<u>אקונומטריקה ב</u>

<u>החוג לכלכלה וניהול</u>

תוכן

פרק 1 - מבחני ספציפיקציה- מבחן LM (כופלי לגרנג')

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

לדוגמא:

נניח שיש לנו מודל הכולל 4 משתנים מסבירים <u>(UNRESTRICTED)</u>:

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

לגבי השניים הראשונים אנו בטוחים כי הם רלוונטיים וחייבים להופיע במודל. לגבי השניים האחרונים אנחנו לא בטוחים.

<u>השערות:</u>

$$H0: \beta_3 = \beta_4 = 0$$
$$H1: OTHERWISE$$

<u>המודל המוגבל (RESTRICTED) הינו:</u>

$$Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t}$$

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 <u>רגרסיית העזר</u> שבה מנסים לנבא את הסטייה מקו הרגרסיה עבור כל תצפית:

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

<u>הרציונאל של המבחן הוא:</u>

אחת מן ההנחות הקלאסיות של מודל הרגרסיה היא כי המשתנים הב"ת אינם מתואמים עם טעות האמידה: $\operatorname{cov}(X_t, u_t) = 0$.

המשתנים המצויים בתוך המודל (השניים הראשונים), לא מתואמים עם טעות האמידה ($R^2_{\hat{u}12}=0$).

לגבי שני המשתנים שיש לנו ספק לגביהם:

אם הם יהיו מתואמים עם טעות האמידה ($R^2_{\hat{u}34}\neq 0$) – זו אינדיקציה ש"שכחנו" להוסיפם למודל וכי הם רלוונטיים (דוחים H0).

אם הם אינם מתואמים עם טעות האמידה ($R^2_{\hat{u}34}=0$), המודל מושלם כמו שהוא (מקבלים את H0).

- **הסבר באמצעות מעגלי וואן:**

<u>חישוב הסטטיסטי:</u>

הכפלת ה- R^2 של רגרסיית העזר במספר התצפיות (T).

$$LM_{stat}=R^2\cdot T$$

<u>כלל הכרעה של H0:</u>

אם $LM_{stat}>\chi^2_m$ נדחה את H0.

m- מס' ההגבלות ב-H0.

לדוגמא:

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

1) **בדקו את הטענה כי לפחות אחד מן המשתנים הנוספים רלוונטי למודל בשתי דרכים.**

2) **איזה מן המשתנים הנוספים רלוונטי למודל?**

****<u>שימו לב</u> כי:**

עבור <u>המשתנים הנוספים למודל</u> -כל המדדים (הבטות, ערכי t וה-Pvalu) ברגרסיית העזר שווים לאלו של הרגרסיה הלא מוגבלת.

עבור <u>המשתנים הקיימים במודל</u> -המדדים אינם שווים בין שתי הרגרסיות.

3) **הסבירו את הקשרים בין שלוש המשוואות: (R) (U) ו (עזר) ואת הקשר בין מבחן WALD ומבחן LM.**

הקשר בין משוואות ה-U, ה-R והעזר:

א. U=R+עזר

ב. $ESS_U = ESS_y$

ג. $ESS_R = TSS_y$

ד. $R_y^2 = 1 - \frac{ESS_y}{TSS_y} = 1 - \frac{ESS_U}{ESS_R} = \frac{ESS_R - ESS_U}{ESS_R}$

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

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

פרק 2 - משתני דמי

הנושא של משתני דמי מטפל בהכנסת משתנים ב"ת איכותיים למודל הרגרסיה.

עד כה כל המשתנים הב"ת שהכנסנו למודל היו כמותיים, כלומר קיבלו ערכים מספריים.

למשל, נניח שאנו סבורים שמס' שנות הלימוד של אדם משפיעות על שכרו:

W_t=**השכר (המשתנה התלוי)**

S_t = **שנות לימוד (המשתנה הב"ת)**

משוואת הרגרסיה: $W_t = \alpha + \beta \cdot S_t$

במקרה זה המשתנה המסביר (כמו גם המוסבר) הוא כמותי.

נניח שאנו סבורים שגם משתנה המגדר משפיע על השכר. משתנה זה איננו כמותי כמו שנות לימוד אלא איכותי שכן הוא לא מקבל ערכים מספריים אלא ערכים קטגוריאליים כ"גבר" או "אישה".

נשאלת השאלה כיצד נכניס אותו לתוך משוואת הרגרסיה?

נגדיר משתנה D שיקבל את הערך 0 אם מדובר ב"אישה" ואת הערך 1 אם מדובר ב"גבר".

משתנה כזה נקרא משתנה דמי (dummy variable).

לעומת משתנה רגיל ש"פועל" תמיד, משתנה זה "יפעל" רק אם מדובר בגבר.

ניתן להכניס את משתנה הדמי למודל בשלושה אופנים שונים:

(1) משתנה דמי לחותך- המין משפיע על השכר ההתחלתי בלבד

(2) משתנה דמי לשיפוע- המין משפיע על התוספת לשכר בגין שנות הלימוד

(3) משתנה דמי לכל הפונקציה- המין משפיע גם על החותך וגם על השיפוע

(1)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: $H0: \alpha_1 = 0$

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

? 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 באוכלוסיה?

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.

<u>פונקצית רגרסיה המכילה משתנים איכותיים בלבד</u>

המגדר הוא המשתנה היחיד במשוואה:

$$W_t = \alpha_0 + \alpha_1 D + u_t$$

החותך מייצג כאן את השכר ה<u>ממוצע</u> 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: $H0: \alpha_1 = 0$ **(מבחן זהה למבחן t להבדל בין ממוצעים).**

? 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 <u>הממוצע</u> בין נשים וגברים?

(2) משתנה דמי לשיפוע

המגדר משפיע על התוספת לשכר בגין שנות הלימוד.

$$W_t = \alpha + \beta_0 S_t + \beta_1 D S_t + u_t$$

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: $H0:\beta_1=0$

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

? 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.

(3) משתנה דמי לכל הפונקציה

המין משפיע גם על החותך וגם על השיפוע. הווה אומר, גם על השכר ההתחלתי וגם על התוספת לשכר ההתחלתי בגין שנות הלימוד.

המודל: $W_t = \alpha_0 + \alpha_1 D + \beta_0 S_t + \beta_1 D S_t + u_t$

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)**

<u>בדיקת השערות למשתני הדמי:</u>

באמצעות מבחן WALD יש לבדוק: $H0:\alpha_1=\beta_1=0$

H1: לפחות אחד הפרמטרים שונה מ-0

אם דוחים את השערת האפס, יש לבצע מבחני t עבור כל אחד מהפרמטרים בנפרד:

$HO:\alpha_1=0$ **ו-** $H0:\beta_1=0$

<u>מבחן CHOW</u>

דרך נוספת לבדיקת ההבדל בין הקטגוריות, <u>בלא יצירת משתני דמי</u>:

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 שהשתמשנו בו מקודם):

המודל המוגבל (R) לא לוקח בחשבון את השפעת המגדר ולכן לא יכלול את המדגם המאוחד כי אין צורך בשתי רגרסיות נפרדות:

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

המודל הלא מוגבל (U) כולל את שני חלקי המדגם:

$$ESS_U = ESS_f + ESS_m$$
$$DF_U = DF_f + DF_m$$

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

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

1. **אם דחינו את HO 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.**

? חוקר רצה לבדוק את הטענה שסוג הכביש משפיע על מס' תאונות הדרכים בקטעי כביש בינעירוניים, בהינתן נפח התנועה.

החוקר בדק האם <u>הפונקציה</u> של מס' התאונות בהינתן נפח התנועה, שונה בין כבישים מהירים לבין כבישים שאינם מהירים. לשם כך אמד החוקר את ארבע המשוואות הבאות:

(1) **כבישים מהירים בלבד**

(2) **כבישים לא - מהירים בלבד**

(3) **שני סוגי הכביש (כל המדגם)**

(4)

כאשר : = מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה

=נפח התנועה בקטע כביש t ליום באלפים

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

תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

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

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

3) 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.

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

5) מהי הרגרסיה "תחת   " 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
```

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:** $H0: \alpha_1 = 0$	**מבחן t להפרש השיפועים:** $H0: \beta_1 = 0$	**מבחן WALD להפרש בין הפונקציות (החותכים והשיפועים):** $H0: \alpha_1 = \beta_1 = 0$ ****ניתן לבדוק את ההשערה בדבר הבדל בין הפונקציות גם במבחן CHOW.** **אם דוחים את HO יש לברר את מקור ההבדל באמצעות מבחני t (אפשרי רק ב-WALD):** $H0: \alpha_1 = 0$ $H0: \beta_1 = 0$

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים

כאשר המשתנה האיכותי כלל שני ערכים בלבד (למשל, מגדר: גבר, אישה) הסתפקנו במשתנה דמי אחד.

במקרים רבים המשתנה האיכותי כולל יותר משני ערכים/קטגוריות. במקרה כזה 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**

$\alpha_0 + \alpha_i$ **:החותך בקטגוריה i.**

<u>בדיקת השערות:</u>

השערות:

$$H0: \alpha_1 = \alpha_2 = \alpha_3 = 0$$
$$H1: OTHERWISE$$

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 HO 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:** $H0: \alpha_1 = 0$
2. **האם יש הבדל במחיר ההתחלתי של הירקות בין הקיץ לחורף:** $H0: \alpha_2 = 0$
3. **האם יש הבדל במחיר ההתחלתי של הירקות בין הסתיו לחורף:** $H0: \alpha_3 = 0$

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

1. **מהי השערת האפס לבדיקת הטענה?**
2. **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.

1. **מהי השערת האפס לבדיקת הטענה?**
2. **מהו המבחן הסטטיסטי המתאים? פרטו.**

(2) <u>משתני דמי לשיפוע</u>

<u>הטענה:</u> יש הבדל בין עונות השנה בתוספת למחיר הירקות בגין המחיר לצרכן

<u>המודל:</u> $V_t = \alpha + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t$

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.**

<u>בדיקת השערות:</u>

השערות:

$$H0: \beta_1 = \beta_2 = \beta_3 = 0$$
$$H1: OTHERWISE$$

המבחן הסטטיסטי : מבחן WALD:

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

(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 HO במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין השיפועים על ידי מבחני t.

(3) <u>משתני דמי לכל הפונקציה</u>

<u>הטענה:</u>יש הבדל בין עונות השנה בפונקצית הרגרסיה לניבוי מחיר הירקות באמצעות המחיר לצרכן.

<u>המודל:</u> $V_t = \alpha_0 + \alpha_1 D_{1i} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + \beta_0 P_t + \beta_1 (D_{1i} P_t) + \beta_2 (D_{2i} P_t) + \beta_3 (D_{3i} P_t) + u_t$

<u>בדיקת השערות:</u>

השערות:

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

המבחן הסטטיסטי: מבחן WALD:

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

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

אם דוחים את HO, יש לבדוק במבחני WALD האם ההבדל הוא בין החותכים או בין השיפועים:

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

באם דוחים את H0 יש להמשיך לבדוק באמצעות מבחני t:

$$HO:\beta_j=0\qquad H0:\alpha_j=0$$

משתני דמי עבור שני משתנים איכותיים

נתבונן בדוגמא שבה יש שני משתנים איכותיים המשפיעים על פונקצית 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) <u>הבדל בחותך ללא אינטראקציה</u>

<u>המודל:</u> $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \beta \cdot S_t + u_t$

במודל זה- אין השפעה משולבת של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים לא תלוי בגזע (זהה עבור שחורים ועבור לבנים) ולהיפך-ההבדל בשכר ההתחלתי בין לבנים לשחורים לא תלוי במגדר (זהה עבור נשים וגברים).

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

1. **הבדל בשכר ההתחלתי בין גברים לנשים :** $HO:\alpha_1 = 0$
2. **הבדל בשכר ההתחלתי בין שחורים ללבנים:** $H0:\alpha_2 = 0$

(2) <u>הבדל בחותך עם אינטראקציה</u>

<u>המודל:</u> $W_t = \alpha_0 + \alpha_1 G + \alpha_2 R + \alpha_3 G \cdot R + \beta \cdot S_t + u_t$

במודל זה הטענה היא כי קיימת, בנוסף להשפעה של מגדר וגזע בנפרד על השכר, גם השפעה משולבת (אינטראקציה) של מגדר וגזע על השכר ההתחלתי.

במילים אחרות, ההבדל בשכר ההתחלתי בין גברים ונשים תלוי בגזע (שונה אם מדובר בשחורים או בלבנים) ולהיפך.

במודל זה, לעומת הקודם, נוספת ההשערה לבדיקת השפעת האינטראקציה בין מגדר לגזע על השכר ההתחלתי:

3. $HO:\alpha_3 = 0$

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:

$HO: \gamma_1 - \gamma_3 = \gamma_2$ **או** $HO: \gamma_1 - \gamma_2 = \gamma_3$

התוצאות שיתקבלו כאן יהיו כמובן זהות לחלוטין לתוצאות שהתקבלו בדרך הקודמת:

$$WALD = t^2$$
$$PF = Pt$$

? שאלה מס' 1

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

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

	השכלה נמוכה	השכלה גבוהה	הפרש
לבנים	α_0	$\alpha_0+\alpha_3$	α_3
שחורים	$\alpha_0+\alpha_2$	$\alpha_0+\alpha_1$	$\alpha_1-\alpha_2$
הפרש	α_2	$\alpha_1-\alpha_3$	

(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) הי 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:

$$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + \nu$$ **מהם ה-Z-ים?**

(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').

<u>שאלה מס' 2</u>

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

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

כאשר : S משתנה דמי : 1= עבור נשים, 0=גברים

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

בטרם ניגשים לפיתרון השאלה יש להכין את טבלת העזר הבאה:

	השכלה גבוהה (E=1)	השכלה נמוכה (E=0)	הפרש
נשים (S=1)	$\alpha_0+\alpha_1+\alpha_2+\alpha_3+(\beta_0+\beta_1+\beta_2+\beta_3)EXP_t$	$\alpha_0+\alpha_1+(\beta_0+\beta_1)EXP_t$	חותך: $\alpha_2+\alpha_3$ שיפוע: $\beta_2+\beta_3$
גברים (S=0)	$\alpha_0+\alpha_2+(\beta_0+\beta_2)EXP_t$	$\alpha_0+\beta_0 EXP_t$	חותך: α_2 שיפוע: β_2
הפרש	חותך: $\alpha_1+\alpha_3$ שיפוע: $\beta_1+\beta_3$	חותך: α_1 שיפוע: β_1	

א. רשמו את הפונקציה לחישוב:

1. **תחזית לוג השכר עבור גבר בעל השכלה נמוכה ו- 10 שנות ניסיון.**
2. **תחזית לוג השכר ההתחלתי עבור נשים משכילות.**
3. **לאחר כמה שנות ניסיון ישתווה השכר של נשים משכילות לזה של גברים משכילים?**

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:

1. **אין השפעה של מגדר והשכלה על השכר.**
2. **השפעת ההשכלה אינה תלויה במגדר.**
3. **אין השפעות השכלה אצל גברים.**
4. **אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.**

<u>פרק 3 - הפרה של ההנחות הקלאסיות</u>

ארבעת הנושאים האחרונים שנלמד עוסקים במצב של הפרת אחת ההנחות הקלאסיות הדרושות לאמידת הפרמטרים בשיטת OLS.

הנושא הראשון- <u>הטרוסקדסטיות</u>, עוסק בהפרת הנחה מס׳ 5 המדברת על שונות קבועה ויחידה לאורך קו הרגרסיה: $V(u_t)=\sigma^2$.

הנושא השני- <u>מתאם סידרתי</u>, עוסק בהפרת הנחה מס׳ 6 המדברת על אי תלות בין הטעויות: $\text{cov}(u_t,u_s)=0$

הנושא השלישי-<u>מודלים דינמיים</u>, והרביעי-<u>משוואות סימולטניות</u>, עוסקים בהפרת הנחה מס׳ 4 המדברת על 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.**

פרק 4 - הטרוסקדסטיות

הטרוסקדסטיות הוא מצב שבו מופרת הנחה מס' 5, הנחת ההומוסקדסטיות, הגורסת כי השונות של ה"קריזה" 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 הקריזה של כל תצפית היא בעלת שונות אחרת:

$V(u_t)=\sigma_t^2$

למשל כאשר בודקים את הקשר שבין הכנסה ותצרוכת מגלים כי ברמות הכנסה נמוכות אין גמישות רבה בהוצאות ולכן התצפיות מרוכזות סביב קו הרגרסיה. לעומת זאת ברמות גבוהות של הכנסה התצפיות מפוזרות יותר ויש שונות גבוהה יותר בתצרוכת. השונות, אם כן, איננה אחידה סביב קו הרגרסיה והיא תלויה בתצפית- זהו מצב של הטרוסקדסטיות.

ההשלכות של הטרוסקדסטיות על אומדי OLS

מבין התכונות של אר"פ (ליניאריות, חוסר הטיה, עקיבות ויעילות) היחידה שמופרת בהינתן הטרוסקדסטיות היא: יעילות של אומדי הרבועים הפחותים.

זאת מכיוון שבכדי להוכיח יעילות של האומדים השתמשנו בהנחה מס' 5 של שונות קבועה. לכן הפרתה גוררת פגיעה ביעילות האומד.

מבחנים לזיהוי הטרוסקדסטיות

החשד לקיומה של בעיית הטרוסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות. גרף השאריות גם מאפשר לנו לבחון את הצורה הפונקציונאלית של השונות, כלומר באיזה אופן השונות משתנה בין תצפית לתצפית.

קיימות שתי שיטות לזיהוי הטרוסקדסטיות: מבחן GQ (Goldfeld-Qauandt) ומבחן White.

ההבדל בין שני המבחנים הוא בהחלטה על הצורה הפונקציונאלית של השונות. כלומר על האופן שבו השונות משתנה בין תצפית לתצפית, (על פי דיאגראמת הפיזור).

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) <u>מבחן GQ</u>

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).**

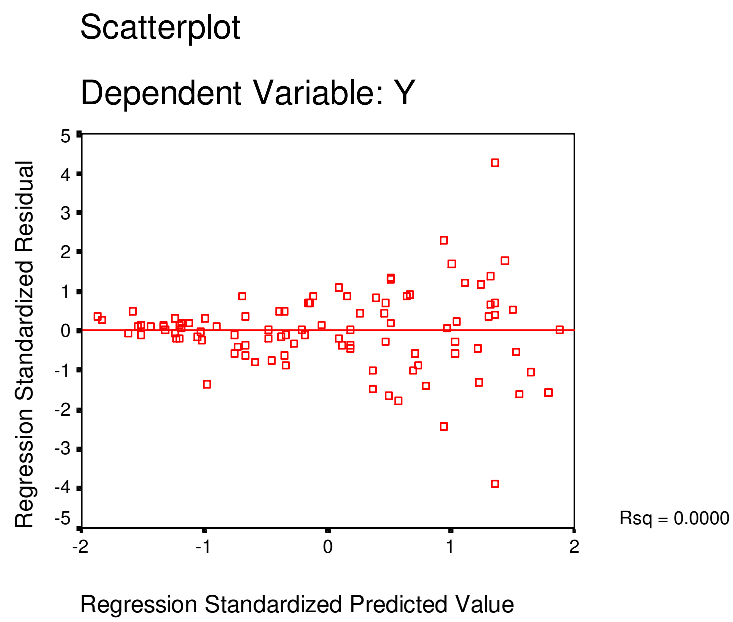
- **סטטיסטי זה מתפלג** $F_{(\alpha; T_1 - K - 1, T_2 - K - 1)}$

- **כלל ההכרעה: אם** $F_{stat} > F_C$ **אז דוחים את H0.**

- **השערות:** $HO: \sigma_1^2 = \sigma_2^2$
 $H1: \sigma_1^2 > \sigma_2^2$

לדוגמא: **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 ככל שעולים ברמת ההכנסה כך השונות בהוצאות הפרט עולה. כלומר השונות איננה אחידה סביב קו הרגרסיה אלא תלויה ברמת ההכנסה- זהו מצב של הטרוסקדסטיות.

בכדי לבצע מבחן GQ :

- **התצפיות של המשתנה הכנסה 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	---	----			

<u>השערות:</u>

$$HO:\sigma_1^2=\sigma_2^2$$
$$H1:\sigma_1^2>\sigma_2^2$$

<u>סטטיסטי המבחן:</u>

$$F_{stat}=\frac{ESS_1/T_1-K-1}{ESS_2/T_2-K-1}=\frac{2{,}934{,}638}{166{,}452.9}=17.62$$

<u>כלל הכרעה:</u>

$F_{stat}=17.62>F(0.05;14,14)=2.48$ **לכן יש סיבה מספקת לדחות את HO ברמת מובהקות של 5%.**

<u>מסקנה:</u>

יש עדות להטרוסקדסטיות בנתונים.

? על מנת לבחון את פונקצית 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 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).

2) <u>מבחן White</u>

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:

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

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 ב SAS -RES).**
- **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:**

$$HO: \alpha_j = \beta_j = \gamma_{jj} = 0$$
$$H1: OTHERWISE$$

נדגים 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: $\hat{u}_t^2 = \alpha_0 + \alpha_1 X_t + \beta_1 X_t^2 + v_t$

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

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

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

השערות:

$$H0: \alpha_1 = \beta_1 = 0$$
$$H1: OTHERWISE$$

סטטיסטי המבחן:

$$LM_{stat} = T_y \cdot R_y^2 = 48 \cdot 0.3907 = 18.75$$

<u>כלל הכרעה:</u>

$LM_{stat} = 18.75 > \chi^2_{2(0.05)} = 5.991$ **לכן יש סיבה מספקת לדחות את H0 ברמת מובהקות של 5%.**

<u>מסקנה</u>

יש עדות לקיום של הטרוסקדסטיות בנתונים.

****<u>הערה חשובה:</u> אם חלק מערכי** $\hat{u}_t^2$ **יצאו שליליים (והרי שונות לא יכולה להיות שלילית) יש להוציא לוג ולאמוד את** $\ln \hat{u}_t^2$ **כך התחזית תצא תמיד חיובית.**

? חוקר מניח כי מכירות של חנות הן פונקציה של שיטחה, דמי שכירות והאפשרות של מכירת עיתונים.

נסמן:

Y_SALES=מכירות חודשיות (ש"ח)

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	1	-------	----	-------	--- --
X1_SQUARE	1	-------	-----	------	--- ---
X1_SQUARE^2	1	-------	-----	-----	--- ---
X1_SQUARE*X2_RENT	1	-------	-----	------	-- ---
X2_RENT	1	-------	--------	-----	-- --
X2_RENT^2	1	--------	--------	-----	-- ---
X2_RENT*PAPERS	1	-------	------	-----	-- ---
PAPERS	1	------	------	-----	- ---

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:___________________________________

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 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 :

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

בצורתה המפורשת המשוואה החדשה נראית כך:

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

קיבלנו מודל חדש שבו :

הקריזה היא : $\frac{u_t}{\sqrt{Z_t}}$

המשתנה המוסבר: $\frac{Y_t}{\sqrt{Z_t}}$

המשתנים המסבירים הם : $\frac{X_t}{\sqrt{Z_t}}$ **ו-** $\frac{1}{\sqrt{Z_t}}$

במודל זה <u>אין חותך.</u>

הפרמטרים β ו- α הם של המשוואה המקורית (שימו לב כי ל- α אין משמעות של חותך).

למשוואה זו אין משמעות כלכלית אך היא מאפשרת לנו לאמוד את הפרמטרים α ו- β בשיטת OLS כך שנקבל אומדים $\hat{\alpha}$ ו- $\hat{\beta}$ יעילים.

ניתן להוכיח כי הטרנספורמציה הליניארית של השונות הפכה אותה לקבועה:

$$V(\frac{u_t}{\sqrt{Z_t}}) = \frac{1}{Z_t} V(u_t) = \frac{1}{Z_t} \cdot \sigma^2 \cdot Z_t = \sigma^2$$

הערה: אם ביצענו את מבחן WHITE והוצאנו log לנתונים אז $Z_t = e^{\ln \hat{u}_t^2}$ **כך שיש לשקלל את הרגרסיה ב:** $W_t = \frac{1}{\sqrt{e^{\ln \hat{u}_t^2}}}$.

? שאלה מס' 1

נתון המודל : **(1)** $Y_t = \alpha + \beta \cdot X_t + U_t$

ונתון כי : $VAR(U_t) = \sigma^2 / Z_t^2$ **(** 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:

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

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

שאלה מס' 2

עני על שאלה מס' 1, כאשר נתון כי $VAR(U_t) = \sigma^2 \cdot X_t^2$.

<u>שאלה מס' 3</u>

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. פיתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה הבאה:

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

כאשר: $W_t =$ ___________

ד. ם 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**

פרק 5 - מתאם סידרתי

מתאם סידרתי עוסק במצב שבו מופרת ההנחה הקלאסית מס' 6 -אי תלות בין הטעויות: $\operatorname{cov}(u_t, u_s) = 0$

זהו מצב שבו קיימת תלות סטטיסטית בין הטעויות במודל: $\operatorname{cov}(u_t, u_s) \neq 0$.

תלות כזו בין הטעויות קיימת בדרך כלל כאשר הנתונים הנאספים הם נתוני סדרות עיתיות ולא נתוני חתך בהם עסקנו עד כה.

נתוני חתך- מתייחסים לפרטים שונים בתקופת זמן נתונה.

נתוני סדרות עיתיות- מתייחסים לאותו הפרט לאורך זמנים שונים.

בנתוני החתך סביר שמאחר ומדובר בפרטים שונים -הטעויות בניבוי שלהם תהיינה בלתי תלויות. לעומת זאת בנתוני סדרות עיתיות, מאחר ומדובר באותו הפרט הנמדד בזמנים שונים סביר דווקא שהטעויות בניבוי שלו תהיינה תלויות אחת בשנייה.

אם אנחנו מדברים, למשל, על פונקציות תצרוכת:

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

אם מדובר בנתוני חתך- החלטות התצרוכת של פרט t לא אמורות להשפיע על החלטותיו של פרט s והנחה 6 תתקיים:

$$\operatorname{cov}(u_t, u_s) = 0$$

אולם אם מדובר על נתוני סדרה עיתית: החלטותיו של פרט מסוים היום סביר שיושפעו מהחלטות שעשה בעבר או שישפיעו על החלטות שיעשה בעתיד:

$$\operatorname{cov}(u_t, u_{t-s}) \neq 0$$

כאשר : u_t **- "קריזה" בזמן t .**

u_{t-s} **-"קריזה" בזמן t-s .**

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.

****<u>שימו לב</u> 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

הגדרנו מתאם סדרתי כהפרה של הנחה מס' 6:

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

$$\text{cov}(u_t, u_{t-s}) \neq 0$$

נשאלת השאלה כיצד נראית הפרה זו?

<u>מתאם סדרתי מסדר ראשון:</u>

ההנחה היא כי יש מתאם בין קריזות במרחק אחד, כלומר u_t תלוי ישירות רק ב- u_{t-1} :

$$\text{cov}(u_t, u_{t-1}) \neq 0$$

את המתאם בין ה"קריזות" 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)**

(2) $1 < \rho < -1$ **(כי אם חורג מ-1 ה"קריזה" 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:**

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

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

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

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

מלבד α **ו-** β **נרצה לאמוד גם את** ρ**.**

<u>**מתאם סדרתי מסדר שני:**</u>

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

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

<u>**מתאם סדרתי מסדר P:**</u>

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

u_t **מושפע מתקופות שונות בעבר.**

תכונות המתאם הסדרתי

ניתן להוכיח כי מכיוון שכל טעות בזמן מסוים מתואמת עם הטעות הסמוכה לה בזמן:

$$r_{(u_t,u_{t-s})} = \rho^s$$

מכיוון שכך 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:

$$E(u_t) = 0$$

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

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

? **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.

מבחנים לזיהוי מתאם סדרתי

קיימים שני מבחנים סטטיסטיים לזיהוי קיומו של מתאם סדרתי:

מבחן DW (דרבין ווטסון) ומבחן LM.

1) <u>**מבחן DW (דרבין ווטסון) לקיום מתאם סדרתי מסדר ראשון:**</u>

- **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___ _ $\rho > 0$ ______]_**?**__[__________ $\rho = 0$ ______]__**?**__[______ $\rho < 0$ _**4**

d_L d_U $4 - d_U$ $4 - d_L$

- **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.

<u>**השערות:**</u>

$$HO: \rho = 0$$
$$H1: \rho > 0, \rho < 0$$

<u>**חישוב הסטטיסטי**</u>

$$DW_{stat} \cong 2 \cdot (1 - \hat{\rho})$$

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

לדוגמא: 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	-7.468	0.0000

Durbin-Watson D 0.150

האם קיים מתאם סדרתי?

נתבונן בטבלת DW כאשר נתון כי K=1 ו-T=153 :

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

לפי הטבלה (עבור T=100) : $\begin{aligned} d_L &= 1.65 \\ d_U &= 1.69 \end{aligned}$

נציב בטווח :

$$[\_\_\_\_ \rho > 0 \_\_\_\_\_]\_\_\_\_[\_\_\_\_\_\_\_\_\_\_ \rho = 0 \_\_\_\_\_]\_\_\_\_[\_\_\_\_\_ \rho < 0 \_]$$

0   1.65   1.69   2.31   2.35   4

DW=0.150

מסקנה :

יש עדות לקיום מתאם סדרתי חיובי מסדר ראשון.

? 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**

T=100 וידוע כי:

$$\begin{aligned} u_t &= 0.9 u_{t-1} \\ d_L &= 1.57 \\ d_U &= 1.65 \end{aligned}$$

האם קיים מתאם סדרתי ברמת מובהקות של 5% ?

<u>למבחן DW יש שתי בעיות עיקריות:</u>

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

<u>בנוסף לכך על מספר תנאים להתקיים כדי שאפשר יהיה להשתמש במבחן DW :</u>

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

2) <u>מבחן LM</u>

לעומת מבחן DW מבחן LM מתאים גם לבחון קיומו של מתאם סדרתי מסדרים גבוהים יותר מסדר ראשון.

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

ניתן לרשום את 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:** $LM_{stat} = T_y \cdot R_y^2$
- **נדחה את H0 כאשר:** $LM_{stat} > \chi_m^2$ **כאשר m = סדר המתאם הסדרתי.**

אם נדחה את H0 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:

$$HO:\rho_1=\rho_2=...=\rho_s=0$$
$$H1: \text{אחרת}$$

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

לדוגמא: 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	-00096
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 באוכלוסיה?

פיתרון 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:

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

כך "נטרלנו" את המתאם הסדרתי:

$\Leftarrow u_t = \rho \cdot u_{t-1} + \varepsilon_t$ $\varepsilon_t = u_t - \rho \cdot u_{t-1}$ **נזכור כי** 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.

? 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

<u>**אמידת ρ בשיטת קוקרן אורקוט**</u>

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

התהליך מתבצע באופן הבא:

אמידת המשוואה המקורית בה אנו מניחים כי קיים מתאם סדרתי.

חישוב $\hat{u}_t$ ו- $\hat{u}_{t-1}$ וקבלת האומד $\hat{\rho}$ על ידי אמידת המשוואה : $\hat{u}_t=\rho\hat{u}_{t-1}+\varepsilon_t$

הבעיה היא כי האומד איננו יעיל כיוון שהאומדים לפרמטרים במשוואה המקורית אינם יעילים (בשל קיומו של מתאם סדרתי). לכן נמשיך ונבצע את הפעולות הבאות:

הצבת האומד הזה ברגרסיית ההפרשים וחילוץ $\hat{\alpha}$ ו- $\hat{\beta}$ חדשים. שוב נחשב $\hat{u}_t$ ו- $\hat{u}_{t-1}$ תוך שימוש באומדים החדשים ונקבל אומד $\hat{\rho}$ חדש.

נחזור על התהליך הזה מספר פעמים נוספות עד שנגיע להתכנסות (עד שהאומד $\hat{\rho}$ ישתווה לזה שקיבלנו בתחילה).

אין צורך לבצע תהליך זה ידנית משום שרוב תוכנות המחשב מבצעות הליך זה באופן מיידי ומדווחות על התוצאות הסופיות.

התהליך הממוחשב נקרא אוטו רגרסיה (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 אוטו רגרסיה מסדר שני) וכו'

אם משתנה AR מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

לדוגמא: 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			
AR(1)	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?

❓ שאלה מס׳ 1

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

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-2103.53	1128.38	-1.864	0.069
Y	1	0.71944	0.0510	14.086	0.000
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.

<u>שאלה מס׳ 2</u>

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?

פרק 6 - סיכום בעיית המתאם הסדרתי וההטרוסקדסטיות

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

<u>פרק 7 - מודלים דינמיים</u>

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

ניתן לדבר על שלוש סוגים של השפעות בהקשר של המודל הדינמי (מיכפלים):

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):

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

<u>מציאת המכפילים במודל הדינמי:</u>

לעיתים אנו מתבקשים למצוא את המכפילים של מודל דינמי מסוים (לטווח קצר, בינוני וארוך) ואף מתבקשים לבדוק השערות ביחס אליהם.

אדגים את מציאת המכפילים על המודל הדינמי הבא: $Y_t = \alpha + \beta_0 X_t + \beta_1 Y_{t-1} + u_t$

1) מכפל טווח קצר (מיידי):

X היום משפיע על Y היום דרך β_0 **:** $\frac{\partial Y_t}{\partial X_t} = \beta_0$

2) מכפל ביניים מסדר j (מכפיל דינמי):

בכדי למצוא את המכפיל הדינמי "נגלגל" את המודל 2 תקופות אחורה:

$$Y_t = \alpha + \beta_0 X_t + \beta_1 Y_{t-1} + u_t$$
$$Y_t = \alpha + \beta_0 X_t + \beta_1 (\alpha + \beta_0 X_{t-1} + \beta_1 Y_{t-2} + u_{t-1}) + u_t$$
$$Y_t = \alpha + \beta_0 X_t + \beta_1 \alpha + \beta_1\, \beta_0 X_{t-1} + \beta_1^2 Y_{t-2} + \beta_1 u_{t-1} + u_t$$
$$Y_t = \alpha + \beta_0 X_t + \beta_1 \alpha + \beta_1 \beta_0 X_{t-1} + \beta_1^2 (\alpha + \beta_0 X_{t-2} + \beta_1 Y_{t-3} + u_{t-2}) + u_t$$
$$Y_t = \alpha + \beta_0 X_t + \beta_1 \alpha + \beta_1 \beta_0 X_{t-1} + \beta_1^2 \alpha + \beta_1^2 \beta_0 X_{t-2} + \beta_1^3 Y_{t-3} + \beta_1^2 u_{t-2} + u_t$$

ההשפעה של X_t **על** Y_t **היא:** β_0

ההשפעה של X_{t-1} **על** Y_t **היא:** $\beta_0 \beta_1$

ההשפעה של X_{t-2} **על** Y_t **היא:** $\beta_0 \beta_1^2$

מכאן ניתן להסיק שבאופן כללי ההשפעה של X_{t-j} **על** Y_t **היא:**

$$\frac{\partial Y_t}{\partial X_{t-j}} = \beta_0 \beta_1^j$$

חישוב מכפל הביניים בדרך קצרה:

β_1	β_0	
Y_{t-1}	X_t	Y_t
Y_{t-2}	X_{t-1}	Y_{t-1}
Y_{t-3}	X_{t-2}	Y_{t-2}

ניתן לראות כי X_{t-2} **משפיע על** Y_t **דרך** β_0 **,** β_1 **,** β_1 **כלומר** $\beta_0\beta_1^2$.

סביר שככל ש-X רחוק יותר בזמן כך השפעתו קטנה יותר ולכן $\beta_1 < 1$.

3) מכפל טווח ארוך:

מכפל טווח ארוך מהווה את סך ההשפעות:

$$\frac{\partial Y^*}{\partial X^*} = \sum_{j=0}^{P} \beta_0\beta_1^j = \beta_0 + \beta_0\beta_1 + \beta_0\beta_1^2 + \ldots + \beta_0\beta_1^p =$$
$$= \beta_0(1 + \beta_1 + \beta_1^2 + \ldots + \beta_1^p)* =$$
$$= \beta_0 \cdot \frac{1}{1-\beta_1}$$

*** סכום של טור הנדסי יורד.**

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

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 ב"קריזות", כלומר במתאם הסדרתי.

רגרסיית ההפרשים, המהווה פיתרון למתאם הסדרתי, היא למעשה מודל דינמי:

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

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

$$Y_t = \alpha_0 + \beta_0 X_t + \beta_1 X_{t-1} + \beta_2 Y_{t-1} + \varepsilon_t$$

<u>לסיכום:</u> 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} ישמש 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 אל"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 <u>אינם</u> משתנים מקריים).

$$\hat{\beta} = \frac{S_{XY}}{S_{XX}}$$

$$\hat{\beta} = \frac{S_{Y_t Y_{t-1}}}{S_{Y_{t-1} Y_{t-1}}}$$

$$E(\hat{\beta}) = \beta + E\left(\frac{\sum (Y_{t-1} - \bar{Y})u_t}{\sum (Y_{t-1} - \bar{Y})^2}\right) \neq \beta$$

מאחר והמשתנה המקרי נשאר בתוך התוחלת, לא ניתן להשתמש בהנחה מספר 1:
$E(u_t)=0$ **. לכן הביטוי השני איננו מתאפס והאומד הוא אומד מוטה.**

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 $COV(Y_{t-1},u_t)\neq 0 \;\Leftarrow$ **והאומד איננו עקיב.**

<u>לסיכום: ההשלכות על אר"פ:</u>

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 מסכמות-מתאם סדרתי ומודלים דינמיים

<u>שאלה מס' 1</u>

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:

1. לא קיים מתאם סדרתי

2. קיים מתאם סדרתי והוא:_______________

3. לא ניתן לקבוע אם המתאם הסדרתי מובהק.

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

מוטים	**נכון/ לא נכון/ לא ניתן לדעת**
ליניאריים	**נכון/ לא נכון/ לא ניתן לדעת**
יעילים	**נכון/ לא נכון/ לא ניתן לדעת**
עקיבים	**נכון/ לא נכון/ לא ניתן לדעת**

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

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

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

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

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 ?

<u>שאלה מס׳ 2</u>

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

$$M_t = \alpha + \beta \cdot P_t + U_t \qquad \textbf{(1)}$$

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 **:**

1. **לא ניתן לחשב את הסטטיסטי DW מהנתונים הקיימים**
2. **קיים מתאם סדרתי שלילי**
3. **קיים מתאם סדרתי חיובי**
4. **לא קיים מתאם סדרתי**
5. **לא ניתן לקבוע אם המתאם הסדרתי מובהק.**

ב. סמנו את התשובה הנכונה בהכרח:

1. **האומדים ליניאריים חסרי הטיה, עקיבים אך לא יעילים**
2. **האומדים ליניאריים חסרי הטיה, עקיבים ויעילים**
3. **האומדים מוטים אך עקיבים**
4. **האומדים חסרי הטיה, אך לא עקיבים**
5. **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:

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

כאשר M_{t-1} **הינה 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 ג' 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-i

ט. בדקו את הטענה כי המכפיל הדינמי לשינוי P בתקופה t-1 ($\frac{\partial M_t}{\partial P_{t-1}}$) 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?

שאלה 2 - 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

שאלה 2 - 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

$m1 = m_{t-1}$

שאלה ב - פרק 3

pendent Variable: res Residual

$RES = \hat{\varepsilon}_t$ = של אמידת משוואה (2)

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

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

שאלה 3 - פרק 1 - משוואה (1)

pendent Variable: lny

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	71.06922	14.21384	40.35	<.0001
Error	397	139.83633	0.35223		
Corrected Total	402	210.90554			

Root MSE	0.59349	R-Square	0.3370
Dependent Mean	7.14247	Adj R-Sq	0.3286
Coeff Var	8.30934		

Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	6.44240	0.10765	59.84	<.0001
d1		1	-0.28522	0.13600	-2.10	0.0366
d2		1	-0.77210	0.13276	-5.82	<.0001
d3		1	0.43817	0.06804	6.44	<.0001
exp	exp	1	0.07116	0.00860	8.28	<.0001
exp2		1	-0.00144	0.00015662	-9.17	<.0001

<u>**פרק 8 - משוואות סימולטניות**</u>

עוסקות בהפרה של הנחה 4 המדברת על כך שהמשתנים הב"ת במשוואת הרגרסיה אינם משתנים מקריים ולכן לא מתואמים עם הטעויות: $\operatorname{cov}(x,u)=0$.

במילים אחרות ה-Xים במשוואה נחשבו <u>משתנים אקסוגניים</u>- משפיעים על Y אך לא מושפעים ממנו בחזרה.

זהו סוג של משתנים כלכליים הנשלט על ידי קובעי המדיניות או גורמים חיצוניים אחרים.

דוגמא למשתנה אקזוגני: שער חליפין קבוע (הנקבע באופן אקזוגני על ידי הבנק המרכזי) המשפיע על כמות הכסף במשק אך לא מושפע ממנה בחזרה.

לעומת זאת, קיים סוג נוסף של משתנים כלכליים ב"ת הקרויים <u>משתנים אנדוגניים</u>- משפיעים על Y אך גם מושפעים ממנו בחזרה.

מאחר ומשתנים אלו הם גם מסבירים וגם מוסברים, הם נחשבים כמשתנים מקריים, המתואמים עם הטעויות במודל: $\operatorname{cov}(x,u)\neq 0$.

לדוגמא: התצרוכת הפרטית מושפעת בדר"כ מרמת ההכנסה, וזו מושפעת מן הביקושים השונים לתוצר, ולכן גם ההכנסה וגם התצרוכת נקבעים ביחד באופן אנדוגני במערכת.

משוואות המבנה(משוואות סימולטניות)

מערכת משוואות הכוללות משתנים מסבירים אנדוגניים ואקסוגניים הנמצאות בשיווי משקל.

בדר"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.

לדוגמא:

(1)

(2)

X ו-Y הם משתנים אנדוגניים הנקבעים סימולטנית במערכת (למשל, תצרוכת והכנסה) ואילו ה-Zים הם משתנים אקזוגניים (כמו למשל, שער הדולר ומגדר):

$\operatorname{cov}(x,u) \neq 0$, $\operatorname{cov}(y,u) \neq 0$ **ואילו-** $\operatorname{cov}(z_{jt},u) = 0$

יתכן והמערכת תכלול משוואה נוספת המתארת מצב של שיווי משקל בין שתי המשוואות.

לדוגמא:

נניח שאנו רוצים לאמוד את הביקוש לתירס ואת ההיצע של התירס במדינה. הכמות המבוקשת והכמות המוצעת מושפעות מן המחיר ומשפיעות עליו.

מערכת משוואות הביקוש וההיצע של התירס הנה מהצורה:

משוואת הביקוש: $Qd_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_t$

משוואת ההיצע: $Qs_t = \beta_0 + \beta_1 P_t + \beta_2 Z_t + v_t$

משוואת שיווי המשקל: $Qd_t = Qs_t = Q$

כאשר:

Qd_t= **כמות ביקוש של תירס**

Qs_t= **כמות היצע של תירס**

P_t= **מחיר התירס**

X_t= **הכנסה** Z_t= **מי גשם**

שלושת משוואות אלו מכונות המשוואות המבניות שכן הן מתארות את מבנה שוק התירס.

תחילה יש להחליט מהם המשתנים האנדוגניים והאקסוגניים במערכת:

- **משתנה תלוי הוא תמיד אנדוגני (Q_t).**

- **השאלה היא איזה מהמשתנים המסבירים הוא אנדוגני: כזכור הכמות המבוקשת והמוצעת מושפעות מן המחיר אך גם משפיעות עליו. לפיכך מחיר התירס הוא אנדוגני (P_t).**

- **המשתנים האנדוגניים נקבעים סימולטנית במערכת ומתואמים עם הטעויות במערכת:** $\text{cov}(p_t,u_t)\neq 0; \text{cov}(Q_t,u_t)\neq 0$

- **המשתנים האקזוגניים במערכת: מי גשם (Z_t) שהוא כמובן משתנה חיצוני שלא מושפע מן הביקוש וההיצע וכמו כן, בהנחה שהמשק לא מייצר רק תירס אז גם הכנסה (X_t) היא משתנה אקזוגני למערכת:** $\text{cov}(z_t,u_t)=0; \text{cov}(x_t,u_t)=0$

המטרה היא , כמו תמיד, לאמוד בצורה יעילה את הפרמטרים (אלפות ובטות) ולבצע בדיקת השערות.

הבעיה היא כי ברגע שקיים משתנה מסביר אנדוגני במערכת, אמידת כל אחת מהמשוואות המבניות בנפרד בשיטת OLS תניב אומדים לא ליניאריים, מוטים, לא יעילים ולא עקיבים.

השלכות על אל"פ

הנחה 4 שימשה אותנו להוכחת ליניאריות, חוסר הטיה ועקיבות.

לכן הפרתה של הנחה זו 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).

נלמד 3 שיטות אמידה אחרות: ILS, TSLS ו-IV. אך ראשית יש לבטא את המשוואות המבניות בצורה אחרת הנקראת: "הצורה המצומצמת".

הצורה המצומצמת של מודל עם משוואות סימולטניות

משוואות הצורה המצומצמת הן פיתרון עבור המשתנים האנדוגניים במערכת: הגדרת המשתנים האנדוגניים כפונקציה של המשתנים האקזוגניים במערכת בלבד.

מספר המשוואות המצומצמות הוא כמספר המשתנים האנדוגניים במערכת (במקרה זה שניים).

התהליך האלגברי הוא כדלקמן:

- **נשווה בין שתי המשוואות בשיווי משקל:** $Qd_t = Qs_t$

$$\alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_t = \beta_0 + \beta_1 P_t + \beta_2 Z_t + v_t$$

- **נחלץ את P:**

$$P_t = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1} - \frac{\alpha_2}{\alpha_1 - \beta_1} X_t + \frac{\beta_2}{\alpha_1 - \beta_1} Z_t + \frac{v_t + u_t}{\alpha_1 - \beta_1}$$

- **הצורה המצומצמת:**

$$P_t = \lambda_0 + \lambda_1 X_t + \lambda_2 Z_t + \varepsilon_{1t}$$

כאשר:

$$\lambda_0 = \frac{\beta_0 - \alpha_0}{\alpha_1 - \beta_1}$$

$$\lambda_1 = -\frac{\alpha_2}{\alpha_1 - \beta_1}$$

$$\lambda_2 = \frac{\beta_2}{\alpha_1 - \beta_1}$$

$$\varepsilon_{1t} = \frac{v - u}{\alpha_1 - \beta_1}$$

- **כדי לבטא את Q במונחים של משתנים אקזוגניים בלבד, נצטרך להציב את P באחת המשוואות (לא משנה איזה כי הן נמצאות בשיווי משקל):**

$$Q_t = \alpha_0 + \alpha_1(\lambda_0 + \lambda_1 X_t + \lambda_2 Z_t + \varepsilon_t) + \alpha_2 X_t + u_t$$

$$Q_t = \alpha_0 + \alpha_1 \lambda_0 + (\alpha_1 \lambda_1 + \alpha_2) X_t + \alpha_1 \lambda_2 Z_t + \alpha_1 \varepsilon_t + u_t$$

- **הצורה המצומצמת עבור Q:**

$$Q_t = \mu_0 + \mu_1 X_t + \mu_2 Z_t + \varepsilon_{2t}$$

כאשר:

$$\begin{aligned}\mu_0 &= \alpha_0 + \alpha_1\lambda_0 \\ \mu_1 &= \alpha_1\lambda_1 + \alpha_2 \\ \mu_2 &= \alpha_1\lambda_2 \\ \varepsilon_2 &= \alpha_1\varepsilon_1 + u_t\end{aligned}$$

<u>משוואות הצורה המצומצמת שהתקבלו:</u>

$$P_t = \lambda_0 + \lambda_1 X_t + \lambda_2 Z_t + \varepsilon_{1t}$$

$$Q_t = \mu_0 + \mu_1 X_t + \mu_2 Z_t + \varepsilon_{2t}$$

<u>תכונות המשוואות מהצורה המצומצמת</u>

- **מס' המשוואות הוא כמספר המשתנים האנדוגניים במערכת (P ו-Q).**
- **המשתנה המוסבר הוא אנדוגני וכל המסבירים אקסוגניים.**
- **המשתנים המסבירים הם זהים בכל המשוואות (X ו-Z).**
- **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.**

<u>אמידת הפרמטרים של משוואות המבנה באמצעות משוואות הצורה המצומצמת</u>

משוואות הצורה המצומצמת מאפשרות, כאמור, לאמוד את הפרמטרים (ה- λ ות וה- μ ים) בשיטת OLS אבל אנחנו מעוניינים למעשה לאמוד את הפרמטרים של המשוואות המקוריות- משוואות המבנה (ה- α ות וה- β ות).

מתוך הפרמטרים של הצורה המצומצמת נחלץ את הפרמטרים של משוואות המבנה: מתוך ה- λ ות נחלץ את ה- α ות ומתוך ה- μ ים נחלץ את ה- β ות.

בתהליך החילוץ של הפרמטרים המבניים ייתכנו 3 מצבים:

1) <u>**אין זיהוי:**</u> **לא ניתן לחלץ את הפרמטרים המבניים (ה- α ות וה- β ות) מתוך הפרמטרים של הצורה המצומצמת (ה- λ ות וה- μ ים). זה קורה כשיש פחות משוואות מנעלמים (כלומר פחות λ ות ו- μ ים מ- α ות ו- β ות) אז יש אינסוף פתרונות.**

2) <u>**זיהוי מדויק:**</u> **יש רק דרך אחת לחלץ את הפרמטרים המבניים מהפרמטרים של הצורה המצומצמת, זה קורה כשיש בדיוק אותו מספר משוואות ונעלמים (פיתרון יחיד).**

3) <u>**זיהוי יתר:**</u> **יש יותר מדרך אחת לחלץ את הפרמטרים המבניים מתוך הפרמטרים של הצורה המצומצמת. זה קורה כשיש יותר משוואות מנעלמים (יותר מפתרון אחד).**

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**

בדוגמא שלנו:

משוואת הביקוש: $Qd_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_t$

משוואת היצע: $Qs_t = \beta_0 + \beta_1 P_t + \beta_2 Z_t + v_t$

<u>**עבור משוואת הביקוש:**</u>

g-1: 2-1=1

K-k: 3-2=1

מכיוון ש: $g-1=K-k$

הזיהוי של ה- α **ות מתוך משוואות הצורה המצומצמת יהיה <u>מדויק</u>.**

<u>עבור משוואת ההיצע:</u>

g-1: 2-1=1

K-k: 3-2=1

מכיוון ש: $g-1=K-k$

הזיהוי של ה- β **ות מתוך משוואות הצורה המצומצמת יהיה <u>מדויק</u>.**

? חוקר רצה לאמוד את פונקצית הביקוש ואת פונקצית 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 בש"ח בתקופה *t*.**

INCOMEt **- הכנסת הצרכנים באלפי ש"ח בתקופה *t*.**

Lt **- מחיר שעת עבודה בש"ח 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, וקבעו האם ניתן לזהות את מקדמי המשוואות, באילו שיטות ומהן תכונות האומדנים שיתקבלו?

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

ג. עיינו בדפי הפלט המתאימים 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)

שיטות לפיתרון משוואות סימולטניות

קיימות 3 שיטות לפיתרון משוואות סימולטניות: ILS, TSLS ו-IV.

1) <u>שיטת ריבועים פחותים עקיפה (ILS)</u>

א. יש להציג את מערכת משוואות המבנה בצורתה המצומצמת.

ב. יש לאמוד בשיטת OLS את הפרמטרים של המשוואות בצורה המצומצמת (ה- λ ות וה- μ ים).

ג. יש לחלץ מן הפרמטרים של המערכת המצומצמת את הפרמטרים של הצורה המבנית (ה- α ות וה- β ות).

משום שתהליך החילוץ איננו ליניארי האומדים המבניים המתקבלים הם מוטים אך עקיבים.

כאשר <u>הזיהוי מדויק</u> האומדים יהיו גם אסימפטוטית יעילים (במדגמים גדולים).

כאשר <u>הזיהוי הוא יתר</u>: האומדים לא יהיו יעילים.

בדוגמא שלנו:

<u>מערכת המשוואות המבניות:</u>

משוואת הביקוש: $Qd_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_t$

משוואת ההיצע: $Qs_t = \beta_0 + \beta_1 P_t + \beta_2 Z_t + v_t$

<u>מערכת המשוואות בצורתה המצומצמת:</u>

$$P_t = \lambda_0 + \lambda_1 X_t + \lambda_2 Z_t + \varepsilon_{1t}$$

$$Q_t = \mu_0 + \mu_1 X_t + \mu_2 Z_t + \varepsilon_{2t}$$

<u>נאמוד את הפרמטרים של הצורה המצומצמת</u> בשיטת OLS ונקבל את האומדים:

$$\hat{\lambda}_0, \hat{\lambda}_1, \hat{\lambda}_2, \hat{\mu}_0, \hat{\mu}_1, \hat{\mu}_2$$

אומדים אלו יקיימו את כל תכונות אר"פ.

התמודדנו קודם לכן עם <u>שאלת הזיהוי</u> וקבענו כי עבור שתי המשוואות <u>הזיהוי הוא מדויק</u>, כלומר קיים פיתרון אחד עבור כל אומד.

כעת נוכל <u>לחלץ את האומדים המבניים</u> מתוך האומדים של הצורה המצומצמת.

אלו האומדים של הצורה המצומצמת:

$$\hat{\mu}_0 = \hat{\alpha}_0 + \hat{\alpha}_1\hat{\lambda}_0 \qquad \hat{\lambda}_0 = \frac{\hat{\beta}_0 - \hat{\alpha}_0}{\hat{\alpha}_1 - \hat{\beta}_1}$$

$$\hat{\mu}_1 = \hat{\alpha}_1\hat{\lambda}_1 + \hat{\alpha}_2 \qquad \hat{\lambda}_1 = -\frac{\hat{\alpha}_2}{\hat{\alpha}_1 - \hat{\beta}_1}$$

$$\hat{\mu}_2 = \hat{\alpha}_1\hat{\lambda}_2 \qquad \hat{\lambda}_2 = \frac{\hat{\beta}_2}{\hat{\alpha}_1 - \hat{\beta}_1}$$

אנחנו צריכים לבטא את האומדים המבניים: $\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ במונחי האומדים המצומצמים:

נתחיל מ- $\hat{\alpha}_1$: $\hat{\alpha}_1 = \frac{\hat{\mu}_2}{\hat{\lambda}_2}$

$\hat{\alpha}_2$:

$$\hat{\alpha}_2 = \hat{\mu}_1 - \hat{\alpha}_1\hat{\lambda}_1$$

$$\hat{\alpha}_2 = \hat{\mu}_1 - \frac{\hat{\mu}_2}{\hat{\lambda}_2}\hat{\lambda}_1$$

$\hat{\alpha}_0$:

$$\hat{\alpha}_0 = \hat{\mu}_0 - \hat{\alpha}_1\hat{\lambda}_0$$

$$\hat{\alpha}_0 = \hat{\mu}_0 - \frac{\hat{\mu}_2}{\hat{\lambda}_2}\hat{\lambda}_0$$

נעבור ל- λ ות:

נתחיל מ- $\hat{\beta}_1$:

$$\hat{\beta}_1 = \frac{\hat{\alpha}_2 + \hat{\lambda}_1\hat{\alpha}_1}{\hat{\lambda}_1}$$

$$\hat{\beta}_1 = \frac{\hat{\mu}_1 - \frac{\hat{\mu}_2}{\hat{\lambda}_2}\hat{\lambda}_1 + \hat{\lambda}_1\frac{\hat{\mu}_2}{\hat{\lambda}_2}}{\hat{\lambda}_1}$$

$$\hat{\beta}_1 = \frac{\hat{\mu}_1}{\hat{\lambda}_1}$$

$\hat{\beta}_2$:

$$\hat{\beta}_2 = \hat{\lambda}_2(\hat{\alpha}_1 - \hat{\beta}_1)$$
$$\hat{\beta}_2 = \hat{\lambda}_2(\frac{\hat{\mu}_2}{\hat{\lambda}_2} - \frac{\hat{\mu}_1}{\hat{\lambda}_1})$$
$$\hat{\beta}_2 = \hat{\mu}_2 - \frac{\hat{\lambda}_2\hat{\mu}_1}{\hat{\lambda}_1}$$

$\hat{\beta}_0$:

$$\hat{\beta}_0 = \hat{\lambda}_0(\hat{\alpha}_1 + \hat{\beta}_1) + \hat{\alpha}_0$$
$$\hat{\beta}_0 = \hat{\lambda}_0(\frac{\hat{\mu}_2}{\hat{\lambda}_2} + \frac{\hat{\mu}_1}{\hat{\lambda}_1}) + \hat{\mu}_0 - \frac{\hat{\mu}_2}{\hat{\lambda}_2}\hat{\lambda}_0$$
$$\hat{\beta}_0 = \hat{\lambda}_0\frac{\hat{\mu}_1}{\hat{\lambda}_1} + \hat{\mu}_0$$

לסיכום:

$$\hat{\alpha}_0 = \hat{\mu}_0 + \frac{\hat{\mu}_2}{\hat{\lambda}_2}\hat{\lambda}_0 \qquad \hat{\beta}_0 = \hat{\lambda}_0\frac{\hat{\mu}_1}{\hat{\lambda}_1} + \hat{\mu}_0$$
$$\hat{\alpha}_1 = \frac{\hat{\mu}_2}{\hat{\lambda}_2} \qquad \hat{\beta}_1 = \frac{\hat{\mu}_1}{\hat{\lambda}_1}$$
$$\hat{\alpha}_2 = \hat{\mu}_1 - \frac{\hat{\mu}_2}{\hat{\lambda}_2}\hat{\lambda}_1 \qquad \hat{\beta}_2 = \hat{\mu}_2 - \frac{\hat{\lambda}_2\hat{\mu}_1}{\hat{\lambda}_1}$$

אומדים אלו יהיו אמנם מוטים אך עקיבים ויעילים אסימפטוטית כך שניתן יהיה לבצע בדיקת השערות במדגמים גדולים.

כעת משנתונות לנו תוצאות האמידה של המשואות המצומצמות, נוכל לחשב את האומדים של המשוואות המבניות.

למשל:

תוצאות האמידה של מערכת המשוואות בצורתה המצומצמת:

$$P_t = 1.2 + 3.4X_t + 4Z_t$$

$$Q_t = 3 + 1.3X_t + 2Z_t$$

חשב את האומדים לשיפוע מחיר התירס הן במשוואת הביקוש והן במשוואת ההיצע.

<u>תשובה:</u>

נחשב את α_1 ואת β_1 :

$$\hat{\alpha}_1 = \frac{\hat{\mu}_2}{\hat{\lambda}_2} = \frac{2}{4} = 0.5$$
$$\hat{\beta}_1 = \frac{\hat{\mu}_1}{\hat{\lambda}_1} = \frac{1.3}{3.4} = 0.38$$

? 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 הפיתרון 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

2) <u>שיטת ריבועים פחותים בשני שלבים (2SL2)</u>

לשיטה זו שני שלבים:

א. אמידת משוואות הצורה המצומצמת בשיטת OLS ושימוש בתוצאות האמידה כדי לחשב את המשתנים האנדוגניים (המסבירים).

ב. הצבת המשתנים האנדוגניים שהתקבלו במשוואות המבנה ואמדתן ב-OLS.

אם משוואות המבנה <u>מזוהות בדיוק או ביתר</u>- האומדים שיתקבלו יהיו אמנם מוטים אבל עקיבים ויעילים אסימפטוטית.

אם הזיהוי מדויק-האומדים שיתקבלו יהיו זהים לאומדים שהתקבלו בשיטת הריבועים הפחותים העקיפה.

<u>כאשר אין זיהוי:</u> אין אקסוגניים ולכן אין משתנים מסבירים בצורה המצומצמת או שכל האקזוגניים בצורה המצומצמת כבר קיימים במשוואה המקורית ולכן החלפת x ב- $\hat{x}$ תיצור בעיה של מולטיקולינאריות מלאה.

בדוגמא שלנו:

<u>מערכת המשוואות המבניות:</u>

משוואת הביקוש: $Qd_t = \alpha_0 + \alpha_1 P_t + \alpha_2 X_t + u_t$

משוואת ההיצע: $Qs_t = \beta_0 + \beta_1 P_t + \beta_2 Z_t + v_t$

<u>מערכת המשוואות בצורתה המצומצמת:</u>

$$P_t = \lambda_0 + \lambda_1 X_t + \lambda_2 Z_t + \varepsilon_{1t}$$

$$Q_t = \mu_0 + \mu_1 X_t + \mu_2 Z_t + \varepsilon_{2t}$$

א. <u>אמידת $\hat{p}_t$ במשוואה המצומצמת ב-OLS:</u>

$$\hat{P}_t = \hat{\lambda}_0 + \hat{\lambda}_1 X_t + \hat{\lambda}_2 Z_t$$

חישוב $\hat{p}_t$

ב. <u>הצבת $\hat{p}_t$ במשוואות המבנה ואמידתן ב-OLS :</u>

$$Qd_t = \alpha_0 + \alpha_1 \hat{P}_t + \alpha_2 X_t + u_t$$

$$Qs_t = \beta_0 + \beta_1 \hat{P}_t + \beta_2 Z_t + v_t$$

? 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**

1. **מה ניתן יהיה לומר על האומדים שהתקבלו בשיטה זו?**
2. **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

? לפניך המודל הסימולטני הבא:

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

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

$Q_t^D = Q_t^S$

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 ?

3) <u>שיטת משתני עזר (IV)</u>

משתנה עזר הוא משתנה שיחליף את המשתנה המסביר האנדוגני במשוואת המבנה ויעזור לאמוד את הקשר בינו לבין התלוי.

<u>משתנה העזר צריך להיות</u>:

1. משתנה אקזוגני או פונקציה ליניארית של משתנים אקזוגניים:
$\text{cov}(Z,u) = 0$

2. מתואם עם המשתנה האנדוגני אותו הוא מחליף: $\text{cov}(Z,X) \neq 0$.

ככל שהמתאם גבוה יותר, האומד שיתקבל באמצעותו יהיה טוב יותר.

לדוגמא:

נתון המודל: $Y_t = \alpha + \beta X_t + u_t$ **כאשר:** $\text{cov}(X_t,u_t) \neq 0$.

<u>הבעיה:</u> אומדני OLS שיתקבלו יהיו מוטים, לא עקיבים ולא יעילים.

<u>הפיתרון בשיטת IV:</u> אמידת ההשפעה של X על Y עם משתנה אחר Z שמתואם עם X אך לא עם u:

$$\operatorname{cov}(Z, X) \neq 0$$
$$\operatorname{cov}(Z, u) = 0$$

<u>תכונות האומדים שיתקבלו באמצעות משתני העזר:</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.

<u>משתנה העזר הטוב ביותר</u> (והיחיד שיניב אומדים יעילים):

אומדן לאנדוגני שהתקבל מאמידת משוואת הצורה המצומצמת בשלב הראשון של 2SLS.

<u>משתנה לא יוכל לשמש כמשתנה עזר:</u>

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.

<u>משתני עזר שונים יכולים להניב את אותם האומדים לפרמטרים:</u>

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.

לדוגמא:

נתונה משוואת המבנה הבאה:

$$Y_{1t} = \beta_0 + \beta_1 Y_{2t} + \beta_2 Z_{1t} + \beta_3 Z_{2t} + \ldots + \beta_k Z_{(k-1)t} + u_t$$

המשתנים האנדוגניים: Y_1, Y_2

המשתנים האקזוגניים: $Z_1 ... Z_{k-1}$

בשיטת IV נחפש משתנה עזר שיחליף את המשתנה המסביר האנדוגני Y_2

משתנה עזר תקף: Z_k

שכן הוא מקיים את התנאים הבאים:

1) **לא משפיע באופן ישיר על** Y_1 **(אינו נמצא במשוואת המבנה).**

2) $\text{cov}(Z_k, u_t) = 0$ **(משתנה אקזוגני).**

3) $\text{cov}(Z_k, Y_2) \neq 0$ **(מתואם עם המשתנה האנדוגני אותו הוא מחליף).**

ככל שהקשר שלו עם האנדוגני טוב יותר כך משתנה העזר יניב אומדים יעילים יותר לפרמטרים.

__הערה חשובה:__ שיטת האמידה בשני שלבים- 2SLS היא מקרה פרטי של שיטת משתנה העזר:

האומדן לאנדוגני שמתקבל בשלב הראשון מהווה את משתנה העזר היחיד שיניב אומדים יעילים (מבין אפשרויות אחרות של משתני עזר שיניבו אומדים עקיבים בלבד).

בדוגמא שלנו:

ניצור משתנה עזר $\hat{Y}_{2t}$ **שיחליף את** Y_{2t} **על ידי אמידת המשוואה המצומצמת בשלב הראשון של 2SLS**

משוואת הצורה המצומצמת:

$$Y_{2t} = \pi_0 + \pi_1 Z_{1t} + ... + \pi_{k-1} Z_{k-1} + v$$

__שלב ראשון 2SLS :__ נריץ רגרסיה של המשוואה המצומצמת בשיטת OLS כדי לאמוד את Y_{2t} **:**

$$\hat{Y}_{2t} = \hat{\pi}_0 + \hat{\pi}_1 Z_{1t} + ... + \hat{\pi}_{k-1} Z_{k-1}$$

$\hat{Y}_{2t}$ מהווה משתנה עזר העומד בכל התנאים הדרושים:

1) אינו נמצא במשוואת המבנה ($\hat{Y}_{2t} \neq Y_{2t}$).

2) הוא פונקציה של משתנים אקזוגניים בלבד ולכן לא מתואם עם הטעויות במשוואת המבנה ($\operatorname{cov}(\hat{Y}_{2t}, u) = 0$), כלומר, הוא מהווה משתנה אקזוגני במשוואת המבנה בניגוד למשתנה אותו הוא מחליף- Y_{2t}.

3) מקיים קשר עם Y_{2t} ($\operatorname{cov}(\hat{Y}_{2t}, Y_{2t}) \neq 0$). למעשה, זהו האומד שהקשר שלו עם Y_{2t} יהיה החזק ביותר מבין משתני העזר האפשריים ולכן היחיד שיניב אומדים יעילים לפרמטרים.

<u>בשלב השני של 2SLS:</u> נריץ רגרסיה בשיטת OLS של משוואת המבנה תוך שימוש במשתנה העזר $\hat{Y}_{2t}$ במקום Y_{2t}:

$$Y_{1t} = \beta_0 + \beta_1 \hat{Y}_{2t} + \beta_2 Z_{1t} + \beta_3 Z_{2t} + \ldots + \beta_k Z_{(k-1)t} + u_t$$

ההנחות הקלאסיות מתקיימות בשלב השני מכיוון ש- $\operatorname{cov}(\hat{Y}_{2t}, u) = 0$.

$\hat{\beta}_1$ מהרגרסיה הזו הוא האומד הטוב ביותר בשיטת IV של β_1.

? נתונות המשוואות הבאות:

(1)

(2)

נתון כי: משתנים אנדוגניים ו- משתנים אקסוגניים.

חוו דעתכם על כל אחת מהטענות הבאות, והסבירו:

א. ניתן להשתמש ב- כמשתנה עזר לאמידת משוואה מס' (1).

ב. ניתן להשתמש ב- כמשתנה עזר לאמידת משוואה מס' (2).

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

ד. שימוש ב- 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'.**

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 ($\operatorname{cov}(Y_{t-1},u_t)=0$) אז Y_{t-1} אקסוגני.

מבחנים סטטיסטיים:

1) 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.**

לדוגמא:

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 - $Cov(Y_i,\varepsilon_i)\neq 0$.

הסבירו כיצד תשתמשו ב – *Hausman Test* כדי לבחון את ההשערה: $H_0:Cov(Y_i,\varepsilon_i)=0$?

<u>תשובה:</u>

- **נאמוד בשיטת הריבועים הפחותים את המשוואה המצומצמת: $y_i=\gamma_1+\gamma_2 s_i+v_i$**
- **נחזה את השאריות: $\hat{v}_i$.**

- **נאמוד את משוואת המבנה המקורית בתוספת הטעות כמשתנה מסביר נוסף:**

$D_i = \beta_1 + \beta_2 y_i + \beta_3 s_i + \delta\hat{v}_i + u_i$ **ונבצע מבחן להשערה :** $H0 : \delta = 0$

אם דוחים את H0 נוכל להסיק כי המשתנה אנדוגני ולהפך.

2) 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.**

לדוגמא:

נתונה מערכת המשוואות הסימולטניות הבאות –

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

כאשר X_1, X_2, X_3 **הינם משתנים אקסוגניים.**

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?**

<u>תשובה:</u>

נריץ את המשוואה המצומצמת הראשונה: $Y_{1i} = \pi_{11}X_{1i} + \pi_{12}X_{2i} + \pi_{13}X_{3i} + \tilde{u}_i$.

נבצע מבחן F להשערה: $H0: \pi_{11} = \pi_{12} = 0$.

נאמר ש: X_{1i} **ו -** X_{2i} **הם משתני עזר חלשים אם** $F_{stat} < 10$.

תרגילים מסכמים

? 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: $\text{cov}(Z_j, u_t) = 0$ **עבור** $j = 1,...,5$ **(כלומר ה-Z ים אקסוגניים).**

א. אמידת כל אחת מהמשוואות תניב אומדים:

1. **מוטים** **נכון/לא נכון/ לא ניתן לדעת**
2. **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:

1. **תוך שימוש בשיטת ILS ניתן לאמוד את משוואה (1) באופן עקיב וחד ערכי:** **נכון/לא נכון/ לא ניתן לדעת**
2. **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:

1. Z_5
2. $\frac{Z_1 + Z_5}{2}$
3. $2Z_1 + 3Z_2 + Z_3$
4. $Z_3 + Z_4$
5. $3Z_3 + 4Z_4$
6. $3Z_3 + 3Z_4$
7. 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):

1. **1 ו-2**
2. **4 ו-6**
3. **5 ו-6**

5-14 .4

ח. האם משתנה עזר 1 (Z_5) יניב אומדים יעילים ?

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

? 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 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.

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.

? נתונה מערכת המשוואות הסימולטניות הבאה –

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

כאשר X_1, X_2, X_3 הינם משתנים אקסוגניים.

א. חלצו את מערכת המשוואות המצומצמת ה (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)

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

ב. הראו שבהינתן אומדים עקיבים ל - $\pi_{11}, \ldots, \pi_{23}$, 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?

? 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 - $Cov(Y_i, \varepsilon_i) \neq 0$.

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 $H_0 : Cov(Y_i, \varepsilon_i) = 0$?

פרק 9 - סיכום תכונות אר"פ

אם המודל שאותו אנו אומדים מנסח נכון את הקשר המתמטי בין המשתנים וכולל את כל המשתנים הרלוונטיים להסבר התופעה ורק אותם ולא קיים קשר חזק או מלא בין המשתנים הבלתי תלויים במודל וכל ההנחות הקלאסיות מתקיימות, אזי האומדים שאותם נקבל הם א.ח.ה, הם BLUE (יעילים) והם עקיבים.

הבעיה במודל	כולל את כל המשתנים הרלוונטיים	לא כולל משתנים שאינם רלוונטיים	אין קשר חזק או מלא בין הב"ת	כל ההנחות הקלאסיות מתקיימות	פגיעה בתכונות אר"פ
אין בעיה	√	√	√	√	**אין פגיעה**
השמטת משתנים רלוונטיים	×	√	√	√	**פגיעה בכל התכונות- *מוטים, לא יעילים ולא עקיבים**
הוספת משתנים לא רלוונטיים	√	×	√	√	**אין פגיעה**
מולטיקוליניאריות חלקית	√	√	×	√	**אין פגיעה**
מולטיקוליניאריות מלאה	√	√	×	√	**אר"פ אינם מוגדרים**
הטרוסקדסטיות	√	√	√	$V(u_t)=\sigma_t^2$ ×	**יעילות**
מתאם סדרתי	√	√	√	$\mathrm{cov}(u_t,u_s)\neq 0$ ×	**יעילות**
מודלים דינמיים	√	√	√	$\mathrm{cov}(x,u)\neq 0$ ×	**מוטים** **היעילות והעקיבות תלויים בקיום מתאם סדרתי**
משוואות סימולטניות	√	√	√	$\mathrm{cov}(x,u)\neq 0$ ×	**אומדים** מוטים אך עקיבים.** **היעילות תלויה בשיטה.**

***אלא אם אין מתאם בין המשתנה הב"ת שהושמט לאלו המצויים במודל.**

****בהנחה שהמשוואות מזוהות.**