

אקונומטריקה ב



תוכן העניינים

1.	מבחן ML	1
2.	בעיות ספציפיקציה	5
3.	תיאוריה מולטיקוליניאריות	6
4.	סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות	9
5.	משתנה דמי	14
6.	תיאוריה הפרת ההנחות קלאסיות	31
7.	הטרוסקדסטיות	32
8.	מתאם סדרתי	40
9.	סיכום מתאם סדרתי והטרוסקדסטיות	51
10.	מודלים דינאמיים	52

אקונומטריקה 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.

$$H_0 : \beta_3 = \beta_4 = 0$$

$$H_1 : \text{OTHERWISE}$$

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

$$\text{הרגרסיה: } Y_t - \hat{Y}_t = \hat{u}_t$$

כעת אומדים את רגרסיית העזר שבה מנסים לנבא את הסטייה מקו הרגרסיה עבור

$$\text{כל תצפית: } \hat{u}_t = \delta_0 + \delta_1 x_{1t} + \delta_2 x_{2t} + \delta_3 x_{3t} + \delta_4 x_{4t} + \omega_t$$

חישוב הסטטיסטי: (R^2 של רגרסיית העזר * מספר התצפיות) $LM_{stat} = R^2 \cdot T$

כלל הכרעה: אם: $LM_{stat} > \chi_m^2$ נדחה את H_0 (מס' ההגבלות ב- H_0).

• שימו לב כי:

עבור המשתנים הנוספים למודל – כל המדדים (הבטות, ערכי t ו- P value) 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	Standard	T for H0:	
		Estimate	Error	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: U , R ועזר ואת הקשר בין מבחן WALD ומבחן LM.

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

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

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

ד. שחזרו בעזרת שתי המשוואות הראשונות (R ו- U) 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. $U=R+\text{עזר}$

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. $WALD_{stat} = 26.962$

אקונומטריקה ב

פרק 2 - בעיות ספציפיקציה

תוכן העניינים

1. תיאוריה 5

בעיות ספציפיקציה:

רקע:

טעויות ספציפיקציה הן טעויות בניסוח משוואת הרגרסיה.

1. הוספת משתנה לא רלוונטי:

$$\text{למשל, המודל האמיתי: } Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$

$$\text{המודל הנאמד (הטעות): } Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon$$

אם נקבל את H_0 במבחן t למובהקות β_3 נסיק כי המשתנה איננו רלוונטי ונאמוד את המודל מחדש הפעם ללא המשתנה השלישי. אולם, גם אם לא נוכל לאמוד מחדש, הימצאותו של משתנה שאיננו רלוונטי במודל הרגרסיה איננה פוגמת ברלוונטיות של המשתנים האחרים במודל ולא בתכונות החיוניות למבחני המובהקות שלהם.

2. השמטת משתנה רלוונטי:

$$\text{למשל, המודל האמיתי: } Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$$

$$\text{המודל הנאמד (הטעות): } Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$$

בהיעדר x_2 , בדיקות ההשערות לפרמטרים של המודל הטעותי אינן תקפות:

אומד לשונות הפרמטרים	אומד ל- α	אומד ל- β_1	
מוטה (כלפי מעלה)	מוטה אלא אם: $\bar{x}_2 = 0$	חסר הטיה	$S_{12} = 0$
	מוטה	מוטה <u>כיוון ההטיה:</u> חיובי: S_{12} ו- β_2 שווי סימן שלילי: S_{12} ו- β_2 מנוגדי סימן	$S_{12} \neq 0$

אקונומטריקה ב

פרק 3 - תיאוריה מולטיקוליניאריות

תוכן העניינים

1. כללי 6

מולטיקוליניאריות:

רקע:

מולטיקוליניאריות היא תופעה סטטיסטית בעייתית המתייחסת למתאם בין המשתנים המסבירים במודל.

נבחין בין מולטיקוליניאריות מלאה לחלקית.

מולטיקוליניאריות מלאה:

מתאם מלא בין המשתנים המסבירים במודל.

הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של

המשתנה המסביר השני: $x_1 = a + bx_2$ (x_1 הוא קומבינציה ליניארית מלאה של x_2)

מכאן ש: $r_{12} = 1$.

- שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת

(למשל: $x_1 = x_2^2$), אז בהכרח: $r_{12} \neq 1$.

במצב של מולטיקוליניאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני. מדוע זה בעייתי?

כיוון שלא ניתן לאמוד את המודל שכן אר"פ אינם מוגדרים.

פתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

מולטיקוליניאריות חלקית:

כאשר יש מתאם גבוה מאוד בין משתנים מסבירים במודל (אך לא מושלם) עלולה להיווצר בעיה של מולטיקוליניאריות חלקית.

מכיוון שיש מתאם גבוה בין המשתנים הב"ת לא נוכל לבדוד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני המשתנה התלוי.

כל אחד מהמשתנים הב"ת "יגזול" מן ההשפעה הייחודית שיש למשתנה הב"ת השני על המשתנה התלוי, כך שבסופו של דבר, למרות שהמודל עם שני המשתנים הב"ת יהיה מובהק, התרומה הייחודית של כל משתנה ב"ת לניבוי התלוי לא תהיה מובהקת.

זיהוי מולטיקוליניאריות חלקית :

1. כאשר קיימת סתירה בין התוצאה במבחן F למובהקות המודל (המודל מובהק) לבין מבחני t למובהקות השיפועים (אף אחד מן השיפועים אינו מובהק).

הסתירה נוצרת כתוצאה מהגדלת השונות של כל אחד מהשיפועים בשל המתאם הגבוה בין ה"ב", באופן שלא מאפשר לדחות את השערת האפס

$$\text{למובהקות השיפועים: } t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}, \quad S_{\hat{\beta}_1}^2 = \frac{MSE}{SSX_1(1-r_{12})}$$

2. רגישות לספציפיקציה – הורדת משתנה ב"ת שאיננו מובהק תהפוך משתנים ב"ת אחרים במודל למובהקים. אם אין בעיה של מולטיקוליניאריות, הורדת משתנים ב"ת שאינם רלוונטיים מהמודל, לא אמורה להשפיע על מובהקותם של המשתנים הב"ת האחרים.

3. סימנים הפוכים – כאשר השיפועים של המשתנים הב"ת מקבלים סימנים הפוכים מכיוון ההשפעה שלהם על המשתנה התלוי. אם למשל, x_1 משפיע חיובית על Y ואילו x_2 משפיע שלילית על Y אבל הם יופיעו במשוואת הרגרסיה עם סימנים הפוכים ($\hat{\beta}_1$ שלילית ואילו $\hat{\beta}_2$ חיובית), יש לחשוד שקיימת בעיה.

השלכות של מולטיקוליניאריות חלקית :

מולטיקוליניאריות חלקית איננה פוגעת בתכונות של אר"פ (הם נותרים ליניאריים, חסרי הטיה, יעילים ועקיבים) ולא באומד השונות של האומדים (שנותר חסר הטיה) כך שבדיקת השערות תוך שימוש באומדים הללו תהיה תקפה (זאת בניגוד למולטיקוליניאריות מלאה).
במובן הזה, בעיה של מולטיקוליניאריות חלקית דומה לבעיה של הוספת משתנה ב"ת שאיננו רלוונטי.

פתרונות למולטיקוליניאריות חלקית :

1. ברוב המקרים נשקול להוריד את אחד המשתנים.
יחד עם זאת, כאשר המובהקות של המשתנים היא גבולית: $1 < t_{\hat{\beta}} < 2$, יתכן ונותר את שניהם בתוך המודל כיוון שבסך הכל יש עליה ב- $AdjR^2$ (לפי חוק חיטובסקי).
2. ניתן לעיתים לאחד את שני המשתנים למשתנה אחד.

שלבי בדירת ההשערות :

1. מבצעים מבחן F לבדיקת מובהקות המודל.
2. במידה והמודל מובהק, מבצעים מבחן t למובהקות כל אחד מהשיפועים.
3. ביצוע מבחן WALD לבדיקת כל השיפועים שלא יצאו מובהקים :
 - א. אם מקבלים את H_0 : אין סתירה בין מבחן WALD למבחני t - אין בעיה של מולטיקוליניאריות חלקית, נוריד את קבוצת המשתנים הלא רלוונטיים מהמודל.
 - ב. אם דוחים את H_0 : יש סתירה בין מבחן WALD למבחני t - קיימת בעיה של מולטיקוליניאריות חלקית, יש להוריד מן המודל כל פעם משתנה אחד ולבצע מבחן WALD בלעדיו, עד שמזהים את המשתנה / משתנים שיש להוריד מהמודל.

אקונומטריקה ב

פרק 4 - סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות

תוכן העניינים

1. כללי 9

סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות:

רקע:

הבעיה	הגדרה	זיהוי	השלכות	פתרון												
הוספת משתנה לא רלוונטי	המודל האמיתי : $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$ המודל הנאמד (הטעות): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$	קבלת H_0 במבחן t למובהקות β_2	ניתן לבצע בדיקת השערות אר"פ $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$ חסרי הטיה אומדי השונות $(S_{\hat{\alpha}}^2, S_{\hat{\beta}_1}^2, S_{\hat{\beta}_2}^2)$ חסרי הטיה	הורדת* המשתנה												
השמטת משתנה רלוונטי	המודל האמיתי : $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ המודל הנאמד (הטעות): $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$	דחיית H_0 במבחן t למובהקות β_2	<table><tr><th>בהיעדר : x_2</th><th>אומד ל- β_1</th><th>אומד ל- α</th><th>אומד לשונות הפרמטרים</th></tr><tr><td>$S_{12} = 0$</td><td>חסר הטיה</td><td>מוטה אלא אם: $\bar{x}_2 = 0$</td><td>מוטה חיובית</td></tr><tr><td>$S_{12} \neq 0$</td><td>מוטה חיובית: S_{12} ו- β_2 שויי סימן מוטה שלילית: S_{12} ו- β_2 מנוגדי סימן</td><td>מוטה</td><td>מוטה חיובית</td></tr></table> לא ניתן לבצע בדיקת השערות	בהיעדר : x_2	אומד ל- β_1	אומד ל- α	אומד לשונות הפרמטרים	$S_{12} = 0$	חסר הטיה	מוטה אלא אם: $\bar{x}_2 = 0$	מוטה חיובית	$S_{12} \neq 0$	מוטה חיובית: S_{12} ו- β_2 שויי סימן מוטה שלילית: S_{12} ו- β_2 מנוגדי סימן	מוטה	מוטה חיובית	הוספת המשתנה
בהיעדר : x_2	אומד ל- β_1	אומד ל- α	אומד לשונות הפרמטרים													
$S_{12} = 0$	חסר הטיה	מוטה אלא אם: $\bar{x}_2 = 0$	מוטה חיובית													
$S_{12} \neq 0$	מוטה חיובית: S_{12} ו- β_2 שויי סימן מוטה שלילית: S_{12} ו- β_2 מנוגדי סימן	מוטה	מוטה חיובית													
מולטיקוליניאריות מלאה	מתאם מלא בין המשתנים המסבירים במודל $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ כאשר : $r_{12} = \pm 1$	אם: $x_1 = a + bx_2$ אז : $r_{12} = 1$	לא ניתן לבצע בדיקת השערות אר"פ $(\hat{\alpha}, \hat{\beta}_1, \hat{\beta}_2)$ בלתי מוגדרים.	הורדת אחד המשתנים												
מולטיקוליניאריות חלקית	מתאם חזק בין המשתנים המסבירים במודל $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$ כאשר : $0.7 < r_{12} < 1$	א. סתירה בין מבחן F ל- t ב. רגישות לספציפיקציה ג. סימנים הפוכים	ניתן לבצע בדיקת השערות אין פגיעה בתכונות אר"פ ושונותם	הורדת** אחד המשתנים או איחודם												

- * במידה והמובהקות גבולית ($1 < t_{\hat{\beta}} < 2$) נשקול להשאיר משתנה לא רלוונטי כי מעלה את $AdjR^2$ (חוק חיטובסקי).
- ** במידה ומובהקותם גבולית ($1 < t_{\hat{\beta}} < 2$) נשקול להשאיר את שניהם בשל העלייה ב- $AdjR^2$ (חוק חיטובסקי).

שאלות:

(1) להלן מודל של שכר W_t , כפונקציה של שנות לימוד S_t :

$$1. W_t = \alpha + \beta \cdot S_t + u_t$$

להלן מודל של שכר W_t , כפונקציה של שנות לימוד S_t ושל גיל A_t :

$$2. W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + v_t$$

כל האומדים חיוביים ומובהקים וקיים קשר שלילי בין גיל להשכלה.

א. $\hat{\beta}_1$ במשוואה (1) הוא:

i. אומד חסר הטיה.

ii. אומד מוטה שלילית.

iii. אומד מוטה חיובית.

iv. אומד מוטה, אך לא ניתן לדעת את כיוון ההטיה.

ב. ניתן להשתמש במבחן t לבדיקת מובהקות

השיפוע במשוואה (1). נכון/לא נכון/לא ניתן לדעת

ג. בנוסף למשתנים במשוואה השנייה, החליט החוקר להוסיף גם את

משתנה הוותק, EXP_t . מכיוון שלא היו בידו נתונים על הוותק, החליט

החוקר להעריכו עבור כל עובד על ידי הגיל של העובד פחות 24 שנים

(מתוך ההנחה שהחיים המקצועיים מתחילים בגיל זה לערך).

להלן משוואה מס' 3:

$$3. W_t = \alpha + \beta_1 \cdot S_t + \beta_2 \cdot A_t + \beta_3 \cdot EXP_t + w_t$$

חוזה דעתך על המשוואה השלישית.

(2) נתונות ארבע משוואות הרגרסיה הבאות (כאשר הסטיות במודל האמיתי

מקיימות את הנחות הרגרסיה הקלאסיות):

$$1. X_{2t} = \lambda + \delta \cdot X_{1t} + V_t : \sum \hat{V}_t^2 = \sum (X_{2t} - \bar{X}_{2t})^2$$

$$2. Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$$

(10.3) (19.8)

$$3. Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + \beta_3 \cdot X_{3t} + W_t$$

(9.9) (17.3) (0.37)

$$4. Y_t = \alpha + \beta_1 \cdot X_{1t} + \sum_t$$

(6.3)

(המספרים בסוגריים הם ערכי t של אומדני המקדמים).

לגבי הטענות הבאות, קבעו לגבי כל טענה אם היא נכונה או לא, והסבירו:

א. האומד של β_1 במשוואה (2) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.

ב. האומד של β_1 במשוואה (3) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.

- ג. האומד של β_1 במשוואה (4) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.
- ד. האומדן $\hat{\beta}_1$ במשוואה (4) זהה ל- $\hat{\beta}_1$ במשוואה (2).
- ה. השונות התיאורטית של האומדן $\hat{\beta}_1$ במשוואה (4) זהה לשונות התיאורטית של $\hat{\beta}_1$ במשוואה (2), אך אומדני השונות שונים.
- ו. האומד ל- α במשוואה (4) הינו חסר הטיה.
- ז. האומד ל- α במשוואה (3) הינו חסר הטיה.
- ח. R^2 של משוואה (2) גדול מ- R^2 של משוואה (3).
- ט. $\bar{R}^2$ של משוואה (2) גדול מ- $\bar{R}^2$ של משוואה (3).

$$(3) \quad Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$$

חו דעתכם על הטענות הבאות (כל סעיף עומד בפני עצמו):

$$R^2 = 0.92 \quad Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t \quad (0.5) \quad (0.3)$$

- הערכים בסוגריים הם ערכי t .
למובהקות הבטות יש טעות במודל
כי המודל מובהק והמקדמים לא :
נכון/לא נכון/לא ניתן לדעת
- ב. בהנחה כי מתקיים : $X_{1t} - 2X_{2t} = 1$ לא ניתן לאמוד
את המודל בשיטת הריבועים הפחותים :
נכון/לא נכון/לא ניתן לדעת
- ג. בהנחה כי מתקיים : $X_{1t} = X_{2t}^2$ לא ניתן לאמוד
את המודל בשיטת הריבועים הפחותים :
נכון/לא נכון/לא ניתן לדעת
- ד. הוכיחו תשובותיכם לסעיפים א' ו-ב'.
- ה. בהנחה כי מתקיים : $r_{12} = 0.98$
- i. לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים :
נכון/לא נכון/לא ניתן לדעת.
 - ii. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.
 - iii. בהנחה שהמודל יצא מובהק אולם הבטות אינן מובהקות וערכי t למובהקות הבטות הן כדלקמן : $t_{\hat{\beta}_1} = 1.31$, $t_{\hat{\beta}_2} = 1.45$, מה יהיה הפתרון הטוב ביותר, לדעתכם, לבעיה במודל (אליה התייחסתם בסעיף ii)?
1. להוריד את x_1 .
 2. להוריד את x_2 .
 3. להוריד את שני המשתנים.
 4. להותיר את שני המשתנים.

תשובות סופיות:

- (1) א. ii. ב. לא נכון. ג. קיימת בעיית מולטיקוליניאריות מלאה.
 (2) א. לא נכון. ב. לא נכון. ג. נכון. ד. נכון. ה. נכון.
 ו. לא נכון. ז. נכון. ח. לא נכון. ט. נכון.
 (3) א. לא נכון. ב. נכון. ג. לא נכון. ד. הוכחה. ה. i. לא נכון.
 ii. מולטיקוליניאריות חלקית. iii. 4.

אקונומטריקה ב

פרק 5 - משתנה דמי

תוכן העניינים

1. כללי 14

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 זה 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 WALT 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 WALT) :
 המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן יכול את
 המדגם המאוחד : $W_t = \alpha + \beta X_t + u_t$

המודל הלא מוגבל (U) כולל את שני חלקי המדגם :
 $ESS_U = ESS_f + ESS_m$
 $DF_U = DF_f + DF_m$

$$CHOW_{stat} = \frac{\frac{ESS_R - (ESS_f + ESS_m)}{DF_R - (DF_f + DF_m)}}{\frac{ESS_f + ESS_m}{DF_f + DF_m}} = WALT_{stat}$$

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 גרסיות נפרדות אנו בודקים הבדל בכל הפונקציה ואילו שיטת משתני הדמי מאפשרת לבדוק הבדל רק בחותך או רק בשיפוע.

סיכום ביניים :

משתנה דמי לכל הפונקציה	משתנה דמי לשיפוע	משתנה דמי לחותך	
$Y_t = \alpha_0 + \alpha_1 D + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha + \beta_0 X_t + \beta_1 DX_t + u_t$	$Y_t = \alpha_0 + \alpha_1 D + \beta \cdot X_t + u_t$	המודל
קיים הבדל בין הקטגוריות במשוואת הרגרסיה כולה (בחותר ובשיפוע).	קיים הבדל בין הקטגוריות בתוספת ל-Y בגין X (בשיפוע).	קיים הבדל בין הקטגוריות ב-Y ההתחלתי (בחותר).	ההשערה במילים
מבחן WALT 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- $H_0 : \alpha_1 = 0$ WALT): $H_0 : \beta_1 = 0$	מבחן t להפרש השיפועים: $H_0 : \beta_1 = 0$	מבחן t להפרש החותכים: $H_0 : \alpha_1 = 0$	בדיקת ההשערה

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:

$$V_t = \text{מדד מחירי הירקות.}$$

$$p_t = \text{מדד המחירים לצרכן.}$$

1. משתני דמי לחותך:

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

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

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:

$$H_0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$

השערות:

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי – מבחן WALD:

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

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

- 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.

$$\text{המודל: } V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t$$

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:

$$H_0: \beta_1 = \beta_2 = \beta_3 = 0$$

השערות:

$$H_1: \text{OTHERWISE}$$

המבחן הסטטיסטי – מבחן WALD:

$$(U) \quad V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1t} P_t) + \beta_2 (D_{2t} P_t) + \beta_3 (D_{3t} P_t) + u_t$$

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

- 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 :

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

בדיקת השערות :

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

המבחן הסטטיסטי - מבחן WALD :

(U) :

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

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

אם דוחים את H_0 , יש לבדוק במבחני WALD האם ההבדל הוא בין החותכים

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

אם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחני t :

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

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 :

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

המודל זה – אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.

ניתן לבדוק השערות על כל אחד מהמשתנים הב"ת האיכותיים בנפרד :

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

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

2. הבדל בחותך עם אינטראקציה :

$$W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$$

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 :

$$H_0 : \alpha_3 = 0$$

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.

$$W_t = \gamma_0 + \gamma_1 D_1 + \gamma_2 D_2 + \gamma_3 D_3 + \delta \cdot S_t + u_t$$

נעזר בטבלה בכדי לנסח את ההשערות לבדיקת האינטראקציה :

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

ההשערות לבדיקת קיום האינטראקציה : $H_0 : \gamma_1 - \gamma_3 = \gamma_2$ או $H_0 : \gamma_1 - \gamma_3 = \gamma_2$ התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך

$$WALD = t^2$$

$$PF = Pt$$

שאלות:

משתנה דמי לחותך:

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

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

$$(S.E) \quad (134) \quad (56) \quad (24)$$

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.

$$W_t = 5200 + 1120 \cdot D$$

$$S_{\hat{\alpha}_1} = 63$$

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:

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

$$(68) \quad (23) \quad (25)$$

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:

$$1. \quad NUM_t = \gamma_1 + \delta_1 \cdot AVGD_t + \varepsilon_{1t} \quad \text{כבישים מהירים בלבד.}$$

$$2. \quad NUM_t = \gamma_2 + \delta_2 \cdot AVGD_t + \varepsilon_{2t} \quad \text{כבישים לא מהירים בלבד.}$$

$$3. \quad NUM_t = \gamma_3 + \delta_3 \cdot AVGD_t + \varepsilon_{3t} \quad \text{שני סוגי הכביש (כל המדגם).}$$

$$4. \quad NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_2 \cdot AVGD_t + \beta_3 \cdot (AVGD \cdot TYPE)_t + U_t$$

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 WALT?

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	Mean	F Value	Pr > F
		Squares	Square		
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	Standard	t Value	Pr > t
		Estimate	Error		
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	Mean	F Value	Pr > F
		Squares	Square		
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	Standard	t Value	Pr > t
		Estimate	Error		
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 בפרט להלן:

Analysis of Variance					
		Sum of	Mean		
Source	DF	Squares	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	Standard	t Value	Pr > t
		Estimate	Error		
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 :
- 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 WALT.
- 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 :
- $$\ln(Y)_t = \lambda_0 + \lambda_1 S + \lambda_2 E + \delta_1 EXP + \delta_2 EXP^2 + \omega_t$$
- 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 נח. 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. $H_0 : \beta_2 + \beta_3 = 2 \cdot \beta_2$

ה. $H_0 : \beta_3 = \beta_2$

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

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

ii. WALD.

(6) א. נכון. ב. יש עדות לכך. ג. יש עדות לכך.

ד. $H_0 : \alpha_3 = \alpha_1 - \alpha_2$ או $H_0 : \alpha_2 = \alpha_1 - \alpha_3$

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

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

ii. $\ln(MWAGE) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$ iii. $EXP_t = \frac{-(\alpha_1 + \alpha_3)}{\beta_1 + \beta_3}$

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

iii. $H_0 : \alpha_2 = \beta_2 = 0$ iv. $H_0 : \beta_2 + \beta_3 = 0$

אקונומטריקה ב

פרק 6 - תיאוריה הפרת ההנחות קלאסיות

תוכן העניינים

1. כללי 31

הפרת ההנחות הקלאסיות:

רקע:

ארבעת הנושאים הבאים עוסקים במצב של הפרת אחת ההנחות הקלאסיות
הדרושות לאמידת הפרמטרים בשיטת OLS :

- הטרוסקדסטיות (הפרת הנחה מס' 5) – שונות קבועה ויחידה לאורך קו
הרגרסיה : $V(u_t) = \sigma^2$.
- מתאם סידרתי (הפרת הנחה מס' 6) – אי תלות בין הטעויות : $\text{cov}(u_t, u_s) = 0$.
- מודלים דינמיים ו-משוואות סימולטניות (הפרת הנחה מס' 4) – אי תלות בין
המשתנים הב"ת לטעויות : $\text{cov}(x, u) = 0$.

בכל אחד מן הנושאים נלמד :

- מהן ההשלכות של הפרת ההנחות הללו על אומדי הריבועים הפחותים.
- מהם המבחנים הסטטיסטיים המשמשים לזיהוי קיומה של ההפרה.
- כיצד נתקן את משוואת הרגרסיה כך שניתן יהיה לאמוד את הפרמטרים
בשיטת OLS.

אקונומטריקה ב

פרק 7 - הטרוסקדסטיות

תוכן העניינים

1. כללי 32

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 שונות 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 שונות 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: $H_0: \sigma_1^2 = \sigma_2^2$
 $H_1: \sigma_1^2 > \sigma_2^2$

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:

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

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

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: $H_0: \alpha_j = \beta_j = \gamma_{jj} = 0$
 $H_1: OTHERWISE$

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 (Z_t) יש לנטרל. לשם כך ניצור משתנה חדש W_t שיהווה השורש ההופכי

$$W_t = \frac{1}{\sqrt{Z_t}}$$

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

$$Y_t = \alpha + \beta X_t + u_t$$

המשוואה המקורית: $Y_t W_t = \alpha \cdot W_t + \beta (X_t W_t) + u_t W_t$

בצורתה המפורשת המשוואה החדשה נראית כך: $\frac{Y_t}{\sqrt{Z_t}} = \alpha \cdot \frac{1}{\sqrt{Z_t}} + \beta \cdot \frac{X_t}{\sqrt{Z_t}} + \frac{u_t}{\sqrt{Z_t}}$

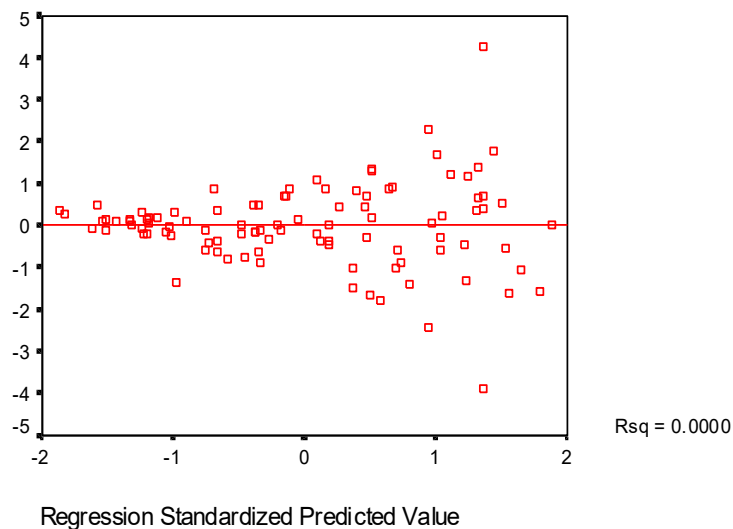
שאלות:

מבחן GQ:

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

Scatterplot

Dependent Variable: Y



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

Source	DF	Sum of Squares	Mean 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

Source	DF	Sum of Squares	Mean 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 מיון את

התצפיות בסדר עולה של מספר העובדים, השמיט $\frac{1}{3}$ מהתצפיות האמצעיות

והריץ שתי גרסיות 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 (ש).
 $X1_SQUARES$ - 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					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model					
Error					
Corrected Total					
		Root MSE	R-Square	0.086942	
		Dependent Mean	Adj R-Sq		
Parameter Estimates					
Variable	DF	Parameter Estimate	Standard 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: $VAR(U_t) = \frac{\sigma^2}{Z_t^2}$ (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 נתוןכי: $V(u_t) = \sigma_t^2 = \begin{cases} X_t \sigma^2 & \Leftrightarrow t \geq 50 \\ X_t^2 \sigma^2 & \Leftrightarrow t \leq 50 \end{cases}$. (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: $H_0: \sigma_1^2 = \sigma_2^2$, חישוב סטטיסטי: 17.62, כלל הכרעה: 2.48.
 $H_1: \sigma_1^2 > \sigma_2^2$

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

(3) יש עדות לכך, השערות: $H_0: \alpha_1 = \beta_1 = 0$, חישוב סטטיסטי: 18.75,
 $H_1: OTHERWISE$

כלל הכרעה: 5.991

(4) קיום הטרוסקדסטיות בנתונים.

$WHITE(LM)$.

RES^2 .

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

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

$LM_{stat} = 1.73$.

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. $W_t = \frac{1}{\sqrt{X_t^2}}$ $t \leq 50$, $W_t = \frac{1}{\sqrt{X_t}}$ $t \geq 50$. ד. לא.

אקונומטריקה ב

פרק 8 - מתאם סדרתי

תוכן העניינים

1. כללי 40

מתאם סדרתי:

רקע:

מתאם סדרתי עוסק במצב שבו מופרת ההנחה (מס' 6) של אי תלות בין הטעויות:
 $cov(u_t, u_s) = 0$ ונוצרת תלות סטטיסטית בין הטעויות במודל: $cov(u_t, u_s) \neq 0$.
 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 סדרתי 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 סדרתי חיובי בין הטעויות ולמשתנים יש
 מגמת זמן (X עולה או יורד עם הזמן) אומד השונות (ESS) יהיה מוטה כלפי מטה
 ואז נקבל: 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} : $cov(u_t, u_{t-1}) \neq 0$.

את המתאם בין הטעויות מסדר ראשון ניתן לנסח באופן הבא: $u_t = \rho \cdot u_{t-1} + \varepsilon_t$
 כך ש:

- $\rho \neq 0$ (כי אם $\rho = 0$ אין מתאם סדרתי).
- $-1 < \rho < 1$ (כי אם חורג מ-1 הטעות הולכת וגדלה עם הזמן).
- 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

$$E(\varepsilon_t) = 0$$

$$V(\varepsilon_t) = \sigma_\varepsilon^2 \quad \text{בניגוד ל-}(u_t) \text{ כך ש:}$$

$$\text{cov}(\varepsilon_t, \varepsilon_{t-s}) = 0$$

המודל יכול שתי משוואות-המשוואה העיקרית והגדרת המתאם הסדרתי (מסדר

$$Y_t = \alpha + \beta X_t + u_t \quad \text{ראשון):}$$

$$u_t = \rho \cdot u_{t-1} + \varepsilon_t$$

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:

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

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: $\rho_{u_t, u_{t-1}} > \rho_{u_t, u_{t-2}}^2 > \rho_{u_t, u_{t-3}}^3 > \dots > \rho_{u_t, u_{t-s}}^s$
בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

$$E(u_t) = 0$$

$$V(u_t) = \sigma_u^2 = \frac{\sigma_\varepsilon^2}{1 - \rho^2}$$

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

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 $K = \text{מס' המשתנים ה"ת במודל}$ ו- $T = \text{מס' התצפיות}$ 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.

$$\begin{aligned} H_0: \rho &= 0 \\ H_1: \rho &> 0, \rho < 0 \end{aligned} \quad \text{השערות:}$$

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

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

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

המודל הלא מוגבל :

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

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 : $\hat{u}_t / x_1, x_2, \dots, x_k; \hat{u}_{t-1}$.
- נחשב סטטיסטי LM : $LM_{stat} = T \cdot R^2$.
- נדחה את H_0 כאשר : $LM_{stat} > \chi_m^2$ כאשר $m = \text{סדר המתאם הסדרתי}$.
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 :

$$\begin{aligned} H_0 : \rho_1 = \rho_2 = \dots = \rho_s = 0 \\ H_1 : \text{OTHERWISE} \end{aligned}$$

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

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 הפרשים 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 : $\varepsilon_t = u_t - \rho \cdot u_{t-1}$, $u_t = \rho \cdot u_{t-1} + \varepsilon_t$

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:

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

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

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: $1, 2, 3, \dots$.

תוצאות האמידה שהתקבלו:

Dependent variable: CLOSE

Analysis of Variance					
F Value	Prob>F	Mean Square	Sum of Squares	DF	Source
	55.78		-----	1	Model
			-----	151	Error
			-----	152	C Total
		0.181	R-square	----	Root MSE
		-----	Adj R-sq	----	Dep Mean
				----	C.V.
Parameter Estimates					
T for H0:		Standard Error	Parameter Estimate	DF	Variable
Parameter=0	Prob> T				
91.047	0.0000	0.0148	1.3474	1	INTERCEP
0.0000	-7.468	0.0001	-0.00075	1	TIME
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 $d_L \leq d_\alpha \leq d_U$ 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	0.82	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

$$Y_t = \alpha + \beta X_t + \varepsilon_t \quad (3) \quad \text{המודל הינו:}$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

u_t - הפרעה מקרית קלאסית.

$T = 100$ וידוע כי:

$$u_t = 0.9u_{t-1}$$

$$d_L = 1.57$$

$$d_U = 1.65$$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

מבחן LM:

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

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

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

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

תוצאות האמידה שהתקבלו:

Dependent variable: RES

		Analysis of Variance			Source
F Value	Prob>F	Mean Square	Sum of Squares	DF	
			-----	2	Model
			-----	150	Error
			-----	152	C Total
		0.855	R-square	-----	Root MSE
		-----	Adj R-sq	-----	Dep Mean
				-----	C.V.

Parameter Estimates

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

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			
Durbin-Watson D				2.235	0.000

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).

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

בניסיון לבדוק האם מתקיים קשר מהסוג הבא: $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \rho_3 u_{t-3} + \varepsilon_t$.

$$\hat{u}_t = \gamma_1 \hat{u}_{t-1} + \gamma_2 \hat{u}_{t-2} + \gamma_3 \hat{u}_{t-3} + \gamma_4 Y_t + \omega_t$$

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						
	Error	:T for H0 Parameter=0	Standard Error	Parameter Estimate	Prob> T	Variable
	1128.38	-1.864		0.069 -2103.53	1	INTERCEP
	0.0510	14.086		0.000 0.71944	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: $\ln \hat{Y}_t = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$, $\hat{\rho} = 0.2$.

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:

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

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

(2) יש עדות לכך.

(3) יש עדות לכך.

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

(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.

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

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

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

אקונומטריקה ב

פרק 9 - סיכום מתאם סדרתי והטרוקדסטיות

תוכן העניינים

1. כללי 51

סיכום מתאם סדרתי והטרוקדסטיות:

רקע:

הטרוסקדסטיות	מתאם סדרתי	
	למשל: $Y_t = \alpha + \beta X_t + u_t$	המשוואה העיקרית של המודל
$V(u_t) = \sigma^2$	$\text{cov}(u_t, u_{t-s}) = 0$	ההנחה הקלאסית המופרת
$V(u_t) = \sigma_t^2$	$\text{cov}(u_t, u_{t-s}) \neq 0$	המצב לאחר ההפרה
$V(u_t) = W_t \sigma^2$	$u_t = \rho u_{t-1} + \varepsilon_t$	המשוואה המאפיינת את ההפרה
מתקבלים אומדים חסרי הטיה ועקיבים, אך תכונת היעילות נפגעת.		מה קורה אם אומדים OLS-ב
מבחן GQ מבחן White	מבחן DW מבחן LM	זיהוי הבעיה
שיטת WLS	שיטת קוקרן – אורקוט (רגרסיית ההפרשים) הכנסת משתנה מוסבר בפיגור (מודל דינמי)	פתרון הבעיה

אקונומטריקה ב

פרק 10 - מודלים דינאמיים

תוכן העניינים

1. כללי 52

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 (מכפילים):

1. מכפל לטווח קצר (מידי):

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

2. מכפל ביניים מסדר j (מכפיל דינמי):

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

3. מכפל טווח ארוך (מצב עמיד):

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

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

$$Y_t = Y_{t-1} = \dots = Y_{t-p} = Y^*$$

$$X_t = X_{t-1} = \dots = X_{t-p} = X^*$$

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 סטטי. הדינמיות תבוא אז לידי ביטוי בטעויות, כלומר במתאם הסדרתי. גרסיית 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:

$$\hat{\beta} \rightarrow \beta + \frac{COV(Y_{t-1}, u_t)}{V(Y_{t-1})}$$

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

אם יש מתאם סדרתי: $COV(Y_{t-1}, u_t) \neq 0 \Leftarrow$ 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. $Y_t = \alpha + \beta_1 X_t + \beta_2 X_{t-1} + \beta_3 X_{t-2} + u_t$

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	Standard	T for H0:	
		Estimate	Error	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:

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

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:

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

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):

Dependent 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):

Dependent 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

$$M_t = M_{t-1}$$

פלט מס' 3:

Dependent Variable: res Residual

$$RES = \hat{\varepsilon}_t = (2) \text{ מודל נורמלי}$$

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 מידי: $\frac{\partial Y_t}{\partial X_t} = 0$, מכפיל טווח ביניים: $\frac{\partial Y_t}{\partial X_{t-j}} = \beta_1 \beta_2^{j-1}$, מכפיל של

$$\frac{\partial Y^*}{\partial X^*} = \beta_1 \cdot \frac{1}{1 - \beta_2} \quad \text{הטווח הארוך:}$$

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

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

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

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

(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. WALT, $H_0: \beta_1 = 1 - \beta_2$.