מבוא אקונומטריקה

פרק 13 - הטרוסקדסטיות

תוכן העניינים

הטרוסקדסטיות:

רקע:

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

במצב של הטרוסקדסטיות שונות הטעויות של כל תצפית היא שונה: $V(u_t)=\sigma_t^2$.

ההשלכות של הטרוסקדסטיות על אומדי OLS:

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

מבחנים לזיהוי הטרוסקדסטיות:

החשד לקיומה של בעיית הטרוסקדסטיות בנתונים צריך להתעורר כאשר אנו בוחנים את גרף השאריות – באיזה אופן השונות של הטעויות משתנה בין תצפית לתצפית.

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

מבחן GQ מניח כי במקום שונות אחת אחידה של הטעויות לכל התצפיות, קיימות שתי שונויות שונות בלבד. ואילו מבחן White מניח כי לכל תצפית ותצפית שונות שונה של טעויות.

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

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

- סטטיסטי זה מתפלג: $F_{(\alpha;T_1-K-1,T_2-K-1)}$.
- כלל ההכרעה: אם $F_{stat} > F_C$ אז דוחים את H_0.
- ההשערות: $\begin{matrix} H_0: \sigma_1^2 = \sigma_2^2 \\ H_1: \sigma_1^2 > \sigma_2^2 \end{matrix}$.

2. מבחן White:

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

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

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

האומד ל- σ_t^2 הוא $\hat{u}_t^2$.

המבחן הוא מבחן LM:

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

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

פיתרון בעיית ההטרוסקדסטיות – ריבועים פחותים משוקללים (WLS):

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

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

המשוואה המקורית: $Y_t = \alpha + \beta X_t + u_t$

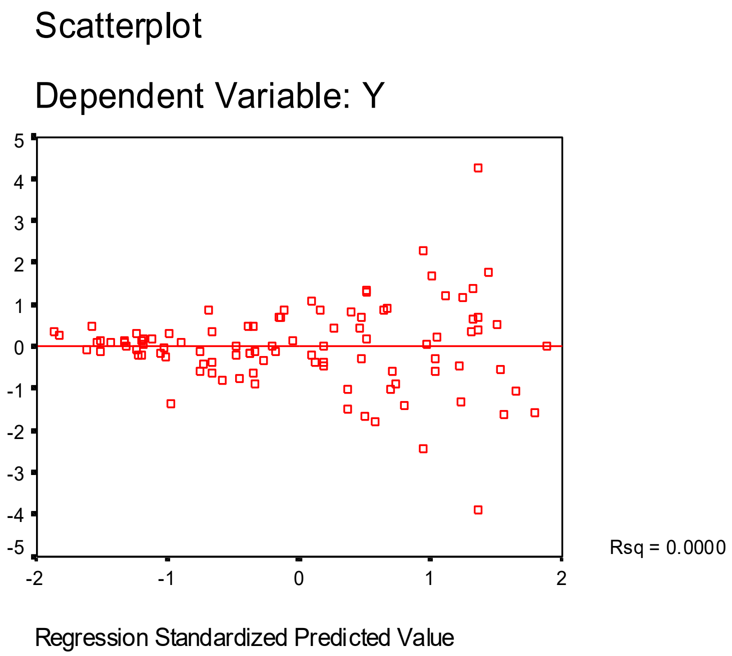
$Y_t W_t = \alpha \cdot W_t + \beta (X_t W_t) + u_t W_t$.

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

שאלות:

מבחן GQ:

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



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

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

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

משוואה (1)

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

Number of Observations Read          16
Number of Observations Used          16

Analysis of Variance

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

משוואה (2)

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

Number of Observations Read          16
Number of Observations Used          16

Analysis of Variance

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

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

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

Q - תפוקה שנתית באלפי שקלים.

L - מספר עובדים.

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

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

ברגרסיה הכוללת את הערכים הנמוכים יחסית של תשומת העבודה הוא

קיבל: $R^2 = 0.403$, $ESS = 279.3$.
ברגרסיה הכוללת את הערכים הגבוהים יחסית של תשומת העבודה הוא
קיבל: $R^2 = 0.238$, $ESS = 493.8$.
האם יש עדות לקיום הטרוסקדסטיות בנתונים? (בצעו את המבחן המתאים:
רשמו השערות, חשבו סטטיסטי מבחן, רשמו כלל הכרעה והגיעו למסקנה).

מבחן WHITE:

3) על אותו הקשר שבין הכנסה לתצרוכת: $Y_t = \alpha + \beta X_t + u_t$.
שאנו חושדים על פי גרף השאריות כי קיים בו מצב של הטרוסקדסטיות.
בכדי לבצע את מבחן WHITE:

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

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
```

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

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

נסמן:

Y_SALES - מכירות חודשיות (₪).

$X1\_SQUARES$ - שטח החנות (מ"ר).

$X2\_RENT$ - דמי שכירות ($).

$PAPERS$ - משתנה איכותי המקבל 1 אם החנות מוכרת גם עיתונים ו-0 אם לא.

החוקר חשד כי קיימת בעיה של הטרוסקדסטיות בנתונים.

החוקר ביצע מבחן לזיהוי הטרוסקדסטיות שתוצאותיו נתונות להלן:

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

                        Analysis of Variance

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

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

הרגרסיה המופיעה בפלט לעיל נועדה לבדיקת: ________________________.

על ידי מבחן: ____________________ .

המשתנה התלוי הינו:________________.

המשתנים הב"ת:___________________.

ההשערות הינן:______________________ .

גודל הסטטיסטי למבחן הינו (רשמו תוצאה מספרית):_____________.

המסקנה המתקבלת היא:_____________________________.

שיטת WLS:

5) נתון המודל:
1. $Y_t = \alpha + \beta \cdot X_t + U_t$.

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

א. מהי הבעיה שנוצרת באמידת משוואה (1)?
ב. מהן תכונות אומדי הריבועים הפחותים של משוואה (1)?

כדי לפתור את הבעיה שנוצרה, נאמדה המשוואה הבאה:
2. $Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + U_t \cdot W_t$.

ג. מהו W_t שבעזרתו ניתן לאמוד את α ו- β בצורה יעילה?
ד. מהו האומד היעיל של σ^2?
ה. האם ניתן להשוות בין המודלים על בסיס R^2? אם לא, האם ניתן להחליט בכל זאת איזה מודל טוב יותר?
ו. חוו דעתכם על הטענות הבאות, ונמקו:
i. אם נתון כי: $Z_t = a + b \cdot \bar{X}$, התשובות לסעיפים א' ו-ב' נשארות ללא שינוי.
ii. המשוואה הנורמאלית: $\sum \hat{\varepsilon}_t = 0$ (כאשר: $\varepsilon_t = U_t \cdot W_t$) היא אחת המשוואות הנורמאליות לאמידת משוואה (2).

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

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

א. במשוואה מס' 1 יש בעיה של: ______________.
ב. אמידת משוואה (1) תניב אומדים בלתי מוטים ועקיבים: נכון/ לא נכון/ לא ניתן לדעת
ג. פתרון הבעיה הקיימת במשוואה (1) ייתכן על ידי אמידת המשוואה הבאה: $Y_t \cdot W_t = \alpha \cdot W_t + \beta \cdot (X_t \cdot W_t) + \omega_t$.
כאשר: $W_t =$ ______.
ד. אם נתון כי: $V(u_t) = \sigma_t^2 = \begin{cases} 3\sigma^2 \Leftrightarrow t \geq 50 \\ \sigma^2 \Leftrightarrow t \leq 50 \end{cases}$
האם ישתנו תשובותיכם לסעיפים א' ו-ב': כן/לא/לא ניתן לדעת

תשובות סופיות:

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

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

3) יש עדות לכך, השערות: $H_0: \alpha_1 = \beta_1 = 0$, $H_1: OTHERWISE$, חישוב סטטיסטי: 18.75,
כלל הכרעה: 5.991.

4) קיום הטרוסקדסטיות בנתונים.
$(LM)WHITE$.
RES^2.
$X1\_SQUARE$, $X1\_SQUARE^2$, $X1\_SQUARE * X2\_RENT$,
$X2\_RENT$, $X2\_RENT^2$, $X2\_RENT * PAPERS$, $PAPERS$.
$LM_{stat} = 1.73$.
אין עדות לכך.

5) א. הטרוסקדסטיות. ב. ראו סרטון. ג. $W_t = \frac{1}{\sqrt{\frac{1}{Z_t^2}}} = Z_t$.
ד. $\sigma^2 = \frac{ESS}{T-K}$. ה. לא, המודל השני. ו. i לא נכון. ii. לא נכון.

6) א. הטרוסקדסטיות. ב. ראו סרטון. ג. $W_t = \frac{1}{X_t}$.
ד. $S^2 = \frac{ESS}{T-k-1}$. ה. ראו סרטון. ו. ראו סרטון.

7) א. הטרוסקדסטיות. ב. נכון.
ג. $t \ge 50$ $W_t = \frac{1}{\sqrt{X_t}}$, $t \le 50$ $W_t = \frac{1}{\sqrt{X_t^2}}$. ד. לא.