ההכנסה, נסמן ב- s_i את שיעור החיסכון במדינה i וב- D_i את האיתנות הנאמדת של השלטון הדמוקרטי. אתם שוקלים לאמוד מודל ליניארי של הכנסה ושיעור החיסכון על איתנות הממשל דמוקרטי: $D_i = \beta_1 + \beta_2 Y_i + \beta_3 s_i + \varepsilon_i$.
אבל אתם חוששים ש- $\text{cov}(Y_i, \varepsilon_i) \neq 0$.
הסבירו כיצד תשתמשו ב – *Hausman Test* כדי לבחון את ההשערה: $H_0 : \text{cov}(Y_i, \varepsilon_i) = 0$?

מבחן לחוזק IV:

7) נתונה מערכת המשוואות הסימולטניות הבאות:

$$Y_{1i} = \gamma Y_{2i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$
$$Y_{2i} = \delta Y_{1i} + \beta_3 X_{3i} + v_i$$

כאשר: X_1, X_2, X_3 הינם משתנים אקסוגנים.
להלן מערכת המשוואות של הצורה המצומצמת:

$$Y_{1i} = \pi_{11} X_{1i} + \pi_{12} X_{2i} + \pi_{13} X_{3i} + \tilde{u}_i$$
$$Y_{2i} = \pi_{21} X_{1i} + \pi_{22} X_{2i} + \pi_{23} X_{3i} + \tilde{v}_i$$

תארו כיצד בודקים ש- X_{1i} ו- X_{2i} אינם משתני עזר חלשים ל- Y_{1i} במשוואה השנייה?

תרגילים מסכמים:

1) נתונות המשוואות הבאות:

1. $Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_{1t} + \alpha_3 Z_{2t} + \alpha_4 Z_{3t} + \alpha_5 Z_{4t} + u_t$.
2. $X_t = \beta_0 + \beta_1 Y_t + \beta_2 Z_{1t} + \beta_3 Z_{2t} + \beta_4 Z_{5t} + v_t$.

נתון כי: $\text{cov}\left(Z_j, u_t\right) = 0$ עבור $j = 1,....,5$ (כלומר ה-Z ים אקסוגוניים).

א. אמידת כל אחת מהמשוואות תניב אומדים:

i. מוטים — נכון/לא נכון/לא ניתן לדעת.

ii. עקיבים — נכון/לא נכון/לא ניתן לדעת.

ב. משוואה 1 — מזוהה בדיוק/ מזוהה ביתר/בלתי מזוהה

משוואה 2 — מזוהה בדיוק/ מזוהה ביתר/ בלתי מזוהה

ג. חווה דעתך על הטענות הבאות:

i. תוך שימוש בשיטת ILS ניתן לאמוד את משוואה 1 באופן עקיב וחד ערכי: נכון/לא נכון/לא ניתן לדעת

ii. תוך שימוש בשיטת ILS ניתן לאמוד את משוואה 2 באופן עקיב וחד ערכי: נכון/לא נכון/לא ניתן לדעת

ד. משוואות הצורה המצומצמת הן:

$Y_t = \lambda_0 + \lambda_1 Z_{1t} + \lambda_2 Z_{2t} + \lambda_3 Z_{3t} + \lambda_4 Z_{4t} + \lambda_5 Z_{5t} + \varepsilon_{1t}$.

$X_t = \mu_0 + \mu_{1t} + \mu_2 Z_{2t} + \mu_3 Z_{3t} + \mu_4 Z_{4t} + \mu_5 Z_{5t} + \varepsilon_{2t}$.

נכון/לא נכון/לא ניתן לדעת.

ה. אמידת משוואות הצורה המצומצמת ב-OLS תניב אומדים חסרי הטיה, עקיבים ויעילים: נכון/לא נכון/אי אפשר לדעת

ו. להלן רשימה של משתני עזר פוטנציאליים:

i. Z_5.

ii. $\frac{Z_1 + Z_5}{2}$.

iii. $2Z_1 + 3Z_2 + Z_3$.

iv. $Z_3 + Z_4$.

v. $3Z_3 + 4Z_4$.

vi. $3Z_3 + 3Z_4$.

vii. Z_1.

עבור כל משתנה רשום באיזה משוואה ניתן להשתמש בו אם בכלל.

ז. איזה מבין משתני העזר הבאים יניבו את אותם האומדים עבור אותה המשוואה (תתכן יותר מתשובה אחת נכונה):

i. i ו-ii.

ii. iv ו-vi.

iii. v ו-vi.

iv. iv ו-v.

ח. האם משתנה עזר i (Z_5) יניב אומדים יעילים?

ט. אם ידוע כי אין מתאם סדרתי, האם X_{t-1} , Y_{t-1} הם אנדוגניים או אקסוגניים?

י. האם הוספה של משתנה אקזוגני נוסף למשוואה 1 תשנה את הזיהוי של משוואה 2?

יא. האם הוספה של משתנה אקסוגני נוסף למשוואה 2 תשנה את הזיהוי של משוואה 1?

יב. הנח כי הוטלו המגבלות הבאות על הפרמטרים המבניים: $\alpha_2 = \beta_2 = 0$. האם ניתן כעת לזהות את יתר הפרמטרים במודל?

2) היצע העבודה של נשים נשואות היה נושא מרכזי במחקר הכלכלי.
לצרוך אמידת היצע זה נבחר המודל הבא:

$$HOURS = \beta_1 + \beta_2 WAGE + \beta_3 EDUC + \beta_4 AGE + \beta_5 KIDSL6 + \beta_6 KIDS618 + \beta_7 NWIFEINC + \varepsilon$$

כאשר:

HOURS - היצע העבודה בשעות.

WAGE - שכר לשעה.

EDUC - מספר שנות הלימוד.

AGE - גיל.

KIDSL6 - מספר הילדים בבית מתחת לגיל 6.

KIDS618 - מספר הילדים בגיל 6-18.

NWIFEINC - הכנסת משק הבית ממקורות שאינם בעבודתה של האישה.

א. מהם הסימנים שתצפו לקבל בכל אחד מהמקדמים?

ב. הסבירו מדוע לא ניתן לאמוד את משוואת ההיצע הנ"ל בשיטת הריבועים הפחותים.

ג. הניחו כי אנחנו משתמשים בניסיון של האישה בשוק העבודה (*EXPER*) ובריבועו ($EXPER^2$) כמשתני עזר למשתנה *WAGE*. הסבירו מדוע משתני העזר הללו עונים על הדרישות שלנו ממשתני עזר.

ד. תארו את השלבים (לא בפקודות מחשב) שתבצעו כדי לקבל את האומדים בשיטת TSLS.

3) נתונה מערכת המשוואות הסימולטניות הבאה:

$$Y_{1t} = \gamma Y_{2i} + \beta_1 X_{1i} + \beta_2 X_{2i} + u_i$$
$$Y_{2i} = \delta Y_{1i} + \beta_3 X_{3i} + v_i$$

כאשר X_1 , X_2 , X_3 הינם משתנים אקסוגנים.

א. חלצו את מערכת המשוואות המצומצמת (Reduced Form Equations) של Y_1 ו- Y_2 (ז"א פתרו את המערכת המבנית עבור שני המשתנים האנדוגניים Y_1 ו- Y_2 על מנת לקבל את הצורה המצומצמת. כתבו את המקדמים והשאריות במערכת המצומצמת למטה כפונקציות של הפרמטרים והשאריות במערכת המבנית).

ב. הראו שבהינתן אומדים עקיבים ל- $\pi_{11},...,\pi_{23}$ ניתן למצוא אומד עקיב ל- γ.

ג. האם γ ניתן לזיהוי כאשר $\beta_3 = 0$?

ד. אילו תנאים צריכים X_{1i} ו- X_{2i} לקיים בכדי להיות משתני עזר ל- Y_{1i} במשוואה השנייה?

ה. תארו כיצד בודקים ש- X_{1i} ו- X_{2i} אינם משתני עזר חלשים ל- Y_{1i} במשוואה השנייה?

4) נניח שאתם מעוניינים לאמוד את הקשר בין כלכלה לפיתוח פוליטי.
לכל מדינה i נסמן ב- Y_i - את הרמה הנאמדת של ההכנסה וב- D_i את האיתנות הנאמדת של השלטון הדמוקרטי. אתם שוקלים לאמוד מודל ליניארי של הכנסה על ממשל דמוקרטי: $D_i = \beta_1 + \beta_2 Y_i + \varepsilon_i$, אבל אתם חוששים ש- $\text{cov}(Y_i, \varepsilon_i) \neq 0$.

א. הסבירו מדוע החשש שההכנסה מתואמת עם השגיאה במשוואה הנ"ל הגיוני?

ב. האם אומד הריבועים הפחותים של β_2 הינו חסר הטיה?

ג. נסמן ב- S_i את שיעור החיסכון במדינה i. הסבירו אלו תנאים צריך משתנה עזר (iv) לקיים. נמקו מדוע S_i מתאים או לא מתאים לשמש כמשתנה עזר.

ד. הסבירו כיצד תשתמשו בשיטת 2SLS כדי לאמוד את β_2. האם האומד המתקבל עקיב?

ה. הסבירו כיצד תשתמשו ב- *Hausman Test* כדי לבחון את ההשערה:
$H_0 : \text{cov}(Y_i, \varepsilon_i) = 0$.

תשובות סופיות:

1) א. $\ln Q_t^D = \alpha_0 + \alpha_1 \ln P_t + \alpha_2 \ln Z_t + \alpha_3 \ln INCOME_t + u_t$

$$\ln Q_t^S = \beta_0 + \beta_1 \ln P_t + \beta_2 \ln L_t + v_t$$

$$Q_t^D = Q_t^S$$

משוואת הביקוש מזוהה במדויק.

משוואת ההיצע מזוהה ביתר.

ב. פונקציית הביקוש, התוצאות מתיישבות.

ג. הכמות תרד ב-0.61173%, המחיר יעלה ב-0.453478%.

ד. $\hat{P} = 16.05$, $\hat{Q} = 9.34$.

2) א. ראו סרטון. ב. מזוהות בדיוק. ג. $\hat{Y}_t = \hat{\gamma}_3 + \hat{\gamma}_4 Z_t$, $\hat{C}_t = \hat{\gamma}_0 + \hat{\gamma}_1 Z_t$.

ד. $\hat{\beta} = \frac{\hat{\gamma}_1}{\hat{\gamma}_4}$, $\hat{\alpha} = \frac{\hat{\gamma}_0}{\hat{\gamma}_4} = \frac{\hat{\gamma}_3}{\hat{\gamma}_4}$. ה. $\hat{\beta} = -0.09537$, $\hat{\alpha} = 16,264.46$.

3) א. מוטים אך עקיבים ויעילים במדגמים גדולים.

ב. $\hat{\beta} = -0.09537$, $\hat{\alpha} = 16,264.47$.

4) א. $Q_t = \beta_0 + \beta_1 \cdot \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\beta_1 \cdot \alpha_2}{\alpha_1 - \beta_1} \cdot Z_t + \frac{\beta_1 (u_t - v_t)}{\alpha_1 - \beta_1} + v_t$, BLUE.

ב. $\hat{\beta}_1 = 0.75$, $\hat{\beta}_0 = 1.25$. ג. שלב ראשון: ניתן, שלב שני: לא ניתן.

ד. 0.75.

5) א. לא נכון. ב. נכון. ג. נכון. ד. נכון. ה. נכון.

ו. נכון. ז. לא נכון.

6) ראו סרטון.

7) ראו סרטון.

תרגילים מסכמים:

1) א.i. נכון. ii. לא נכון.

ב. משוואה 1: מזוהה בדיוק, משוואה 2: מזוהה ביתר.

ג.i. נכון. ii. לא נכון. ד. נכון. ה. נכון.

ו.i. 1. ii. 1. iii. 2. iv. 2. v. 2.

vii. 2. ז.i. נכון. ii. נכון. iii. לא נכון. iv. לא נכון.

ח. כן. ט. אקסוגניים. י. לא. יא. כן.

יב. משוואה 1 מזוהה בדיוק ומשוואה 2 מזוהה ביתר.

2) א. מקדם wage חיובי, מקדם educ לא ניתן לדעת, מקדם age יכול להיות חיובי או שלילי, מקדם kidsl6 שלילי, מקדם kids618 חיובי, מקדם nwifeinc שלילי.

ב. ראו סרטון. ג. ראו סרטון. ד. ראו סרטון.

3) א. $\pi_{13} = \frac{\beta_3 \gamma}{1 - \delta\gamma}$, $\pi_{12} = \frac{\beta_2}{1 - \delta\gamma}$, $\pi_{11} = \frac{\beta_1}{1 - \delta\gamma}$,

$$\pi_{23} = \frac{\beta_3}{1 - \delta\gamma} \quad , \quad \pi_{22} = \frac{\beta_2 \delta}{1 - \delta\gamma} \quad , \quad \pi_{21} = \frac{\beta_1 \delta}{1 - \delta\gamma} \quad ,$$

$$\tilde{v}_i = \frac{v_i + \delta u_i}{1-\delta\gamma} \ , \ \tilde{u}_i = \frac{u_i + \gamma v_i}{1-\delta\gamma} .$$

ב. מכיוון ש- $\gamma = \frac{\pi_{13}}{\pi_{23}}$, ניתן לקבל אומד עקיב ל- γ ע"י $\hat{\gamma} = \frac{\hat{\pi}_{13}}{\hat{\pi}_{23}}$.

ג. לא. ד. צריכים להיות מתואמים עם y_{1i} ובלתי מתואמים עם v_i .

ה. ראו סרטון.

4) א. טעות מדידה במשתנה המוסבר, משתנה מושמט, משוואות סימולטניות.

ב. לא. ג. ראו סרטון. ד. עקיב. ה. ראו סרטון.

מבוא לאקונומטריקה ב

פרק 6 - מודלים דינאמיים

תוכן העניינים

1. כללי 57

מודלים דינאמיים:

רקע:

מודל דינמי הוא מודל שיש בו משתנה מוסבר בפיגור, כלומר Y היום מושפע מ-Y של אתמול: $Y_t = \alpha + \beta_0 X_t + \beta_1 Y_{t-1} + u_t$.

המכפילים הדינמיים:

שלוש סוגים של השפעות בהקשר של המודל הדינמי (מכפלים):

1. מכפל לטווח קצר (מיידי):

   איך X היום משפיע על Y היום: $\frac{\partial Y_t}{\partial X_t}$.

2. מכפל ביניים מסדר j (מכפיל דינמי):

   איך X מלפני j תקופות משפיע על Y היום: $\frac{\partial Y_t}{\partial X_{t-j}}$.

3. מכפל טווח ארוך (מצב עמיד):

   איך X משפיע על Y לאורך P תקופות: $\frac{\partial Y^*}{\partial X^*}$.

כאשר X ו-Y נותרים קבועים על פני הזמן (מצב עמיד):

$$Y_t = Y_{t-1} = \ldots = Y_{t-p} = Y^*$$
$$X_t = X_{t-1} = \ldots = X_{t-p} = X^*$$

הקשר בין מתאם סדרתי למודלים דינמיים

המתאם הסדרתי נובע מהשמטה של דינמיות מבנית במודל. המודל המקורי היה צריך להיות מודל דינמי אך נאמד בטעות מודל סטאטי. הדינמיות תבוא אז לידי ביטוי בטעויות, כלומר במתאם הסדרתי.
רגרסיית ההפרשים, המהווה פתרון למתאם הסדרתי, היא למעשה מודל דינמי.

לסיכום:
בכדי לפתור את בעיית המתאם הסדרתי יש לאמוד מלבד את המשתנה המוסבר בזמן t גם את המשתנה המוסבר והמסביר בזמן $t-1$.
המשתנה בפיגור Y_{t-1} נועד לפתור מתאם סדרתי מסדר ראשון, Y_{t-2} משמש לפתירת מתאם סדרתי מסדר שני וכך הלאה.
בכדי לבדוק קיומו של מתאם סדרתי במודל דינמי לא נוכל לבצע מבחן DW אלא רק מבחן LM.

השלכות על אר"פ של משתנה מוסבר בפיגור כמשתנה מסביר:

בניגוד למשתנה מסביר רגיל (X), Y_{t-1} הינו משתנה מקרי.
משום כך אר"פ ברגרסיה הכוללת משתנים כאלה הם מוטים (להזכירכם בהוכחת חוסר הטיה של האומדים השתמשנו בהנחה מס' 4 הגורסת כי המשתנים המסבירים אינם משתנים מקריים).
בנוסף לכך העקיבות של האומדים תלויה בקיום מתאם סדרתי:

$$\hat{\beta} \to \beta + \frac{COV\left(Y_{t-1}, u_t\right)}{V\left(Y_{t-1}\right)}.$$

אם אין מתאם סדרתי: $COV\left(Y_{t-1}, u_t\right) = 0$ $\Leftarrow$ האומד עקיב.

אם יש מתאם סדרתי: $COV\left(Y_{t-1}, u_t\right) \neq 0$ $\Leftarrow$ האומד איננו עקיב.

לסיכום – ההשלכות על אר"פ:

1. האומדים מוטים ולכן ניתן לבצע בדיקת השערות רק במדגמים גדולים (T>30).
2. אם אין מתאם סדרתי $\Leftarrow$ האומדים עקיבים ויעילים (ניתן לבצע בדיקת השערות במדגמים גדולים).
   אם יש מתאם סדרתי $\Leftarrow$ האומדים אינם עקיבים ואינם יעילים (לא ניתן לבצע בדיקת השערות גם במדגמים גדולים).

שאלות:

חישוב מכפלים:

1) חשבו את שלושת סוגי המכפלים של המודלים הדינמיים הבאים:

א. $Y_t = \alpha + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + u_t$.

ב. $Y_t = \alpha + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + u_t$.

תרגול מסכם:

2) המודל הבא הורץ ב-SAS עם מדגם בעל 100 תצפיות: $Y_t = \alpha + \beta X_t + u_t$.

א. מהפלט עולה DW=0.195 לפיכך:

i. לא קיים מתאם סדרתי.

ii. קיים מתאם סדרתי והוא:_____.

iii. לא ניתן לקבוע אם המתאם הסדרתי מובהק.

ב. לפי תשובתך לסעיף א' חווה דעתך על תכונות האומדים:

i. מוטים נכון/ לא נכון/ לא ניתן לדעת

ii. ליניאריים נכון/ לא נכון/ לא ניתן לדעת

iii. יעילים נכון/ לא נכון/ לא ניתן לדעת

iv. עקיבים נכון/ לא נכון/ לא ניתן לדעת

ג. אמידה של איזו משוואה תפתור באופן מלא את הבעיה שנוצרה במודל:

i. $Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + u_t$.

ii. $Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-1} + u_t$.

iii. $Y_t = \alpha + \beta_1 X_t + \beta_2 Y_{t-1} + u_t$.

ד. בדוק את ההשערה כי לפי מודל (3) השפעת X על Y הולכת ופוחתת עם הזמן.
מצורף החלק הרלוונטי מהפלט:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	0.42	0.06	7.00	0.000
X	1	0.25	0.03	8.33	0.000
Y1	1	0.85	0.05	17.00	0.000

ה. מהו המכפיל הדינמי בתקופה t-8?

3) הקשר בין כמות הכסף לבין רמת האינפלציה במשק נאמד בסדרה עתית על ידי המשוואה הבאה:

1. $M_t = \alpha + \beta \cdot P_t + U_t$.

כאשר:

M_t - כמות הכסף במשק בחודש t.

P_t - מדד המחירים לצרכן במשק בחודש t.

משוואה (1) נאמדה בפלט מס' 1.

א. לפי מבחן על הסטטיסטי DW, נראה כי ב- U_t:
   i. לא ניתן לחשב את הסטטיסטי DW מהנתונים הקיימים.
   ii. קיים מתאם סדרתי שלילי.
   iii. קיים מתאם סדרתי חיובי.
   iv. לא קיים מתאם סדרתי.
   v. לא ניתן לקבוע אם המתאם הסדרתי מובהק.

ב. סמנו את התשובה הנכונה בהכרח:
   i. האומדים ליניאריים חסרי הטיה, עקיבים אך לא יעילים.
   ii. האומדים ליניאריים חסרי הטיה, עקיבים ויעילים.
   iii. האומדים מוטים אך עקיבים.
   iv. האומדים חסרי הטיה, אך לא עקיבים.
   v. כל התשובות אינן נכונות.

ג. אומד השונות מוטה ובדיקת השערות
לא תקפה: נכון/לא נכון/אי אפשר לדעת

חוקר טען כי הבעיה שנוצרה במשוואה (1) תיפתר ע"י אמידת המשוואה הבאה:

2. $M_t = \alpha_1 + \beta_1 \cdot P_t + \beta_2 \cdot M_{t-1} + \varepsilon_t$.

כאשר:

M_{t-1} - כמות הכסף בשנה הקודמת.

משוואה (2) נאמדה בפלט מס' 2, כמו כן נאמדה על ידי החוקר המשוואה המופיעה בפלט מס' 3.

ד. הרגרסיה המופיעה בפלט מס' 3 נועדה לבדיקת: _____.
במשוואה: _____.
על ידי מבחן: _____.
גודל הסטטיסטי למבחן הינו (רשום תוצאה מספרית): _____.

ה. לאור תשובתך לסעיף ד' טענת החוקר: נכונה/לא נכונה/אי אפשר לדעת

ו. האומד ל- β_1 במשוואה (2) הוא מוטה אך עקיב: נכון/לא נכון/אי אפשר לדעת

ז. ניתן להשתמש בסטטיסטי DW לבדיקת מתאם סדרתי במשוואה (2): נכון/לא נכון/אי אפשר לדעת.

ח. חשבו את המכפיל הדינמי לשינוי P בתקופה $t-1$.

ט. בדקו את הטענה כי המכפיל הדינמי לשינוי P בתקופה $t-1$ $\left(\frac{\partial M_t}{\partial P_{t-1}}\right)$ הינו 90% מהמכפיל המיידי בטווח הקצר.

י. רשמו את השערת האפס עבור הטענה כי המכפיל בט"א שווה ל-1 . מהו המבחן הסטטיסטי המתאים לבחינת ההשערה?

פלט מס' 1 – משוואה (1):

pendent Variable: m

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	44.03828	44.03828	5145.80	<.0001
Error	103	0.88148	0.00856		
Corrected Total	104	44.91976			

Root MSE	0.09251	R-Square	0.9804
Dependent Mean	8.53854	Adj R-Sq	0.9802
Coeff Var	1.08344		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	1.49372	0.09862	15.15	<.0001
p	1	1.69267	0.02360	71.73	<.0001

Durbin-Watson D 0.208

פלט מס' 2 – משוואה (2):

pendent Variable: m

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	42.19946	21.09973	15988.1	<.0001
Error	101	0.13329	0.00132		
Corrected Total	103	42.33275			

Root MSE	0.03633	R-Square	0.9969
Dependent Mean	8.55393	Adj R-Sq	0.9968
Coeff Var	0.42469		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	1	0.40374	0.06790	5.95	<.0001
m1	1	0.81811	0.03857	21.21	<.0001
p	1	0.28127	0.06633	4.24	<.0001

$M1 = M_{t-1}$

פלט מס' 3:

pendent Variable: res Residual $RES = \hat{\varepsilon}_t$ = (ב) שאיזה משוואה שלנו

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	0.00032896	0.00010965	0.08	0.9695
Error	99	0.13173	0.00133		
Corrected Total	102	0.13206			

Root MSE	0.03648	R-Square	0.0025
Dependent Mean	0.00033853	Adj R-Sq	-0.0277
Coeff Var	10775		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	0.03298	0.08192	0.40	0.6881
p		1	0.02640	0.07979	0.33	0.7415
m1		1	-0.01672	0.04705	-0.36	0.7230
res1		1	-0.01304	0.11039	-0.12	0.9062

$RES1 = \hat{\varepsilon}_{t-1}$

תשובות סופיות:

1) א. מכפיל מיידי: $\frac{\partial Y_t}{\partial X_t} = 0$, מכפיל טווח ביניים: $\frac{\partial Y_t}{\partial X_{t-j}} = \beta_1\beta_2^{j-1}$, מכפיל של הטווח הארוך: $\frac{\partial Y^*}{\partial X^*} = \beta_1 \cdot \frac{1}{1-\beta_2}$.

ב. מכפיל מיידי: $\frac{\partial Y_t}{\partial X_t} = \beta_0$, מכפיל טווח ביניים: $\frac{\partial Y_t}{\partial X_{t-j}} = \beta_2^{j-1}\left(\beta_1 + \beta_0\beta_2\right)$,

מכפיל של הטווח הארוך: $\frac{\partial Y^*}{\partial X^*} = \frac{\beta_1 + \beta_0\beta_2}{1-\beta_2}$.

2) א. ii , חיובי. ב.i. לא נכון. ii. נכון. iii. לא נכון.
iv. נכון. ג. ii. ד. יש עדות לכך.

ה. $\frac{\partial Y_t}{\partial X_{t-j}} = 0.068$.

3) א. iii. ב. i. ג. נכון.
ד. קיום מתאם סדרתי במשוואה (2).
מבחן LM.
$LM_{stat} = 0.2575$.

ה. נכונה. ו. נכון. ז. לא נכון. ח. $\frac{\partial M_t}{\partial P_{t-i}} = \beta_1\beta_2^i$.

ט. יש עדות לכך. י. $H_0: \beta_1 = 1-\beta_2$, WALD.

מבוא לאקונומטריקה ב

פרק 7 - שאלות חזרה למבחן מבוססות על תוכנת STATA

תוכן העניינים

1. כללי 63

שאלות חזרה למבחן מבוססות על תוכנת STATA:

שאלות:

1) כדי לבדוק האם יש קשר בין שכר המורים וההוצאה לתלמיד בבי"ס ציבוריים נאמד המודל הבא: $Pay_i = \beta_1 + \beta_2 Spend + u_i$. כאשר pay הוא שכר המורים ו-Spend מייצג את ההוצאה לתלמיד שנמדדו בדולרים לשנה:

```
. reg pay  spending
```

Source	SS	df	MS
Model	608555015	1	608555015
Residual	264825250	49	5404596.94
Total	873380265	50	17467605.3

Number of obs	= 51
F(1, 49)	= 112.60
Prob > F	= 0.0000
R-squared	= 0.6968
Adj R-squared	= 0.6906
Root MSE	= 2324.8

pay	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
spending	3.307585	.3117043	10.61	0.000	2.681192	3.933978
_cons	12129.37	1197.351	10.13	0.000	9723.204	14535.54

א. פרשו את הרגרסיה (רמז: משמעות האומדים לפרמטרים, האם התוצאות מובהקות?). האם יש הגיון כלכלי לתוצאות שהתקבלו?

ב. מצאו תחזית נקודתית לשכר בממוצע אם ההוצאה לתלמיד היא 5000$.

ג. נבדקו מדדים תיאוריים לגבי המשתנה הב"ת:

```
. su  spending
```

Variable	Obs	Mean	Std. Dev.	Min	Max
spending	51	3696.608	1054.761	2297	8349

ד. הסבירו את משמעות ההנחות הקלאסיות במונחי המשתנים.

ה. האם ההוצאה הממוצעת לתלמיד מסבירה הרבה משונות שכר המורים?

ו. מהו הפער החזוי בשכר המורים בין שני בתי ספר שהפער בהוצאה לתלמיד ביניהם הוא 20$?

ז. האם ניתן לומר על סמך התוצאות כי המודל מובהק? אם כן, באיזה רמת מובהקות?

ח. חוקר ביצע מבחן שמטרתו לבדוק האם המודל יוצא מראשית הצירים:

i. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן.

ii. ערך t סטטיסטי לבדיקת המבחן.

iii. מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת סמך של 95%.

ט. חוקר ביצע מבחן שמטרתו לבדוק כי עבור עליה בדולר אחד בהוצאה השנתית הממוצעת לתלמיד שכר המורים הממוצע עולה בלפחות 3 דולר:

i. השערת האפס וההשערה האלטרנטיבית לבדיקת המבחן.

ii. ערך t סטטיסטי לבדיקת המבחן.

iii. ערך t קריטי לבדיקת המבחן.

iv. מסקנת המבחן: דוחים/לא דוחים את H_0 ברמת ביטחון של 95%.

י. חשבו רווח בר סמך לאמידת השיפוע ברמת סמך של 90%.

יא. חוקר רצה לאמוד את המשוואה בשקלים במקום בדולרים ולכן אמד את המשוואה הבאה: $\hat{pay}1 = \alpha_0 + \alpha_1 spend1$.

הניחו כי שער הדולר עומד על 3.5 ₪.

i. מהי הפקודה ב-stata ליצירת המשתנים החדשים?

ii. מה יהיה האומדן ל- α_1 ? מה יהיה האומדן לסטיית התקן של α_1 ?

על אותם הנתונים נאמדה גם המשוואה הבאה:

$Pay_i = \beta_0 + \beta_1 Spend + \beta_2 Spend2 + u_i$, כאשר: $spend2 = spend^2$.

יב. כתוב את הפקודה ב-STATA ליצירת המשתנה החדש.

יג. מתוצאות האמידה התקבל ש:

3. $\hat{pay}_i = 23128.44 + 1.223 spend_i - 0.0000635 spend2$

(9.55) (10.55) (8.78)

הערכים בסוגריים הם ערכי t סטטיסטי.

השערת האפס לבדיקה האם יש השפעה ליניארית של ההוצאה על התשלום למורים? T סטטיסטי?

מסקנת הבדיקה: דוחים/לא דוחים את H_0 .

יד. מהי הפקודה ב-STATA לבדיקת המתאם בין שני המשתנים הב"ת?

טו. בהינתן תוצאות המודל כיצד נראה פרופיל השכר כפונקציה של ההוצאה על התלמיד (אין צורך בציור מדויק רק במגמה).

טז. מהי ההוצאה לתלמיד שאחריה שכר המורים מתחיל לרדת?

יז. התקבל מתאם של 0.97 בין שני המשתנים הב"ת. לנוכח המתאם הגבוה החוקר טען כי האומדים במשוואה (3) הם מוטים.

הטענה נכונה/הטענה אינה נכונה.

יח. לאור תוצאות האמידה של משוואה (3) ניתן להסיק כי האומדים של משוואה (1) הינם מוטים.

הטענה נכונה/הטענה איננה נכונה.

יט. הניחו כי יש הטרוסקדסטיות במשוואה (3).

מה יהיו ההשלכות לגבי האומדים של המשוואה:

i. האומדים יהיו חסרי הטיה ויעילים.

ii. האומדים יהיו חסרי הטיה אך לא יעילים.

iii. האומדים יהיו מוטים אך יעילים.

iv. האומדים יהיו מוטים ולא יעילים.

2) נתונים 2 המודלים הבאים:

1. $Y_t = \alpha + \beta X_t + u_t$.
2. $Y_t = \beta X_t + u_t$.

א. בהנחה כי R^2 של המודל הראשון גבוה מזה של המודל השני, האם ניתן לומר שהמודל הראשון טוב יותר?
הניחו כי מודל (1) הוא המודל האמיתי.

ב. מהן תכונות האומד לשיפוע של מודל (2)? מוטה/חסר הטיה?

ג. האם ניתן לומר כי האומד הינו עקיב?

ד. חוו דעתכם על הטענה: "על סמך משפט גאוס מרקוב, ניתן להסיק כי אומד הריבועים הפחותים של משוואה (1) הינו יעיל יותר מאומד הריבועים הפחותים של משוואה (2)". הסבירו את תשובתכם.

ה. חשבו את שונות האומד.

ו. לאיזה משני המודלים יהיה אומד לשיפוע גבוה יותר? לראשון/לשני/לא ניתן לדעת.

3) על מנת לאמוד את הקשר שבין השכלה (בשנים) להכנסה (באלפי שקלים) נאמדו שני המודלים הבאים:

```
. reg wage educ
```

Source	SS	df	MS
Model	7888.51144	1	7888.51144
Residual	31092.9858	998	31.1552964
Total	38981.4972	999	39.0205177

Number of obs	1000
F(1, 998)	253.20
Prob > F	0.0000
R-squared	0.2024
Adj R-squared	0.2016
Root MSE	5.5817

wage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	1.138517	.0715497	15.91	0.000	.998112	1.278922
_cons	-4.912181	.9667875	-5.08	0.000	-6.80935	-3.015011

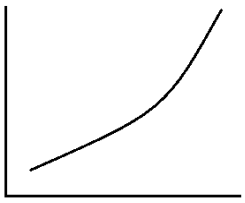
```
. reg lwage educ
```

Source	SS	df	MS
Model	65.5213155	1	65.5213155
Residual	239.767622	998	.240248118
Total	305.288937	999	.305594532

Number of obs	1000
F(1, 998)	272.72
Prob > F	0.0000
R-squared	0.2146
Adj R-squared	0.2138
Root MSE	.49015

lwage	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
educ	.1037608	.0062831	16.51	0.000	.0914313	.1160904
_cons	.7883743	.0848975	9.29	0.000	.6217761	.9549724

א. ידוע כי ניתן לתאר את הקשר בין שכר להשכלה על ידי הגרף הבא:



באיזה מודל כדאי לבחור? תנו 3 נימוקים לבחירתכם.

ב. מהי התשואה להשכלה על סמך המודל הנבחר?

ג. על פי מודל (1), מהי גמישות השכר ביחס להשכלה בנקודת הממוצעים: (7,12)?

ד. על בסיס נתוני המדגם חושב השכר הממוצע עבור התצפיות שלהן 16 שנות השכלה.

```
. su  wage if educ==16
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wage	186	13.30328	7.575015	2.54	60.19

הממוצע הוא 13.303.

שימו לב: סימן ה- "=" משמש להשוואה בין משתנים (למשל: gen price_new= price*1.2), לעומת זאת, סימן ה- "==" משמש להגדרת ערך משתנה (l educ if educ == 8).

i. מהי התחזית לשכר לעובד עם 16 שנות השכלה עבור כל אחד מהמודלים?

ii. איזה מהמודלים נותן תחזית נקודתית מדויקת יותר:
המודל הראשון/המודל השני/ לא ניתן להשוות.

ה. איזה מודל מסביר חלק גדול יותר של השונות של המשתנה התלוי:
המודל הראשון/המודל השני/לא ניתן להשוות.

ו. לפי האומדים בפלט (לא לפי מבחן סטטיסטי) התשואה להשכלה חיובית בהכרח: נכון/לא נכון.

ז. רצו לבחון כיצד השכר מושפע גם מוותק העובד ולכן הכניסו את המשתנה: log(vetek). מה משמעות מקדם השיפוע של המשתנה החדש במודל הראשון ובמודל השני.

ח. בהתייחס למודל השני תוצאות האמידה מראות כי הן התשואה להשכלה והן גמישות הוותק אינם מובהקים. יחד עם זאת הרגרסיה עם שני המשתנים הב"ת יצאה מובהקת. כיצד ניתן להסביר זאת? האם מומלץ להשמיט את שני המשתנים ביחד מהמודל (לאמוד את השכר באמצעות משתנים אחרים)? האם מומלץ לאמוד את המודל ללא משתנה הוותק?

4) על מנת לבחון את פונקציית הייצור של אורז נאמד המודל הבא:

$\ln(PROD) = \beta_1 + \beta_2 \ln(AREA) + \beta_3 \ln(LABOR) + \beta_4 \ln(FERT) + \varepsilon_i$, כאשר:

PROD - כמויות אורז מדושן (נמדד בטונות).

AREA - גודל החלקות בהם האורז נשתל (נמדד בעשרות דונמים).

LABOR - סה"כ ימי עבודה של עובדים ובני משפחה (של החקלאי).

FERT - כמויות דשן בשימוש (נמדד בק"ג).

להלן תוצאות האמידה:

```
. reg lprod larea llabor lfert
```

Source	SS	df	MS	Number of obs =	352
Model	226.084875	3	75.361625	F(3, 348) =	646.51
Residual	40.5653554	348	.116567113	Prob > F =	0.0000
				R-squared =	0.8479
Total	266.65023	351	.759687266	Adj R-squared =	0.8466
				Root MSE =	.34142

lprod	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
larea	.3617359	.0639678	5.65	0.000	.2359237	.4875481
llabor	.4328479	.0668825	6.47	0.000	.301303	.5643928
lfert	.2095023	.0382654	5.47	0.000	.1342417	.2847628
_cons	-1.546786	.2556536	-6.05	0.000	-2.049607	-1.043966

א. בחנו את ההשערה כי גמישות הייצור ביחס לגודל החלקות (AREA) שווה לגמישות ביחס לימי העבודה (LABOR). השתמשו ברמת מובהקות של 5%, נסחו את ההשערה בצורה פורמאלית ודווחו את התוצאות תוך שימוש בפלט הבא:

```
. test  larea= llabor

 ( 1)  larea - llabor = 0

       F(  1,   348) =    0.34
            Prob > F =    0.5592
```

ב. כתבו את הפקודות ב-STATA לאמידת הרגרסיה המוגבלת מהסעיף הקודם.

ג. כעת, בחנו את ההשערה המורכבת מההשערה של סעיף א' + ההשערה כי פונקציית הייצור מקיימת תק"ל ($\beta_2 + \beta_3 + \beta_4 = 1$), תוך שימוש בפלט הבא:

```
. gen l_pr_fe=log( prod/ fert)

. gen l_ar_la_fe=log( area* labor/fert^2)

. reg  l_pr_fe  l_ar_la_fe

      Source |       SS       df       MS              Number of obs =     352
-------------+------------------------------           F(  1,   350) =  439.63
       Model |  51.0075377     1  51.0075377            Prob > F      =  0.0000
    Residual |  40.6079092   350  .116022598           R-squared     =  0.5568
-------------+------------------------------           Adj R-squared =  0.5555
       Total |  91.6154469   351  .261012669           Root MSE      =  .34062

------------------------------------------------------------------------------
   l_pr_fe |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----------+------------------------------------------------------------------
l_ar_la_fe |   .3940824    .018795    20.97   0.000     .3571172    .4310477
     _cons |  -1.402958   .0913195   -15.36   0.000    -1.582562   -1.223354
------------------------------------------------------------------------------

.
```

באיזה מבחן אתם משתמשים?

נסחו את ההשערה, את הסטטיסטי והקריטי.

מהי מסקנתכם?

5) חוקרים ביקשו לאמוד את פונקציית החיסכון המצרפי במשק הישראלי.
מפתח שמות משתנים:
GDS87 - חסכון מקומי גולמי.
GDP87 - תוצר מקומי גולמי.
GC87 - הוצאות הממשלה.
כל הנתונים הינם במיליוני ₪ במחירים קבועים של שנת 1987.
נאמדו 2 המודלים הבאים:

```
. reg  gds87  gdp87
```

Source	SS	df	MS
Model	268713647	1	268713647
Residual	60820138.1	24	2534172.42
Total	329533785	25	13181351.4

Number of obs = 26
F(1, 24) = 106.04
Prob > F = 0.0000
R-squared = 0.8154
Adj R-squared = 0.8077
Root MSE = 1591.9

gds87	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp87	.1852456	.0179896	10.30	0.000	.1481169	.2223742
_cons	-2624.027	1018.32	-2.58	0.017	-4725.737	-522.3182

```
. reg  gds87 gdp87 gc87
```

Source	SS	df	MS
Model	307085221	2	153542611
Residual	22448563.7	23	976024.509
Total	329533785	25	13181351.4

Number of obs = 26
F(2, 23) = 157.31
Prob > F = 0.0000
R-squared = 0.9319
Adj R-squared = 0.9260
Root MSE = 987.94

gds87	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
gdp87	.2955591	.0208369	14.18	0.000	.2524547	.3386635
gc87	-.7784411	.1241513	-6.27	0.000	-1.035268	-.5216146
_cons	4217.674	1260.961	3.34	0.003	1609.177	6826.171

א. כתבו את המודל האקונומטרי שנאמד בכל אחת מהאמידות.

ב. איזה מבין שני המודלים הקודמים הייתם מעדיפים? למה?

ג. מהו ההבדל בין המשמעות של האומדן למקדם המשתנה *GDP87* בשני המודלים?

ד. ביחס לאומד במשוואה 2 האומד $\hat{\alpha}_2$ במשוואה 1 יהיה:
מוטה כלפי מעלה/מוטה כלפי מטה/ חסר הטיה/לא ניתן לדעת

ה. ביחס לאומד במשוואה 3 האומד $\hat{\alpha}_2$ במשוואה 1 יהיה:
יעיל/לא יעיל/ לא ניתן לדעת

ו. בהינתן תוצאות משוואה 1 הקורלציה בין שני המשתנים הב"ת היא:
חיובית/שלילית/אפס/לא ניתן לדעת

ז. החוקר החליט להוסיף משתנה המודד את הצריכה הממשלתית במיליוני דולרים במקום בשקלים: $GC87 ואמד את המשוואה:
3. $GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + \beta_4 \$GC87 + u_t$.
האומדים של משוואה 3 יהיו:
חסרי הטיה ויעילים/מוטים ולא יעילים/חסרי הטיה אך לא יעילים/לא מוגדרים.

ח. החוקר החליט להוסיף מדד נוסף לתוצר המקומי. כתוצאה מהוספת המדד הנוסף משתנה התוצר המקומי הפך להיות לא מובהק.
כיצד ניתן להסביר זאת:
   i. מולטיקוליניאריות מלאה.
   ii. מולטיקוליניאריות חלקית.
   iii. הוספת משתנה לא רלוונטי.
   iv. השמטת משתנה רלוונטי.

ט. החוקר טען כי: "אם נוסיף משתנה נוסף לרגרסיה כלשהי אז האומדן ל- σ^2 לעולם לא יעלה". נכון/לא נכון?

י. החוקר טוען כי "אם נוסיף משתנה נוסף לרגרסיה, אז האומדן ל- $\overline{R}^2$ יעלה בהכרח". נכון/לא נכון.

6) מידע שנאסף לאחרונה על מכירות של 880 בתים בסטוקטון, קליפורניה, נמצא בקובץ *Stockton2*.
המשתנים הם:
PRICE - מחיר בית בדולרים.
SQFT - גודל הבית (square feet).
BEDS - מספר חדרי שינה.
BATHS - מספר חדרי שירותים.
AGE - גיל הבית.
STORIES - הקומה של הבית.
VACANT - משתנה דמי המקבל 1 אם הבית היה פנוי בזמן מכירתו ו-0 אחרת.

להלן תוצאות אמידה המתבססת על המשתנים הנ"ל:

```
. gen price1000= price/1000

. gen lprice1000=ln( price1000)

. gen  sqft100= sqft/100

. reg  lprice1000 sqft100 age beds baths stories vacant
```

Source	SS	df	MS		
Model	92.5168833	6	15.4194805	Number of obs =	880
Residual	32.0322735	873	.03669218	F(6, 873) =	420.24
				Prob > F =	0.0000
Total	124.549157	879	.141694149	R-squared =	0.7428
				Adj R-squared =	0.7410
				Root MSE =	.19155

lprice1000	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
sqft100	.0637441	.0020353	31.32	0.000	.0597495	.0677387
age	-.0024514	.000366	-6.70	0.000	-.0031698	-.0017331
beds	-.0848595	.0133702	-6.35	0.000	-.1111011	-.058618
baths	.0089069	.0181165	0.49	0.623	-.02665	.0444638
stories	-.0182728	.0219074	-0.83	0.404	-.06127	.0247245
vacant	-.0803092	.0132448	-6.06	0.000	-.1063045	-.0543138
_cons	3.994605	.037782	105.73	0.000	3.920451	4.068759

א. מהי משוואת הרגרסיה שנאמדה?
ב. פרשו את תוצאות האמידה. דונו בסימנים ובמשמעויות של כל אחד מהמשתנים.
ג. מה ההבדל במחיר הממוצע בין בית פנוי בזמן מכירתו לבין בית המאוכלס בזמן מכירתו?

ד. חוקר אחר הגדיר את משתנה VACANT= משתנה דמי המקבל 1 אם הבית היה מאוכלס בזמן מכירתו ו-0 אחרת.
האם יש צורך לחשב מחדש את משוואת הרגרסיה? כן/לא.

ה. נניח כי רצו לתרגם את תוצאות המודל מ-sqrft למטרים מרובעים.
יחידה אחת של sqrft שווה ל-0.093 מטר רבוע. מה יהיה הערך של β_1 במודל החדש? מה יהיה ערך t סטטיסטי לדחיית H0 של המשתנה החדש?

ו. מה הפער החזוי בין מחיר של בגודל 130 מטר בת 15 שנה לבין דירה בגודל 115 מטר בת 20 שנה, בהינתן שכל שאר המשתנים נותרים קבועים?

ז. מהי רמת הסמך הנמוכה ביותר עבורה ניתן לדחות את הטענה כי $\beta_5 = 0$?

ח. החוקר החליט לאמוד את המשוואה ללא משתנה VACANT כמשתנה מסביר. להלן תוצאות האמידה:

```
. reg  lprice1000 sqft100 age beds baths stories

      Source |       SS       df       MS              Number of obs =     880
-------------+------------------------------           F(  5,   874) =  477.40
       Model |  91.1678763     5  18.2335753            Prob > F      =  0.0000
    Residual |  33.3812805   874  .038193685           R-squared     =  0.7320
-------------+------------------------------           Adj R-squared =  0.7304
       Total |  124.549157   879  .141694149           Root MSE      =  .19543

------------------------------------------------------------------------------
   lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sqft100 |   .0654365   .0020569    31.81   0.000     .0613995    .0694734
          age |  -.0021886   .0003708    -5.90   0.000    -.0029163   -.0014608
         beds |  -.0823237   .0136344    -6.04   0.000    -.1090837   -.0555638
        baths |   .0062791   .0184781     0.34   0.734    -.0299876    .0425459
      stories |  -.0260982   .0223123    -1.17   0.242    -.0698901    .0176938
        _cons |   3.925237   .0367376   106.85   0.000     3.853132    3.997341
------------------------------------------------------------------------------
```

כיצד השמטת המשתנה השפיעה על משוואת הרגרסיה ועל מקדמיה?

ט. בנוסף אמדו על בסיס המדגם הנוכחי את שתי הרגרסיות הבאות:

```
. reg  lprice1000 sqft100 age beds baths stories if  vacant==0

      Source |       SS       df       MS              Number of obs =     415
-------------+------------------------------           F(  5,   409) =  233.63
       Model |  50.040644     5  10.0081288            Prob > F      =  0.0000
    Residual |  17.5202976   409  .042836913           R-squared     =  0.7407
-------------+------------------------------           Adj R-squared =  0.7375
       Total |  67.5609416   414   .16319068           Root MSE      =  .20697

------------------------------------------------------------------------------
   lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sqft100 |   .0684748   .0030057    22.78   0.000     .0625662    .0743833
          age |  -.001996    .0005387    -3.70   0.000    -.003055    -.000937
         beds |  -.0977677   .019958     -4.90   0.000    -.1370009   -.0585346
        baths |   .0193179   .0251737     0.77   0.443    -.030168    .0688038
      stories |  -.0654738   .033737     -1.94   0.053    -.1317933   .0008457
        _cons |   3.979725   .0554426    71.78   0.000     3.870737    4.088713
------------------------------------------------------------------------------

. reg  lprice1000 sqft100 age beds baths stories if  vacant==1

      Source |       SS       df       MS              Number of obs =     465
-------------+------------------------------           F(  5,   459) =  242.82
       Model |  37.2518359     5  7.45036717            Prob > F      =  0.0000
    Residual |  14.0830841   459  .030682101           R-squared     =  0.7257
-------------+------------------------------           Adj R-squared =  0.7227
       Total |   51.33492   464  .110635603           Root MSE      =  .17516

------------------------------------------------------------------------------
   lprice1000 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
--------------+---------------------------------------------------------------
      sqft100 |   .0593134   .0027696    21.42   0.000     .0538708    .0647561
          age |  -.0028851   .0004996    -5.78   0.000    -.0038668   -.0019034
         beds |  -.0678451   .0178018    -3.81   0.000    -.1028283   -.0328619
        baths |  -.0103401    .026461    -0.39   0.696    -.0623399    .0416596
      stories |   .0265265   .0284661     0.93   0.352    -.0294135    .0824665
        _cons |   3.924588   .0472086    83.13   0.000     3.831817     4.01736
------------------------------------------------------------------------------
```

השוו את תוצאות האמידה של שתי הרגרסיות הנ"ל.

י. ערכו מבחן Chow לבדיקת שקילות (יציבות) המקדמים בשתי הרגרסיות מהסעיף הקודם.

7) חוקר מעוניין ללמוד על הקשר בין הכנסה של משפחה לבין מספר שנות הלימוד של הבעל, מספר שנות הלימוד של האישה והימצאות ילדים קטנים בבית. להלן מפתח שמות המשתנים:

FAMINC - הכנסת המשפחה (דולרים בשנה).

HEDU - מספר שנות הלימוד של הבעל.

WEDU - מספר שנות הלימוד של האישה

$$KID6 = \begin{cases} 1 & \text{אם יש ילדים מתחת לגיל 6} \\ 0 & \text{אחרת} \end{cases},$$

הוא אמד שלוש רגרסיות וקיבל את התוצאות הבאות:

```
. */ Table 1

. reg  faminc Heduc Weduc k16

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  3,   424) =   30.43
       Model |  1.4725e+11     3  4.9082e+10           Prob > F      =  0.0000
    Residual |  6.8384e+11   424  1.6128e+09           R-squared     =  0.1772
-------------+------------------------------           Adj R-squared =  0.1714
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   40160

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   3211.526   796.7026     4.03   0.000     1645.547    4777.504
       Weduc |   4776.907   1061.164     4.50   0.000     2691.111    6862.704
         k16 |  -14310.92   xxxxxxxx    xxxxx   xxxx    xxxxxxxxxx   xxxxxxxxx
       _cons |  -7755.331   11162.93    -0.69   0.488    -29696.91    14186.25
------------------------------------------------------------------------------

. */ Table 2

. reg  faminc Heduc Weduc

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  2,   425) =   40.87
       Model |  1.3405e+11     2  6.7027e+10           Prob > F      =  0.0000
    Residual |  6.9703e+11   425  1.6401e+09           R-squared     =  0.1613
-------------+------------------------------           Adj R-squared =  0.1574
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   40498

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   3131.509    802.908     3.90   0.000     1553.344    4709.674
       Weduc |   4522.641   1066.327     4.24   0.000     2426.711    6618.572
       _cons |  -5533.631   11229.53    -0.49   0.622    -27605.97    16538.71
------------------------------------------------------------------------------

. */ Table 3

. reg  faminc Heduc

      Source |       SS       df       MS              Number of obs =     428
-------------+------------------------------           F(  1,   426) =   61.30
       Model |  1.0455e+11     1  1.0455e+11           Prob > F      =  0.0000
    Residual |  7.2654e+11   426  1.7055e+09           R-squared     =  0.1258
-------------+------------------------------           Adj R-squared =  0.1237
       Total |  8.3109e+11   427  1.9463e+09           Root MSE      =   41297

------------------------------------------------------------------------------
      faminc |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       Heduc |   5155.484   658.4573     7.83   0.000     3861.254    6449.713
       _cons |   26191.27   8541.108     3.07   0.002     9403.308    42979.23
------------------------------------------------------------------------------
```

א. פרשו את משמעויות המקדמים ברגרסיה בטבלה הראשונה, התייחסו למובהקות המקדמים (אין צורך בבדיקה פורמאלית של מובהקות).

ב. התקבל כי ממוצע משתנה KL6 במדגם הוא 0.432. מה משמעות נתון זה?

ג. חשבו את סטית התקן הנאמדת של אומד הריבועים הפחותים ל-β_{KL6} מהמודל הנאמד בטבלה הראשונה. הראו את החישובים שלכם ואת הנוסחאות עליהן אתם מתבססים.
רמז: השתמשו באינפורמציה שניתן לבדוק מובהקות המקדם הנ"ל ביותר מדרך אחת.

ד. הסבירו את הסיבה להבדל המשמעותי בין אומדני β_{HEDU} בשלוש הטבלאות? איך אפשר להסביר את העדר השוני (כמעט העדר שוני) באומדני β_{HEDU} ו-β_{WEDU} בטבלה הראשונה והשנייה?

8) הורצו 2 רגרסיות על מדגם בן 400 תצפיות והתקבלו התוצאות הבאות:

1. $\hat{CM}_i = 263.6416 - 0.0056 \cdot PGNP_i - 2.2316 \cdot FLR_i$
   (s.e) (11.5932) (0.0019) (0.2099) $R^2 = 0.7077$
2. $\hat{CM}_i = 168.3067 - 0.0055 \cdot PGNP_i - 1.768 \cdot FLR_i + 12.8686 \cdot TFR_i$
   (s.e) (11.5932) (0.0019) (0.2099) (?) $R^2 = 0.7474$

כאשר:
$CM = Child\ Mortality$ - מס' מקרי המוות של ילדים מתחת לגיל 5 לכל 1000 לידות חיים.
$PGNP = Per\ Capita\ GNP$ - תוצר לנפש במחירים קבועים בדולרים.
$FLR = Female\ Literacy\ Rate$ - אחוז נשים שיודעות לקרוא ולכתוב.
$TFR = Total\ Fertility\ Rate$ - מס' הלידות הממוצע לאישה במדינה.
הנתונים הם עבור 64 מדינות.

א. כיצד ניתן לפרש את המקדם למשתנה TFR? אפריורית, האם תצפו לקשר חיובי/שלילי בין CM ל-TFR? הסבירו.

ב. האם ערכי המקדמים של המשתנים $PGNP, FLR$ מרגרסיה 1 שונים מאלו ברגרסיה 2? אם כן, מה יכולה להיות הסיבה/ות לשינוי זה?

ג. חוו דעתכם על הטענה הבאה: "כיוון ש: FLR ו-TFR כל כך מתואמים אין לשים אותם באותה הרגרסיה".

ד. באיזה מודל תבחרו מבין השניים? באיזה מבחן סטטיסטי יש להשתמש כדי לענות על שאלה זו? הראו חישובכם. (רמז: הביעו את הסטטיסטי של מבחן F במונחי R^2).

ה. האם תוכלו לחשב את סטיית התקן הנאמדת של המקדם למשתנה TFR? (רמז: היזכרו בקשר בין התפלגות T להתפלגות F).

ו. האם ניתן להשוות את מקדם ההסבר של שתי הרגרסיות? האם ניתן להשוות את מקדם ההסבר המתוקנן? אם כן, השוו ודווחו על התוצאות.

ז. ענו על השאלות התיאורטיות הבאות:

i. איזה מן הגורמים הבאים יכול לגרום לכך שאומדי OLS יהיו מוטים:
   1. הטרוסקדסטיות.
   2. השמטת משתנה מסביר רלוונטי.
   3. מקדם מתאם גבוה מאוד בין שני משתנים מסבירים במודל.

ii. איזה מהגורמים הנ"ל יכול לגרום לכך שסטטיסטי t של OLS לא יהיה תקף?

iii. התייחסו לטענה הבאה: "אם האומדים הינם עקיבים הם יהיו בהכרח גם חסרי הטיה". נכון/לא נכון.

iv. אם נתון ש-u לא מתפלג נורמאלית אז אמידת המשוואה בשיטת OLS תניב אומדים שאינם עקיבים. נכון/לא נכון.

9) חוקרת רצתה לבדוק עונתיות במחירי הירקות. לשם כך הגדירה את משתני הדמי הבאים:

D_1 יקבל את הערך 1 אם מדובר באביב ו-0 אחרת.

D_2 יקבל את הערך 1 אם מדובר בקיץ ו-0 אחרת.

D_3 יקבל את הערך 1 אם מדובר בסתיו ו-0 אחרת.

D_4 יקבל את הערך 1 אם מדובר בחורף ו-0 אחרת.

כאשר:

V_t - מדד מחירי הירקות.

p_t - מדד המחירים לצרכן.

לשם כך אמדה את הרגרסיה הבאה על פני 30 שנה:

$$V_t = \alpha_0 + \alpha_1 D_{1t} + \alpha_2 D_{2t} + \alpha_3 D_{3t} + \alpha_4 \cdot P_t + u_t .$$

א. מדוע לא הכניסה החוקרת למשוואת הרגרסיה את משתנה D_4?

תוצאות האמידה שהתקבלו הן:

$$V_t = 1379.11 + 99.18 D_{1t} + 2209.47 D_{2t} - 476.56 D_{3t} + 489.92 \cdot P_t \quad , \quad R^2 = 0.0844 .$$

נאמדה בנוסף גם המשוואה הבאה: $V_t = \beta_0 + \beta_1 \cdot P_t + \varepsilon_t$.

תוצאות האמידה שהתקבלו: $V_t = 11114.14 + 536.36 \cdot P_t$, $R^2 = 0.0670$.

ב. על סמך התוצאות שהתקבלו, דרגו את העונות לפי רמת המחיר הבסיסית שלהן. הציגו לכל עונה את המיקום שלה ואת רמת המחיר הבסיסית כפי שבא לידי ביטוי במודל.

ג. בדקו את ההשערה כי עונתיות לא משפיע על מחיר הירקות.

ד. כתבו את ההשערות הבאות:

i. מדד מחירי הירקות זהה בחורף ובאביב.

ii. מדד מחירי הירקות גבוה בקיץ מאשר בחורף ביותר מ-600.

ה. האם יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן (בהנחה שהמחיר ההתחלתי של הירקות זהה בין עונות השנה)?

10) חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.
החוקר בדק האם הפונקציה של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את המשוואות הבאות:

1. $NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t}$.

2. $NUM_t = \alpha + \beta_1 \times TYPE_t + \beta_2 \times AVGD_t + \beta_3 \times (AVGD \times TYPE)_t + U_t$.

כאשר:

NUM_t - מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה.

$AVGD_t$ - נפח התנועה בקטע כביש t ליום באלפים.

$TYPE_t$ - משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר, ו- 0 כאשר הכביש לא מהיר.

תוצאות אמידת המשוואות מוצגות להלן:

1. $NUM_t = 0.739 + 0.0233 \cdot AVGD_t$.

2. $NUM_t = 0.14978 + 1.40311 \cdot TYPE_t + 0.002877 \cdot AVGD_t - 0.008 \cdot (AVGD \cdot TYPE)_t$.

1. $ESS = 20963$, $Pt_{\hat{\alpha}} = 0.0019; Pt_{\hat{\beta}} = 0.0001$.

2. $ESS = 20759$, $Pt_{\hat{\alpha}} = 0.6534; Pt_{\hat{\beta}_1} = 0.0067; Pt_{\hat{\beta}_2} = 0.0001; Pt_{\hat{\beta}_3} = 0.1283$.

א. בדקו את טענת החוקר.

ב. מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על 4 אלפי מכוניות ליום בקטע הכביש האמור?

ג. מהי השערת האפס לבדיקת הטענה?

ד. הרגרסיה המוגבלת "תחת H_0" למבחן F (WALD) הינה:

$$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + \nu$$.

כאשר:

$Z_0 =$ _____

$Z_1 =$ _____

$Z_2 =$ _____

$Z_4 =$ _____

11) ברשותכם נתונים על מחירי ארוחת ביג מק במסעדות מקדונלדס ברחבי הארץ ב-1/1/2008 וב-1/1/2009. חלק מהחנויות מנוהלות ע"י הרשת והשאר מנוהלות ע"י זכיינים. אתם מעוניינים לבדוק את השפעתה של פסיקת בית משפט מ-7/2008 אשר מאפשרת לרשת לקבוע מחיר מקסימום עבור מוצרים הנמכרים בחנויות המנוהלות ע"י זכיינים.

לפניך הנתונים הבאים:

$Price$ - מחיר ארוחת ביג-מק.

$$D_{2009}=\begin{cases}0 & year=2008\\1 & year=2009\end{cases}$$

$$D_{Franchise}=\begin{cases}0 & \text{חנות רשת}\\1 & \text{חנות זכיין}\end{cases}$$

נתון המודל הבא המתאר את המחיר:

$$Price_i=\beta_1+\beta_2 D_{2009}+\beta_3 D_{Franchise}+\beta_4 D_{2009}*D_{Franchise}+\varepsilon_i .$$

א. מה המקדם שמגלם את הפרש השכר הצפוי בין חנות רשת לחנות זכיין כשכל יתר המשתנים מוחזקים קבוע?

ב. נסחו את H_0 ו-H_1 לבדיקת ההשערות הבאות:

i. מחיר ארוחת ביג מק ב-2008 היה זהה בחנויות המנוהלות ע"י הרשת ובחנויות המנוהלות ע"י זכיינים.

ii. פסיקת בית המשפט שינתה את פער המחירים בין חנויות המנוהלות ע"י הרשת לבין חנויות המנוהלות ע"י זכיינים בין השנים 2008 ו-2009.

ג. הגדירו את משתנה האינטראקציה על ידי הגדרת 4 הקבוצות הבאות:

D_1 - רשת לפני 2009.

D_2 - רשת אחרי 2009.

D_3 - זכיין לפני 2009.

D_4 - זכיין אחרי 2009.

i. מהי הפקודה הרלוונטיות ב-STATA ליצירת D_1?

ii. כמה ממשתני הדמי יש להכניס לתוך הרגרסיה?

iii. כתבו את משוואת הרגרסיה.

iv. נסחו שוב את ההשערות של סעיף א'.

v. חוקר טוען כי התוצאות שהתקבלו בבדיקת ההשערות על ידי הגדרת משתני הדמי בשני האופנים יהיו זהות. נכון/לא נכון.

12) היצע העבודה של נשים נשואות היה נושא מרכזי במחקר הכלכלי.
לצרוך אמידת היצע זה נבחר המודל הבא:

$$HOURS_i = \beta_1 + \beta_2 WAGE_i + \beta_3 EDUC_i + \beta_4 AGE_i + \beta_5 KIDSL6 + \beta_6 KIDS618 + \beta_7 NWIFEINC_i + u_i$$

כאשר:

$HOURS$ - היצע העבודה בשעות.

$WAGE$ - שכר לשעה.

$EDUC$ - מספר שנות הלימוד.

AGE - גיל.

$KIDSL6$ - מספר הילדים בבית מתחת לגיל 6.

$KIDS618$ - מספר הילדים בגילאים 6-18.

$NWIFEINC$ - הכנסת משק הבית ממקורות שאינם מעבודתה של האישה.

א. מהם הסימנים שתצפו לקבל בכל אחד מהמקדמים?

ב. הסבירו מדוע לא ניתן לאמוד את משוואת ההיצע הנ"ל בשיטת הריבועים הפחותים. איזה מתכונות אר"פ עלולות להיפגע?

ג. אם כתוצאה מאמידת המשוואה בשיטת OLS התקבל כי $\hat{\beta}_2 > 0$, האם ניתן להסיק כי $\text{cov}(wage_i, hours_i) > 0$ במדגם?

ד. החוקר טען שכדאי להשמיט את משתנה ה-wage בשל היותו מתואם עם הטעויות במודל. לטענתו, השמטתו תאפשר לאמוד את המשוואה ב-OLS ולקבל אומדים חסרי הטיה. חוו דעתכם על טענתו.

ה. הניחו כי אנחנו משתמשים בניסיון של האישה בשוק העבודה ($EXPER$) ובריבועו ($EXPER^2$) כמשתני עזר למשתנה $WAGE$. הסבירו מדוע משתני העזר הללו עונים על הדרישות שלנו ממשתני עזר וכיצד ניתן לבדוק זאת.

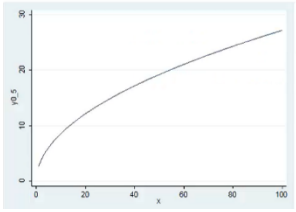
ו. האם ניתן לאמוד את משוואת ההיצע על ידי שימוש במשתני העזר הנ"ל? הסבירו.

ז. תארו את השלבים (לא בפקודות מחשב) שתבצעו כדי לקבל את האומדים בשיטת TSLS.

ח. כתבו את פקודת המחשב לביצוע TSLS ב-STATA.

תשובות סופיות:

1) א. משמעות: ראו סרטון, תוצאות: מובהקות, היגיון כלכלי: יש.
ב. $\hat{pay} = 28,666.87$. ג. $\hat{pay} = 24,355.9$. ד. ראו סרטון.
ה. כן. ו. 66.15. ז. כן, $pv = 0.00$.
ח.i. $H_0 : \beta_1 = 0$ / $H_1 : \beta_1 \neq 0$. ii. $t = 10.13$. iii. דוחים.
ט.i. $H_0 : \beta_2 = 3$ / $H_1 : \beta_2 > 3$. ii. $t = 0.98$. iii. $t_{(0.05,48)} = 2$.
iv. לא דוחים.
י. $(2.90 < \beta_2 < 3.71)$. יא.i. $gen\ pay_1 = pay * 3.5$ / $gen\ spend_1 = spend * 3.5$. ii. אין שינוי.
יב. $gen\ spend_2 = spend * spend$. יג. $H_0 : \beta_2 = 0$, $t = -8.78$, דוחים.
יד. $Corr\ spend\ spend_2$. טו. טז. $X^* = 9,629.92$.



יז. אינה נכונה. יח. נכונה. יט. ii.
2) א. לא ניתן להשוות. ב. מוטה. ג. לא. ד. ראו סרטון.
ה. $V(\hat{\beta}) = \frac{\sigma_u^2}{\sum X_t^2}$. ו. לשני.
3) א. במודל השני. ב. ראו סרטון. ג. $\eta_{y,x} = 1.95$.
ד.i. $WAGE = 13.304$, $WAGE_F = 11.571$. ii. המודל הראשון.
ה. לא ניתן להשוות. ו. לא נכון. ז. ראו סרטון.
ח. לא מומלץ.
4) א. מבחן WALD: $prob(F - ST) = 0.5592 > 0.05$, לא דוחים.
ב. $Gen\ Z_1 = \ln(AREA) + \ln(LABOR)$ / $\text{Reg}\ \ln(PROD)\, Z_1 \ln(FERT)$.
ג. מבחן F.
השערה: $H_0 : \beta_2 + \beta_3 + \beta_4 = 1, \beta_2 = \beta_3 \Leftrightarrow \beta_4 = 1 - 2\beta_2, \beta_2 = \beta_3$ / $H_1 : else$.
סטטיסטי: $F - st = 0.1825$.
5) א. המודל הראשון: $GDS87_t = \alpha_1 + \alpha_2 GDP87_t + \varepsilon_t$.
המודל השני: $GDS87_t = \beta_1 + \beta_2 GDP87_t + \beta_3 GC87_t + u_t$.
ב. המודל השני. ג. ראו סרטון. ד. מוטה כלפי מטה.

ה. לא יעיל. ו. שלילית. ז. לא מוגדרים.
ח. ii. ט. לא נכון. י. לא ניתן לדעת.

6) א.

$$\ln(PRICE1000) = \beta_0 + \beta_1 sqft100 + \beta_2 age + \beta_3 beds + \beta_4 baths + \beta_5 stories + \beta_6 vacant + \varepsilon_i$$

ב. ראו סרטון. ג. $e^{-0.080309} = 0.9228$. ד. לא.
ה. $sqrft1 = \frac{sqrft}{0.093}$, t לא ישתנה. ו. 0.957. ז. 0.404.
ח. ראו סרטון. ט. ראו סרטון. י. ראו סרטון.

7) א. ראו סרטון. ב. פרופורציה. ג. $se(\hat{\beta}_{KL6}) = 5003.818$.
ד. ראו סרטון.

8) א. ראו סרטון, חיובי. ב. שונים, מולטיקולינאריות. ג. ראו סרטון.
ד. השני, מבחן F. ה. כן. ו. מקדם ההסבר: אין טעם להשוות,
מקדם ההסבר המתוקנן: כן. ז. i. 2. ii. 1 ו-2.
iii. לא נכון. iv. ראו סרטון.

9) א. הייתה נוצרת מולטיקולינאריות מלאה.
ב. סתיו (902.58), חורף (1379.11), אביב (1478.29) וקיץ (3588.58).
ג. אין עדות. ד. i. $\begin{matrix} H_0: \alpha_1 = 0 \\ H_0: \alpha_1 \neq 0 \end{matrix}$. ii. $\begin{matrix} H_0: \alpha_2 = 600 \\ H_0: \alpha_2 > 600 \end{matrix}$.
ה. לא ניתן לבדוק.

10) א. ראו סרטון. ב. $NUM_t = 1.532398$. ג. $\begin{matrix} H_0: \beta_2 + \beta_3 = 2 \cdot \beta_2 \\ H_0: \beta_3 = \beta_2 \end{matrix}$.

ד.
$$\begin{matrix} Z_0 = NUM_t \\ Z_1 = TYPE_t \\ Z_2 = (AVGD_t + AVGD \cdot TYPE)_t \\ Z_4 = 0 \end{matrix}$$

11) א. $\beta_3 : D_{franchise}$. ב. i. $\begin{matrix} H_0: \beta_3 = 0 \\ H_1: else \end{matrix}$. ii. $\begin{matrix} H_0: \beta_4 = 0 \\ H_1: \beta_4 \neq 0 \end{matrix}$.
ג. i.
$$\begin{matrix} genD_1 = 1 \; if \; D_{franchise} == 0 \; and \; D_{2009} == 0 \\ replace \; D_1 == 0 \; if \; D_1 == \end{matrix}$$
ii. 3. iii. $\Pr ice = \alpha_1 + \alpha_2 D_1 + \alpha_3 D_2 + \alpha_4 D_3 + u_i$.
iv. i. $\begin{matrix} H_0: \alpha_2 - \alpha_4 = 0 \\ H_1: \alpha_2 - \alpha_4 \neq 0 \end{matrix}$. ii. $\begin{matrix} H_0: \alpha_3 - \alpha_2 - \alpha_4 = 0 \\ H_1: \alpha_3 - \alpha_2 - \alpha_4 \neq 0 \end{matrix}$. v. נכון.

12) א. *wage* חיובי, *educ* לא ניתן לדעת, *age* חיובי או שלילי תלוי במדגם, *kidsl6* שלילי, *kids*618 חיובי, *nwifeinc* שלילי. ב. ראו סרטון.
ג. לא. ד. שגויה. ה. ראו סרטון.
ו. כן. ז. ראו סרטון.
ח.

Ivreg HOURS EDUC AGE KIDSL6 KIDSS618 NWIFEINC (WAGE = EXPER, EXSPER²).