מבוא לאקונומטריקה

תוכן העניינים

מבוא לאקונומטריקה

פרק 1 - מבוא לקורס

תוכן העניינים

מבוא לקורס:

רקע:

הגדרות וסימונים:

משתנה אמפירי – תוצאותיו ידועות מראש (למשל: רמת הכנסה, גיל, מסי שנות לימוד במדגם מסוים).
משתנה מקרי – תוצאותיו לא ידועות מראש (כגון תוצאה בהטלת קובייה או בהטלת מטבע). באקונומטריקה נעסוק בעיקר במשתנים מקריים.

שני סוגי המשתנים יסומנו באות לועזית עם אינדקס (למשל: Y_t או X_t).
קבוע – מקבל ערך אחד בלבד (מסומן באות לועזית ללא אינדקס – למשל a או b).
לכל משתנה מקרי X_t יש תוחלת המייצגת את מרכז ההתפלגות (μ_X או $E(X)$).
השונות – מייצגת את מידת הפיזור של ההתפלגות (σ_X^2 או $V(X)$).
סטית התקן – היא השורש של השונות (σ_X).
שונות משותפת (covariance) – מדד להתפלגות המשותפת של שני משתנים מקריים ומייצגת את הכיוון של הקשר ביניהם ($\mathrm{Cov}(X,Y)$):
$\mathrm{Cov}(X,Y)=0 \Leftrightarrow X,Y$ בלתי מתואמים.
$\mathrm{Cov}(X,Y)>0 \Leftrightarrow$ מתאם חיובי בין המשתנים.
$\mathrm{Cov}(X,Y)<0 \Leftrightarrow$ מתאם שלילי בין המשתנים.
X,Y בלתי תלויים $\Leftarrow X,Y$ בלתי מתואמים.
מקדם המתאם של פירסון – מדד לכיוון ולעוצמת הקשר הליניארי בין שני משתנים: $\eta_{xy}=\dfrac{\mathrm{cov}(x,y)}{\sigma_x\cdot\sigma_y}$.
$-1\le\eta\le1$
$\eta=1$ מתאם ליניארי חיובי מלא בין שני המשתנים.
$\eta=-1$ מתאם ליניארי שלילי מלא בין שני המשתנים.
$\eta=0$ לא קיים מתאם ליניארי בין שני המשתנים.

אמידה:

פרמטר – ערך המשתנה הנחקר המתאר את כל האוכלוסייה.
סטטיסטי/אומד – ערך המשתנה הנחקר המתאר את המדגם.

מדגם	אוכלוסייה
$\bar{X}=\frac{\sum_{i=1}^{n}X_i}{n}$	$E(X)=\mu$
$S_X^2=\frac{S_{XX}}{n}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n}$	$V(X)=\sigma^2=E(X-E(X))^2$
$\frac{S_{XY}}{n}=\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n}$	$\operatorname{cov}(X,Y)=E(X-E(X))(Y-E(Y))$
$r_{XY}=\frac{S_{XY}}{\sqrt{S_{XX}}\sqrt{S_{YY}}}$	$\eta_{XY}=\frac{\operatorname{cov}(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}$

נוסחאות וחוקים בסטטיסטיקה:

יהיו X ו- Y משתנים מקריים, ו- a, b קבועים:

<u>חוקי הסיגמה</u>:

1. $\sum_{t=1}^{T}X_t=X_1+X_2+X_3+\ldots+X_T$.
2. סכום של קבוע: $\sum_{t=1}^{T}a=Ta$.
3. סכום של קבוע כפול משתנה = לקבוע כפול הסכום: $\sum_{t=1}^{T}aX_i=a\sum_{t=1}^{T}X_t$.
4. סכום של סכום/הפרש = לסכום/הפרש הסכומים: $\sum_{t=1}^{T}(X_t\pm Y)_t=\sum_{t=1}^{T}X_t\pm\sum_{t=1}^{T}Y_t$.
5. יש לשים לב כי: $\sum_{t=1}^{T}X_t^2\neq\left(\sum_{t=1}^{T}X_t\right)^2$

$\sum_{t=1}^{T}X_tY_t\neq\sum_{t=1}^{T}X_t\sum_{t=1}^{T}Y_t$.

הגדרות ופיתוחים:

1. סכום הסטיות מהממוצע = 0: $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)=0$.
2. סכום הסטיות הריבועיות מהממוצע (מונה השונות):
$$S_{XX}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=\sum_{t=1}^{T}X_t^2-T\bar{X}^2=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)X_t.$$
3. מונה של השונות המשותפת:
$$S_{XY}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}X_tY_t-T\bar{X}\bar{Y}=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)Y_t=\sum_{t=1}^{T}\left(Y_t-\bar{Y}\right)X_t.$$

חוקי התוחלת:

1. תוחלת של קבוע = קבוע: $E(a)=a$.
2. תוחלת של סכום/הפרש = לסכום/הפרש התוחלות:
$$E(X\pm Y)=E(X)\pm E(Y)$$
$$E\left(\sum(X_i)\right)=\sum E(X_i).$$
3. תוחלת של כפל/חילוק ≠ לכפל/חילוק התוחלות:
$$E(X\cdot Y)\neq E(X)\cdot E(Y)$$
$$E\left(\frac{X}{Y}\right)\neq\frac{E(X)}{E(Y)}.$$
$$E\left(X^2\right)\neq\left[E(X)\right]^2$$
4. השפעת טרנספורמציה ליניארית על התוחלת:
$$E\left(a/\frac{1}{a}X\pm b\right)=a/\frac{1}{a}\cdot E(X)\pm b.$$

חוקי השונות:

1. עבור X ו- Y בלתי תלויים/בלתי מתואמים מתקיים:
שונות של סכום/הפרש = סכום השונויות:
$$V(X\pm Y)=V(X)+V(Y)$$
$$V\sum(X_i)=\sum V(X_i).$$
2. עבור X ו- Y תלויים/מתואמים מתקיים:
שונות של סכום/הפרש ≠ סכום השונויות:
$$V(X\pm Y)=V(X)+V(Y)\pm 2\cdot Cov(X,Y).$$

3. שונות של קבוע = 0: $V(a)=0$, $V(a\pm x)=V(X)$.

4. השפעת טרנספורמציה ליניארית על השונות: $V(aX+b)=a^2V(X)$.

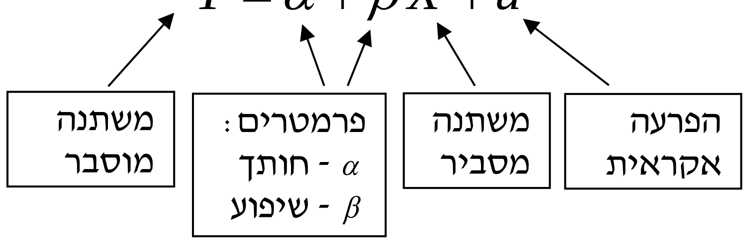
- חוקי התוחלת והשונות מתייחסים למשתנים אמפיריים כאל קבועים (יוצאים מחוץ לתוחלת או לשונות).
  חוקי הסכימה מתייחסים למשתנים אמפיריים כמשתנים הנשארים בתוך הסיגמא (רק הקבועים ייצאו מחוץ לסיגמא).

<u>חוקי השונות המשותפת</u>:

1. שונות משותפת בין משתנה לקבוע = 0: $\text{cov}(X,a)=0$.
2. שונות משותפת של משתנים המוכפלים בקבוע: $\text{cov}(aX,bY)=ab\cdot\text{cov}(X,Y)$.
3. שונות משותפת של משתנה עם עצמו = שונות המשתנה: $\text{cov}(X,X)=V(X)$, $\text{cov}(Y,Y)=V(Y)$.

המודל האקונומטרי:

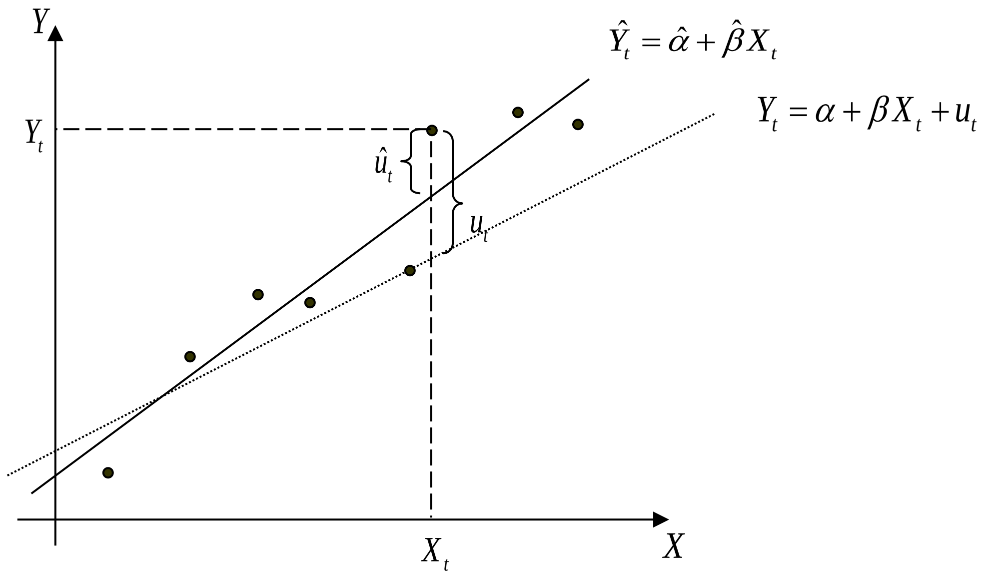
$$Y=\alpha+\beta X+u$$



1. במודל: $Y_t=\alpha+\beta X_t+u_t$, α ו- β הם מספרים קבועים אך לא ידועים. אנו יכולים להעריך אותם ולקבל אומדים (תהליך קבלת האומדנים נקרא אמידה).
2. $\hat{\alpha}$ הוא האומד ל- α ו- $\hat{\beta}$ הוא האומד ל- β.
3. אומדי ריבועים פחותים (אר"פ) הם האומדים שחושבו בשיטת הריבועים הפחותים. מסומנים בד"כ ע"י 'כובע' - $\hat{\beta}$.
   אומדים אחרים מסומנים בד"כ ע"י 'תלתל' - $\tilde{\beta}$.

4. בעוד α ו-β הם קבועים, $\hat{\alpha}$ ו-$\hat{\beta}$ הם משתנים מקריים כיוון שבכל מדגם מתקבלים $\hat{\alpha}$ ו-$\hat{\beta}$ אחרים.
5. את α ו-β ו-u_t לא ניתן לדעת (אלא רק לאמוד מנתוני המדגם) – הקו האמיתי באוכ' לא ידוע.
6. אפשר לדעת את $\hat{u}_t$, שהיא הסטיה מקו הרגרסיה במדגם:

- עבור X_t, הערך הצפוי של המשתנה המוסבר $\left(\hat{Y}_t\right)$ המתקבל לפי הרגרסיה הוא: $\hat{Y}_t = \hat{\alpha} + \hat{\beta}X_t$.
- הסטיה של התצפית (Y_t) מהערך הצפוי לפי הרגרסיה $\left(\hat{Y}_t\right)$ היא: $\hat{u}_t = Y_t - \hat{Y}_t$.



— קו הרגרסיה הנאמד (במדגם)
..... קו הרגרסיה האמיתי באוכלוסייה)
• תצפית בודדת

שאלות:

1) הבא נוכיח את הזהויות הבאות:

א. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=\sum_{t=1}^{T}X_t^2-T\bar{X}^2$.

ב. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)^2=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)X_t$.

ג. $\sum\left(X_t-\bar{X}\right)=0$.

ד. $\sum\frac{\left(X_t-\bar{X}\right)^2}{\sum\left(X_t-\bar{X}\right)X_t}=1$.

ה. $\sum\left(\hat{\alpha}+\hat{\beta}x_i\right)\left(x_i-\bar{x}\right)=\hat{\beta}\left(x_i-\bar{x}\right)^2$.

ו. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}X_tY_t-T\bar{X}\bar{Y}$.

ז. $\sum_{t=1}^{T}\left(X_t-\bar{X}\right)\left(Y_t-\bar{Y}\right)=\sum_{t=1}^{T}\left(X_t-\bar{X}\right)Y_t$.

2) בטא באמצעות: $\operatorname{var}(y),\operatorname{var}(x),\operatorname{cov}(x,y)$ והקבועים a ו- b את הביטויים הבאים:

א. $Var(ax)$.

ב. $Var(x+y)$.

ג. $Var(ax+b)$.

ד. $Cov(x,ay)$.

ה. $Cov(x+a,y+b)$.

ו. מקדם המתאם בין x ל- y.

תשובות סופיות:

1) הוכחה.

2) א. $a^2\operatorname{var}(x)$. ב. $\operatorname{var}(x)+\operatorname{var}(y)+2\operatorname{cov}(x,y)$. ג. $a^2\operatorname{var}(x)$. ד. $a\operatorname{cov}(x,y)$.

ה. $\operatorname{cov}(x,y)$. ו. $r=\frac{\operatorname{cov}(x,y)}{\sqrt{\operatorname{var}(x)}\cdot\sqrt{\operatorname{var}(y)}}$.

מבוא לאקונומטריקה

פרק 2 - אומדי הריבועים הפחותים

תוכן העניינים

אומדי הריבועים הפחותים:

רקע:

Ordinary Least Squares (OLS) – שיטת האמידה של α ושל β לקבלת אומדים $\hat{\alpha}$ ו- $\hat{\beta}$ שיביאו למינימום את סכום ריבועי טעויות האמידה:

$$\min_{\hat{\alpha}\hat{\beta}} \sum \hat{u}_t^2 = \min_{\hat{\alpha}\hat{\beta}} \sum \left(y_t - \hat{y}_t\right)^2 = \min_{\hat{\alpha}\hat{\beta}} \sum \left[y_t - \left(\hat{\alpha} + \hat{\beta} x_t\right)\right]^2 = ?.$$

מתוך גזירת הפונקציה הזו מתקבלים האומדים $\hat{\alpha}$ ו- $\hat{\beta}$.

	מודל עם חותך ושיפוע $Y_t = \alpha + \beta X_t + u_t$	מודל ללא חותך $Y_t = \beta X_t + u_t$	מודל רק עם חותך $Y_t = \alpha + u_t$
חישוב האומדים	$\hat{\beta} = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{t=1}^{T}\left(X_t - \bar{X}\right)\left(Y_t - \bar{Y}\right)}{\sum_{t=1}^{T}\left(X_t - \bar{X}\right)^2}$ $= \frac{\sum_{t=1}^{T}\left(X_t - \bar{X}\right)Y_t}{\sum_{t=1}^{T}\left(X_t - \bar{X}\right)^2}$ $\hat{\alpha} = \bar{Y} - \hat{\beta}\bar{X}$	$\hat{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}$	$\hat{\alpha} = \bar{Y}$
תוחלת האומדים	$E\left(\hat{\beta}\right) = \beta$ $E\left(\hat{\alpha}\right) = \alpha$	$E\left(\hat{\beta}\right) = \beta$	$E\left(\hat{\alpha}\right) = \alpha$
שונות האומדים	$V\left(\hat{\beta}\right) = \frac{\sigma_u^2}{S_{XX}}$ $V\left(\hat{\alpha}\right) = \sigma_u^2\left(\frac{1}{T} + \frac{\bar{X}^2}{S_{XX}}\right)$	$V\left(\hat{\beta}\right) = \frac{\sigma_u^2}{\sum_{i=1}^{T} X_t^2}$	$V\left(\hat{\alpha}\right) = \frac{\sigma_u^2}{T}$

"המשוואות הנורמליות" מתקבלות בתהליך הגזירה של פונקציית הריבועים הפחותים וחייבות להתקיים על מנת שהפונקציה תתקיים $\left(\sum \hat{u}_t^2 = \min\right)$:

עבור המודל הקלאסי (עם חותך):

בגזירה של α : $\sum \hat{u}_t = 0$.

בגזירה של β : $\sum \hat{u}_t \cdot x_t = 0$.

עבור מודל ללא חותך:

בגזירת β בלבד: $\sum \hat{u}_t \cdot x_t = 0$.

מן המשוואות הנורמליות נובעות:

1. התכונות הגיאומטריות:

   א. $\sum \hat{u}_i = 0$.

   ב. $\sum x_i \hat{u}_i = 0$.

- ברגרסיה ללא שיפוע מתקיימת רק התכונה הגיאומטרית הראשונה.
  ברגרסיה ללא חותך מתקיימת רק התכונה הגיאומטרית השנייה.

2. התכונות האלגבריות:

   א. $\operatorname{cov}(x_i, \hat{u}_i) = 0$.

   ב. $\operatorname{cov}(\hat{y}_i, \hat{u}_i) = 0$.

   ג. $\bar{y} = \hat{\alpha} + \hat{\beta}\bar{x} = \bar{\hat{y}}$.

- התכונות האלגבריות תקפות עבור קו הרגרסיה הקלאסי (עם חותך ושיפוע) במדגם בלבד.

ההנחות הקלאסיות של מודל הרגרסיה:

1. קיים קשר ליניארי בין המשתנה המוסבר למשתנה המסביר.
2. X איננו קבוע: $S_{XX} = \sum_{t=1}^{T} \left(X_t - \bar{X}\right)^2 \neq 0$.
3. תוחלת ההפרעה האקראית היא אפס לכל תצפית: $E(u_t) = 0$ לכל t.
4. X_t אינם משתנים מקריים $\Leftarrow$ ניתן להוציא אותם מחוץ לתוחלת ולשונות $\Leftarrow$ $\operatorname{cov}(X_t, u_t) = 0$.

5. הומוסקדסטיות: שונות ההפרעה האקראית קבועה לכל תצפית:
$V(u_t)=\sigma_u^2$ לכל t.

6. u_t ב"ת: $\operatorname{cov}(u_t,u_s)=0$ לכל $t \neq s$.

7. ההפרעות האקראיות מתפלגות נורמלית: $u_t \approx N$.

תכונות האומדים:

אומדי הריבועים הפחותים הם ליניאריים, חסרי הטיה, יעילים ועקיבים.

1. ליניאריות:

אר"פ ניתנים להצגה כטרנספורמציה ליניארית של Y_t.

כדי ש-$\hat{\beta}$ למשל, יהיה אומד ליניארי צריך להתקיים: $\hat{\beta}=\sum W_t \cdot Y_t$.

כאשר W_t היא קומבינציה של ערכי X בדרך כלל. למשל: $\hat{\beta}=\dfrac{\sum X_t \cdot Y_t}{\sum X_t^2}$.

כדי להביא את האומד לצורה: $\tilde{\beta}=\sum w_t \cdot y_i$ נעזר בשוויון: $\dfrac{\sum()}{\sum()}=\sum\dfrac{()}{\sum()}$.

אומד זה ניתן להצגה בצורה הבאה:

$$\hat{\beta}=\sum\frac{X_t}{\sum X_t^2}Y_t=\sum W_t \cdot Y_t$$

$$W_t=\frac{X_t}{\sum X_t^2}$$

לפיכך מדובר באומד ליניארי.

- שימו לב כי:

W_t אסור שיכלול את Y_t.

Y_t אסור שיהיה במכנה או בשורש/חזקה (אלא אם כן במודל הנתון הוא מצוי בשורש/חזקה).

2. חוסר הטיה:
אומד $\hat{\vartheta}$ מסוים יהווה אח"ה לפרמטר ϑ אותו הוא אומד באוכלוסייה אם מתקיים: $E(\hat{\vartheta}) = \vartheta$.

כיצד יודעים אם אומד הוא חסר הטיה?

1. בשלב הראשון יש לבצע עבודת הכנה – מבטאים את האומד באמצעות הפרמטר האמיתי – מציבים במקום ה- Y_t את המודל ומפתחים אלגברית.

- יש לזכור כי:
  u_t, y_t מהווים משתנים מקריים ⇐ נשארים בתוך התוחלת, השונות וה- $\sum$.
  x_t איננו משתנה מקרי (על פי הנחה מס' 4) ⇐ יוצא מחוץ לתוחלת ולשונות אך נשאר בתוך ה- $\sum$ ו- α, β קבועים ⇐ יוצאים מחוץ לתוחלת, לשונות ול- $\sum$.

2. בשלב השני מפעילים תוחלת על האומד המפותח ואם התוחלת שווה לפרמטר האמיתי אז האומד חסר הטייה.

- חוסר הטייה מחייב את התקיימותן של הנחות (3) $E\left(u_t\right)=0$ לכל t ו- (4) $\operatorname{cov}\left(X_t, u_t\right)=0$.

3. יעילות:
יעילות פירושה השונות הקטנה ביותר. ככל שהשונות של האומד קטנה יותר, כך יש הסתברות גבוהה יותר שהוא יהיה קרוב לפרמטר האמיתי באוכלוסייה אותו הוא אומד.
$\hat{\vartheta}_1$ יקרא אומד יעיל יותר מ- $\hat{\vartheta}_2$ אם מתקיים שהשונות שלו קטנה יותר: $V(\hat{\vartheta}_1) < V(\hat{\vartheta}_2)$.

משפט גאוס מרקוב – אר"פ הם בעלי השונות הנמוכה ביותר בקבוצה שלהם (קבוצת האומדים הלינאריים חסרי ההטיה), והם נקראים: B.L.U.E (Best Linear Unbiased Estimation).

כיצד מחשבים שונות של אומד?

חייבות להתקיים הנחות (4) $\text{cov}(X_t, u_t) = 0$, (5) $V(u_t) = \sigma_u^2$ לכל t

ו-(6) $\text{cov}(u_t, u_s) = 0$ לכל $t \neq s$. אם הן מתקיימות, מחשבים את השונות של האיברים המכילים את u_t מהפיתוח הקודם (לפי כללי הסיגמא והשונות).

4. עקיבות:
ככל שהמדגם יגדל כן יתקרב האומד לערך האמיתי של הפרמטר.
אם נגדיל את המדגם לאינסוף תצפיות ונחשב את האומד, הוא יהיה שווה לפרמטר האמיתי באוכלוסייה: $\begin{pmatrix} \hat{\theta} \to \vartheta \\ T \to \infty \end{pmatrix}$.

תנאי הכרחי לעקיבות:
האומד חייב להיות פונקציה של גודל המדגם. במילים אחרות, האומד צריך להיות מושפע מגודל המדגם. ברגע שהאומד עונה על תנאי זה הוא יהיה עקיב. אומד המחושב במדגם סופי בהגדרה לא יוכל להיות עקיב לפרמטר באוכלוסייה.

סיכום: השלבים להוכחת התכונות:

1. הוכחת ליניאריות.
2. הכנת האומד $\Leftarrow$ להציב במקום Y_t את המודל האמיתי.
   במודל עם חותך: $Y_t = \alpha + \beta X_t + u_t$.
   במודל ללא חותך: $Y_t = \beta X_t + u_t$.
3. פיתוח האלגברה.
4. חישוב תוחלת, שונות, עקיבות.

- ליניאריות מהווה תנאי הכרחי לחוסר הטיה.
- ליניאריות וחוסר הטיה מהוות תנאי הכרחי לבחינת היעילות של האומד לפי משפט גאוס-מרקוב.
- עקיבות איננה תלויה בתכונות האחרות, אלא רק בהיותו של האומד פונקציה של גודל המדגם (לא מחושב על מדגם סופי). כך שאומד לא חייב להיות ליניארי או חסר הטיה כדי להיות עקיב.
- העקיבות משפיעה על היעילות של האומד. עבור אומדים התלויים בגודל המדגם: ככל שגודל המדגם גדול יותר כך שונות האומד קטנה והאומד יהיה יעיל יותר לפרמטר באוכ'.

שאלות:

תרגול ממבחנים:

1) נתון המודל: $Y_t = \alpha + \beta X_t + u_t$, $T = 100$, כאשר מתקיימות כל ההנחות הקלאסיות.

נתון האומד: $\tilde{\beta} = \dfrac{\sum_{t=51}^{100} Y_t - \sum_{t=1}^{50} Y_t}{\sum_{t=51}^{100} X_t - \sum_{t=1}^{50} X_t}$.

א. האומד $\tilde{\beta}$ הינו אומד חסר הטיה ל- β.   נכון / לא נכון

ב. האומד $\tilde{\beta}$ הינו אומד עקיב ל- β.   נכון / לא נכון

ג. האומד $\tilde{\beta}$ הינו אומד ליניארי ל- β.   נכון / לא נכון

ד. האומד $\tilde{\beta}$ הינו אומד יעיל ל- β.   נכון / לא נכון

ה. השונות האמיתית של $\tilde{\beta}$ היא?

2) נתון המודל: $Y_t = \beta X_t + u_t$, כאשר כל ההנחות הקלאסיות מתקיימות.
(יש לשים לב המודל ללא חותך).

נתון האומד: $\tilde{\beta} = \dfrac{\sum Y_t}{\sum X_t}$.

א. האומד $\tilde{\beta}$ הינו אומד מוטה ל- β: נכון / לא נכון / אי אפשר לדעת

ב. על סמך משפט גאוס-מרקוב ניתן להסיק כי $\tilde{\beta}$ איננו אומד יעיל יותר מאומד הריבועים הפחותים: נכון / לא נכון / לא ניתן לדעת

ג. מהי השונות האמיתית של $\tilde{\beta}$?

3) נתון המודל: $Y_t = \beta X_t + u_t$, כאשר כל ההנחות הקלאסיות מתקיימות.
(יש לשים לב המודל ללא חותך).

נתון האומד: $\tilde{\beta} = \dfrac{\sum X_t Y_t}{\sum \left(X_t - \bar{X}\right)^2}$.

א. מהי התוחלת של $\tilde{\beta}$?

ב. $E\left(\tilde{\beta}\right) < \beta$.   נכון / לא נכון / אי אפשר לדעת

ג. על סמך משפט גאוס-מרקוב ניתן להסיק כי אומד הריבועים הפחותים הינו אומד יעיל יותר מ- $\tilde{\beta}$.   נכון / לא נכון/ לא ניתן לדעת

ד. מהי השונות האמיתית של האומד $\dfrac{\sum X_t Y_t}{\sum X_t^2}$?

4) בכל השאלות ההנחות הקלאסיות מתקיימות.
האומדים הם אר"פ, והמודל הוא: $Y_t = \alpha + \beta X_t + u_t$.

א. $E(Y_t) = E(\hat{Y}_t)$ נכון / לא נכון / אי אפשר לדעת

ב. $\sum_{t=1}^{T}(X_t - \bar{X})\bar{Y} \neq 0$ נכון / לא נכון / אי אפשר לדעת

ג. אמידת המודל בשיטת הריבועים הפחותים תיתן את התוצאה: $\sum_{t=1}^{T} u_t = 0$. נכון / לא נכון / אי אפשר לדעת

ד. אם נתון ש- $r_{XY} = 0.57$, אזי $\hat{\beta}$:
i. הוא בהכרח שלילי.
ii. הוא בהכרח חיובי.
iii. הוא בהכרח שווה לאפס.
iv. לא ניתן לקבוע את סימנו על סמך הנתונים הקיימים.

ה. סמן את הטענה הנכונה בהכרח:
i. $\sum_{t=1}^{T}(X_t - \bar{Y})\hat{u}_t = 0$.
ii. $S_{XX} = \sum_{t=1}^{T} X_t^2 - (T\bar{X})^2$.
iii. $\sum_{t=1}^{T} X_t u_t = 0$.
iv. אף אחת מהטענות הנ"ל אינה נכונה בהכרח.

ו. אומדי הריבועים הפחותים אינם חסרי הטיה, אם נתון שהשונות של u_t אינה קבועה. נכון / לא נכון / אי אפשר לדעת

ז. אומד חסר הטיה הוא אינו בהכרח גם אומד עקיב. נכון / לא נכון / אי אפשר לדעת

תשובות סופיות:

1) א. נכון. ב. לא נכון. ג. נכון. ד. לא נכון.
ה. $V\left(\tilde{\beta}\right)=\dfrac{100\sigma_u^2}{\left(\sum_{t=51}^{100}X_t-\sum_{t=1}^{50}X_t\right)^2}$.

2) א. לא נכון. ב. נכון. ג. $V\left(\tilde{\beta}\right)=\dfrac{T\sigma_u^2}{\left(\sum X_t\right)^2}$.

3) א. $E\left(\tilde{\beta}\right)=\dfrac{\beta\sum X_t^2}{\sum\left(X_t-\bar{X}\right)^2}$. ב. לא נכון. ג. לא נכון.
ד. $\dfrac{\sigma_u^2}{\sum X_t^2}$.

4) א. נכון. ב. לא נכון. ג. לא נכון. ד. ii.
ה. i. ו. לא נכון. ז. נכון.

מבוא לאקונומטריקה

פרק 3 - מודלים לא ליניאריים

תוכן העניינים

מודלים לא ליניאריים:

רקע:

המודל	משמעות ה- β	השינוי השולי $\left(\frac{\partial Y}{\partial X}\right)$ בכמה ישתנה Y אם נגדיל את X ביחידה?	הגמישות $\left(\frac{\partial Y}{\partial X}\cdot\frac{X}{Y}\right)$ בכמה % ישתנה Y אם נגדיל את X ב-1%?
ליניארי: $Y=\alpha+\beta X+u$	השינוי השולי אם נגדיל את X ביחידה Y ישתנה ב- β יחידות	β	$\frac{\beta X}{Y}$
חצי לוגריתמי: $\ln Y=\alpha+\beta X+u$ $\left(Y=e^{\alpha+\beta X+u}\right)$	שיעור השינוי השולי אם נגדיל את X ביחידה Y ישתנה ב- $100\cdot\beta\%$	βY	βX
לוגריתמי כפול: $\ln Y=\alpha+\beta\ln X+u$ $\left(Y=e^{\alpha}\cdot X^{\beta}\cdot e^{u}\right)$	הגמישות אם נגדיל את X ב-1% Y ישתנה ב- $\beta\%$	$\frac{\beta Y}{X}$	β
לוג ליניארי: $Y=\alpha+\beta\ln X+u$ $\left(e^{y}=e^{\alpha}\cdot X^{\beta}\cdot e^{u}\right)$	אין משמעות כלכלית אם נגדיל את X ב-1% Y ישתנה ב- β	$\frac{\beta}{X}$	$\frac{\beta}{Y}$

- המשתנה שיש בו LN השינוי בו יהיה באחוזים.

תזכורת של חוקי לוגים:

$$LN\left(e^{x}\right)=X$$
$$LN\left(X^{Y}\right)=Y\cdot LN\left(X\right)$$
$$LN\left(X\cdot Y\right)=LN\left(X\right)+LN\left(Y\right)$$
$$LN\left(\frac{X}{Y}\right)=LN\left(X\right)-LN\left(Y\right)$$

שאלות:

1) על מנת לאמד את התשואה להשכלה בישראל בשנים 1948-1990 נאמדו המודלים הבאים:

1. $MWAGE_t = 139.547 + 118.628 \cdot SCL_t$.
2. $MWAGE_t = -1445.08 + 1239.60 \cdot LN(SCL)_t$.
3. $LN(MWAGE)_t = 5.244 + 0.778 \cdot LN(SCL)_t$.
4. $LN(MWAGE)_t = 6.292 + 0.070 \cdot SCL_t$.

א. הסבירו את המשמעות של β בכל אחד מהמודלים.

ב. חשבו את הגמישות בנקודת הממוצעים: (12.311,1600.01) עבור כל אחד מהמודלים.

2) נתונים תוצאות האמידה של המודלים הבאים:

1. $\hat{Y} = e^{4.5} \cdot X^{0.05}$.
2. $\hat{Y} = e^{4.5+0.05X}$.
3. $\hat{Y} = 4.5 + \dfrac{0.05}{X}$.
4. $\hat{Y} = \dfrac{1}{1+e^{4.5+0.05X}}$.

א. כתבו את המודלים בצורה ליניארית בעזרת טרנספורמציה מתאימה.

ב. עבור כל אחד מהמודלים ערכו תחזית נקודתית עבור $X = 6$.

3) נתונים המודלים הבאים עבור התוצר במשק:

1. $Q_i = AK_i^{\beta_1} e^{u_i}$.
2. $Q_i = Ae^{\beta_1 L_i + u_i}$.
3. $Q_i = A + K_i^{\beta_1} + e^{u_i}$.
4. $Q_i = A + \dfrac{\beta_1}{L_i} + u_i$.
5. $Q_i = A + \beta_1 \sqrt{K_i} + u_i$.
6. $Q_i = e^{A+\beta_1 Ki + u_i}$.
7. $Q_i = A\left(\dfrac{K_i}{2} + 7\right)^{\beta_1} e^{u_i}$.
8. $Q_i = A + \beta_1 L_i + u_i$.
9. $Q_i = A + \beta_1 \left(\dfrac{K_i}{L_i}\right) + u_i$.

כאשר:

Q - הוצאות צריכה על מוצר מסוים על ידי פרט מסוים.

A - הוצאות צריכה על המוצר בהינתן רמת הכנסה אפסית.

K - הכנסת הפרט.

L - שנות לימוד.

א. מי מהמודלים הבאים ניתן לאמידה בשיטת OLS?

ב. מי מבין המודלים שלא ניתנים לאמידה בשיטת OLS ניתן להביא למודל ליניארי בפרמטרים ועל כן לאמוד את הפרמטרים שלו?

ג. עבור כל אחד מהמודלים קבעו מיהו המשתנה המוסבר ומיהו המסביר במשוואת הרגרסיה הליניארית.

ד. עקומת אנג'ל מתארת את גמישות הצריכה של הפרט מוצר מסוים ביחס להכנסתו. איזה מהמודלים מתאים כדי לתאר את עקומת אנג'ל?

4) נתון המודל הבא: $Q_i = \frac{A}{K_i^{\beta_1}} e^{u_i}$.

א. האם ניתן לאמוד את המודל בשיטת OLS?

ב. מה המשוואה שצריך לאמוד על מנת לקבל את הפרמטרים למודל זה (כלומר כיצד הופכים את המודל לליניארי בפרמטרים)?

ג. נאמד המודל הבא: $\ln(Q_i) = \alpha_0 + \alpha_1 \ln(K_i) + u_i$, והתקבלו התוצאות הבאות: $\hat{\alpha}_0 = 3\ ,\ \hat{\alpha}_1 = 0.8$.

מהם האומדנים עבור: $A\ ,\ \beta_1$?

5) נתון כי הקשר באוכלוסייה בין X ל- Y נתון על ידי המודל הבא: $\ln Y = \alpha + \beta \ln X + u$. נתון גם כי עבור המודל הנ"ל כל ההנחות הקלאסיות מתקיימות.

כלכלן הציע את האומד הבא עבור β: $\tilde{\beta} = \frac{\sum_{t=1}^{T} (\ln X_t - \ln \bar{X}) \ln Y_t}{\sum_{t=1}^{T} (\ln X_t - \ln \bar{X})^2}$.

א. האם האומד ליניארי?

ב. האם האומד חסר הטיה?

ג. האם האומד *blue*?

ד. מהי שונותו?

תשובות סופיות:

1) א.1. השינוי השולי. 2. אין משמעות כלכלית. 3. גמישות.
4. שיעור השינוי השולי.
ב.1. 0.912. 2. 0.77. 3. 0.778. 4. 0.861.

2) א.1. $\hat{\ln}(Y)=4.5+0.05\cdot\ln(X)$. 2. $\hat{\ln}(Y)=4.5+0.05X$.
3. אין צורך. 4. $\ln\left(\frac{1-\hat{Y}}{\hat{Y}}\right)=4.5+0.05X$.
ב.1. 98.45. 2. 121.51. 3. 4.50833. 4. 0.00816.

3) א. מודלים: 4 , 5 , 8 ו-9.
ב. מודלים: 1 , 2 , 6 ו-7.
ג.1. מסביר: $\ln(K_i)$, מוסבר: $\ln(Q_i)$. 2. מסביר: L_i , מוסבר: $\ln(Q_i)$.
3. אינו ליניארי. 4. מסביר: $\frac{1}{L_i}$, מוסבר: Q_i.
5. מסביר: $\sqrt{K_i}$, מוסבר: Q_i. 6. מסביר: K_i , מוסבר: $\ln(Q_i)$.
7. מסביר: $K_i=\frac{K_i}{2}+7$, מוסבר: $\ln(Q_i)$. 8. מסביר: L_i , מוסבר: Q_i.
9. מסביר: $\frac{K_i}{L_i}$, מוסבר: Q_i.
ד. מודלים: 1 ו-7.

4) א. לא. ב. $\ln(Q_i)=\ln(A)-\beta_1\ln(K_i)+u_i$.
ג. $A=20$, $\beta_1=-0.8$.

5) א. כן. ב. כן. ג. כן. ד. $V(\hat{\beta})=\frac{\sigma_u^2}{SS\ln x}$.

מבוא לאקונומטריקה

פרק 4 - מבחני המובהקות וקריאת פלטים - תוכנת SAS

תוכן העניינים

מבחני המובהקות וקריאת פלטים – תוכנת SAS:

רקע:

פלט ניתוח שונות (Analysis of Variance):

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	k	RSS	$RSS/k = MSR$	$F = \frac{MSR}{MSE}$	PF
	$+$	$+$			
Error	$T-k-1$	ESS	$ESS/(T-k-1) = MSE$		
C Total	$T-1$	TSS			

Root MSE	$\sqrt{MSE} = s_u$	R-square	$R^2 = \frac{RSS}{TSS}$
Dep Mean	$\overline{Y}$	Adj R-sq	$\overline{R}^2 = 1 - \frac{ESS}{TSS} \cdot \frac{T-1}{T-k-1}$
C.V.	$\frac{s_u}{\overline{Y}} \cdot 100$		

פלט מקדמי הרגרסיה (Parameter Estimates):

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	$\hat{\alpha}$	$s_{\hat{\alpha}}$	$\frac{\hat{\alpha}}{s_{\hat{\alpha}}} = t_{(\hat{\alpha}=0)}$	$Pt_{\hat{\alpha}}$
X	1	$\hat{\beta}$	$s_{\hat{\beta}}$	$\frac{\hat{\beta}}{s_{\hat{\beta}}} = t_{(\hat{\beta}=0)}$	$Pt_{\hat{\beta}}$

פלט ה – Covariance of Estimates:

פלט שמתאר את השונות המשותפת (covariance) של האומדנים- $\hat{\alpha}$ ו- $\hat{\beta}$:

Covariance of Estimates		
COVB	**INTERCEP**	X
INTERCEP	$s_{\hat{\alpha}}^2$	$\text{cov}(\hat{\alpha},\hat{\beta})$
X	$\text{cov}(\hat{\alpha},\hat{\beta})$	$s_{\hat{\beta}}^2$

עריכת תחזית וקריאת פלטים (תוכנת SPSS):

אמידה נקודתית:
אמידה נקודתית עבור X_0 מסוים (תחזית).
מחושבת על פי קו הרגרסיה במדגם: $\hat{Y}_0 = \hat{\alpha} + \hat{\beta} \cdot X_0$.

אמידה מרווח ל- $E(Y)$:
אמידת התחזית באוכלוסייה עבור X_0 מסוים. נחשב רווח בר סמך לערך ממוצע של Y באוכ' עבור X_0 מסוים ($E(Y)$) ברמת סמך $1-\alpha$.

נוסחת הרב"ס: $\hat{Y} \pm t_{n-2;1-\frac{\alpha}{2}} \hat{\sigma}_u \sqrt{\frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$.

$\sum (X_i - \bar{X})^2 = S_{XX} = (n-1)S_X^2$, $\hat{\sigma}_u = MSE = \frac{SSE}{n-2}$.

רישום הרב"ס: $p(\_\_\_ \le E(Y) \le \_\_\_) = 1-\alpha$.

אמידה מרווח ל- Y :
אמידת ערך בודד של Y באוכלוסייה עבור X_0 מסוים. נחשב רווח בר סמך לערך בודד של Y באוכ' עבור X_0 מסוים (Y_0) ברמת סמך $1-\alpha$.

נוסחת הרב"ס: $\hat{Y} \pm t_{n-2;1-\frac{\alpha}{2}} \hat{\sigma}_u \sqrt{1 + \frac{1}{n} + \frac{(X_0 - \bar{X})^2}{\sum (X_i - \bar{X})^2}}$.

רישום הרב"ס: $p(\_\_\_ \le Y \le \_\_\_) = 1-\alpha$.

- רב"ס לערך בודד יהיה רחב יותר מאשר רב"ס לערך ממוצע משום שטעות התקן בראשון גדולה מאשר באחרון.

שאלות:

פלט ניתוח שונות:

1) חוקר רצה לבחון את השפעת ההכנסה ($INCOME$) על גובה המס (TAX) (במיליארדי $) שגובה מדינה במערב לפי המודל: $TAX_t = \alpha + \beta \cdot INCOME_t + u_t$.
לשם כך אסף נתונים מ-51 מדינות. להלן התוצאות:

Model: MODEL1

Dependent Variable: TAX

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2046.89694	2046.89694	8798.672	0.0001
Error	49	11.39922	0.23264		
C Total	50	2058.29615			

Root MSE	0.48232	R-square	0.9945
Dep Mean	5.4242	Adj R-sq	0.9943
C.V.	8.88711		

בדקו את ההשערה כי המודל מובהק ברמת מובהקות של 0.05.

פלט מקדמי הרגרסיה:

2) בהמשך לדוגמא הקודמת – בדיקת השפעת ההכנסה על גודל המס, התקבלו גם התוצאות הבאות:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-0.086912	0.08953904	-0.971	0.3365
INCOME	1	0.152232	0.0016229	93.801	0.0001

א. אמדו את המודל: $TAX = \alpha + \beta \cdot INCOME + U$.
מהי המשמעות הכלכלית של β?

ב. האם המודל מובהק? בדקו על סמך הפלט הנ"ל ברמת מובהקות של 0.05.

ג. מהי רמת המובהקות הקטנה ביותר, עבורה עדיין תידחה השערת האפס מסעיף ב'?

ד. בדקו את ההשערה כי ככל שההכנסה עולה כך עולה גם המס (שיפוע β חיובי) ברמת מובהקות של 0.01.

ה. בנו רווח-סמך ברמת סמך של 95% עבור β.

ו. בדקו את ההשערה שתוספת של מיליארד $ להכנסה תגדיל את המס ב-0.2 מיליארד $, ברמת מובהקות של 0.05.

- שימו לב כי:
  במודל עם משתנה מסביר אחד בלבד קיימת זהות בין מבחן F למובהקות המודל לבין מבחן t למובהקות ה-β:

$$F_{(1,T-2;1-\alpha)} = t^2_{\left(T-2;1-\frac{\alpha}{2}\right)}$$

$$F = t^2_{\hat{\beta}}$$

  כלומר: כל החלטה המתקבלת במבחן אחד חייבת להיות זהה להחלטה המתקבלת במבחן השני.

פלט שונויות משותפות:

3) נתון פלט האמידה של המודל: $Y_t = \alpha + \beta X_t + u_t$, שלצורך אמידתו נאספו 240 תצפיות:

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	5.25	0.25	21	0.0000
X	1	0.96	0.12	8	0.0000

Covariance of Estimates

	INTERCEP	X
INTERCEP	0.0625	-0.003
X	-0.003	0.0144

יש לבדוק את ההשערה: $H_0: \alpha = 5\beta$.

שאלה מסכמת:

4) חוקר רצה לבדוק את השפעת הותק בעבודה (EXP) על השכר ($SALARY$) לפי המודל: $\ln(SALARY_t) = \alpha + \beta \cdot EXP_t + u_t$. הוא אסף 403 תצפיות, ואמד את הפרמטרים בתוכנת SAS. להלן חלקים מהפלט ויש להשלימו:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	----	----	5.68015	----	----
Error	----	205.22539	----		
C Total	----	----			

Root MSE	----	R-square	----
Dep Mean	7.14247	Adj R-sq	0.0245
C.V.	10.01602		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	---- (*)	----	----	----
EXP	1	-0.008740	----	----	0.0009

Covariance of Estimates

COVB	INTERCEP	EXP
INTERCEP	0.0047463101	----
EXP	-0.000154685	6.882844 E-6

- נתון נוסף: $EXP = 22$.

א. קיים קשר חיובי מובהק בין ותק ללוג השכר. נכון / לא נכון

ב. שיעור התשואה בשכר לשנת ותק הוא?

ג. תחזית לוג השכר עבור אדם בעל 10 שנות ותק היא?

ביצוע תחזיות:

5) במדגם של 30 דירות מושכרות לסטודנטים ברדיוס של עד 2 ק"מ מסביב למכללה נחקר הקשר בין שכר דירה למספר הסטודנטים הגרים בדירה. להלן התוצאות:

Descriptive Statistics

	Mean	Std. Deviation	N
שכר הדירה	1386.7667	509.46027	30
מספר הסטודנטים	3.0000	1.31306	30

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.602[a]	.362	.339	414.05503

a. Predictors: (Constant), number of students
b. Dependent Variable:rent

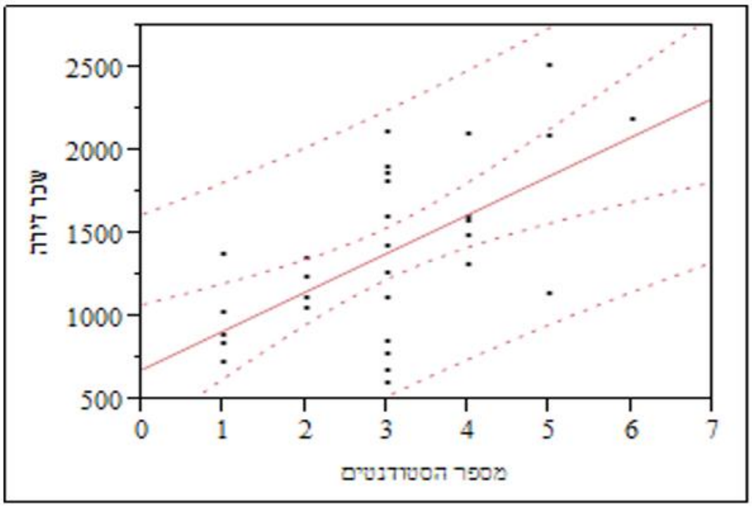
ANOVA[b]

Model		Sum of Squares	Df	Mean Square	F	Sig.
1	Regression	2726579.520	1	2726579.520	15.904	.000[a]
	Residual	4800363.847	28	171441.566		
	Total	7526943.367	29			

a. Predictors: (Constant), number of students
b. Dependent Variable: rent

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	T	Sig.
1	(Constant)	686.207	191.244		3.588	.001
	מספר הסטודנטים	233.520	58.556	.602	3.988	.000



א. חשב אומדן נקודתי לשכר הדירה אותו ישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד.

ב. אמוד את שכר הדירה הממוצע שישלמו סטודנטים החולקים את הדירה עם שותף אחד בלבד, ברמת בטחון של 95%.

ג. אמוד את שכר הדירה שישלם סטודנט יחיד החולק את הדירה עם שותף אחד בלבד, ברמת ביטחון של 95%.

תשובות סופיות:

1) יש עדות לכך.

2) א. ראה סרטון. ב. יש עדות לכך. ג. $Pt_{\hat{\beta}} = 0.0001$.

ד. יש עדות לכך. ה. $P\left(0.1488 \le \beta \le 0.1554\right) = 0.95$.

ו. יש עדות לכך.

3) אין עדות לכך.

4) א. לא נכון. ב. -0.87%. ג. 7.24735.

5) א. 1153.247. ב. $p\left(957.4 \le \mu_{Y_{X=2}} \le 1349.08\right) = 0.95$.

ג. $p\left(282.94 \le Y_{X=2} \le 2023.55\right) = 0.95$.

מבוא לאקונומטריקה

פרק 5 - שינוי יחידות מדידה

תוכן העניינים

שינוי יחידות מדידה:

רקע:

טרנספורמציה ליניארית: הוספה/החסרה של קבוע ו/או הכפלה/חילוק של קבוע של אחד או שני המשתנים (התלוי והב"ת).

- טרנספורמציה ליניארית של המשתנים לא תשפיע על: R^2, F, $t_{\hat{\beta}}$ ו- PF.

השינויים מסוכמים בטבלה הבאה:

	$\hat{\beta}'$	$\hat{\alpha}'$	$S_{\hat{\beta}}$	$S\hat{\alpha}$
הוספת קבוע ל- X: $Y=\alpha'+\beta'(X+d)+v$	$\hat{\beta}'=\hat{\beta}$	$\hat{\alpha}'=\hat{\alpha}-\hat{\beta}d$	$s_{\hat{\beta}'}=s_{\hat{\beta}}$	$s_{\hat{\alpha}'}\neq s_{\hat{\alpha}}$
הוספת קבוע ל- Y: $Y+d=\alpha'+\beta'X+v$	$\hat{\beta}'=\hat{\beta}$	$\hat{\alpha}'=\hat{\alpha}+d$	$s_{\hat{\beta}'}=s_{\hat{\beta}}$	$s_{\hat{\alpha}'}=s_{\hat{\alpha}}$
הכפלת X **פי קבוע:** $Y=\alpha'+\beta'(dX)+v$	$\hat{\beta}'=\frac{\hat{\beta}}{d}$	$\hat{\alpha}'=\hat{\alpha}$	$s_{\hat{\beta}'}=\frac{s_{\hat{\beta}}}{d}$	$s_{\hat{\alpha}'}=s_{\hat{\alpha}}$
הכפלת Y **פי קבוע:** $dY=\alpha'+\beta'X+v$	$\hat{\beta}'=d\hat{\beta}$	$\hat{\alpha}'=d\hat{\alpha}$	$s_{\hat{\beta}'}=ds_{\hat{\beta}}$	$s_{\hat{\alpha}'}=ds_{\hat{\alpha}}$

- $t_{(\hat{\beta}'=0)}=t_{(\hat{\beta}=0)}$ תמיד.
- $t_{(\hat{\alpha}'=0)}=t_{(\hat{\alpha}=0)}$ רק בהכפלות.

שאלות:

1) חוקר ביקש לאמוד את הקשר בין שכר ב-₪ ($MWAGE$) לבין שנות לימוד (SCL) באמצעות 2 מודלים שונים.
להלן תוצאות האמידה:

א. $MWAGE_t = 139.54 + 118.62 \cdot SCL_t$.

ב. $MWAGE_t = -1445.08 + 1239.60 \cdot LN\left(SCL\right)_t$.

חשבו מחדש את מקדמי הרגרסיה וסטטיסטי המבחן F בכל אחד מהמודלים כתוצאה:

1. התברר כי נעשה טעות בחישוב מספר שנות הלימוד, ויש צורך להוסיף 20% למשתנה המקורי.
2. התברר כי הקשר בין שכר לשנות לימוד הוא ריבועי ולכן יש צורך להעלות את המשתנה המקורי של מספר שנות הלימוד בריבוע.

2) בהמשך לנתוני השאלה לדוגמא מהפרק החמישי:
החוקר טען כי יש לבדוק את הקשר בין שכר לוותק ע"י שימוש בשכר נטו (NET) ולא בשכר ברוטו ($SALARY$). (קיים שיעור מס קבוע של 20%).
המודל הוא: $\ln\left(NET_t\right) = \alpha' + \beta' \cdot EXP_t + v_t$.
מה יהיו ערכי האומדים, סטיות התקן שלהם וטיב ההתאמה באמידת מודל זה?

תשובות סופיות:

1) 1. א. $\hat{\alpha}' = \alpha = 139.59$, $\hat{\beta}' = 98.85$, סטטיסטי F לא משתנה.
ב. $\hat{\alpha}' = -1671$, $\hat{\beta}' = 1239.6$, סטטיסטי F לא משתנה.
2. לא ניתן לדעת.

2) $S_{\hat{\alpha}'} = S_{\hat{\alpha}} = 0.0688935$, $S_{\hat{\beta}'} = S_{\hat{\beta}} = 0.0026235$, $\hat{\alpha}' = 7.11161$, $\hat{\beta}' = \hat{\beta} = -0.00874$,
$R^2 = 0.0269$.

מבוא לאקונומטריקה

פרק 6 - הרגרסיה המרובה

תוכן העניינים

רגרסיה מרובה:

רקע:

מבחן T ו- F:

כאשר יש יותר ממשתנה מסביר אחד, מדובר ברגרסיה מרובה.

המודל הקלאסי: $Y_t = \alpha + \beta_1 X_{1t} + \beta_2 X_{2t} + \beta_3 X_{3t} + \ldots + \beta_k X_{kt} + u_t$.

- קבוע α יש אחד.
- מספר ה- β טות כמספר המשתנים הב״ת במודל.

מבחן F למובהקות המודל:

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

סטטיסטי המבחן F וכלל ההכרעה:

$$F = \frac{\frac{RSS}{k}}{\frac{ESS}{T-k-1}} = \frac{\frac{R^2}{k}}{\frac{1-R^2}{T-k-1}} > F\left(k, T-k-1; 1-\alpha\right).$$

מבחן t למובהקות ה- β טות:

מבחן לבדיקת מובהקות β ספציפית:

השערות: $\begin{matrix} H_0 = \beta_1 = 0 \\ H_1 : \beta_1 \neq 0 \end{matrix}$.

סטטיסטי המבחן t וכלל ההכרעה:

$$\left|t_{\hat{\beta}_i}\right| = \left|\frac{\hat{\beta}_i}{S_{\hat{\beta}_i}}\right| > t_{\left(T-k-1; 1-\frac{\alpha}{2}\right)}.$$

השוואה בין מודלים – $\bar{R}^2$ וחוק חיטובסקי:

בכדי להחליט האם כדאי לנו להוסיף למודל משתנה ב"ת מסוים:
נשווה את פרופורציית השונות המוסברת המתוקנת $\bar{R}^2$ בין המודל ללא המשתנה המסביר לבין המודל עם המשתנה המסביר שהוספנו.

- ניתן להשתמש גם באומד המוטה - R^2 להשוואה בין מודלים אם מתקיימים שני התנאים הבאים:
  1. מספר המשתנים זהה.
  2. המשתנה המוסבר זהה.

לפי חוק חיטובסקי – בהוספת משתנה מסביר אחד בלבד למודל ה- $\bar{R}^2$ יעלה אך ורק אם: $|t_{\hat{\beta}}|>1$.

כאשר: $|t_{\hat{\beta}}|<1$ אז $\bar{R}^2$ ירד בהוספת המשתנה והוא גם לא יהיה רלוונטי למודל (מובהק).

כאשר: $|t_{\hat{\beta}}|>2$ אז $\bar{R}^2$ יעלה והמשתנה שהוסף יהיה גם מובהק.

כאשר: $1<|t_{\hat{\beta}}|<2$ אז ה- $\bar{R}^2$ יעלה אך יש לבדוק את רלוונטיות המשתנה שהוסף למודל על פי מבחן t.

שאלות:

מבחן T ו-F:

1) נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

Analysis of Variance

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

Root MSE	------------	R-square	-------------
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

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
X	1	-------------	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-------------	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

א. השלם את הנתונים החסרים בפלט.

ב. האם המודל מובהק? בדקו ברמת מובהקות של 0.05.

ג. האם משתנה W רלוונטי למודל? בדקו ברמת מובהקות של 0.01.

השוואה בין מודלים:

2) במודל לניבוי ההכנסה על פי שנות לימוד וותק במקום העבודה, התקבל: $\bar{R}^2 = 0.266$. הוסף המשתנה היקף המשרה. במבחן למובהקות המשתנה הנוסף התקבל: $t_{\hat{\beta}} = 0.456$. האם ערך $\bar{R}^2$ יעלה/ירד/לא ישתנה בהוספת המשתנה הנוסף למודל?

מבחן Wald ו-T מורכב:

3) נאמד המודל: $Y_t = \alpha + \beta_x X_t + \beta_z Z_t + \beta_w W_t + \beta_s S_t + u_t$ והתקבלו התוצאות הבאות:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	646169.84	161542.46	51835.84	0.0000
Error	199	620.1683	3.1164236		
C Total	203	646790.01			

Root MSE	1.7653395	R-square	0.999041
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

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
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

הועלתה ההשערה כי ההשפעה על Y של משתנה S היא פי 3 מזו של משתנה Z, וכן כי החותך הוא 5.

א. מהי השערת האפס?

ב. מהו המודל המוגבל שאותו צריך לאמוד?

להלן אמידת המודל המוגבל:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	2	646166.01	323083.01		
Error	201	623.9983	3.104469		
C Total	203	646790.01			

Root MSE	1.7619504	R-square	0.999035
Dep Mean	173.6645	Adj R-sq	0.999026
C.V.	1.0145714		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
X	1	0.978491	0.036399	26.88240	0.0000
Z+3S	1	2.999995	0.003669	817.6080	0.0000
W	1	-5.043109	0.071218	-70.81249	0.0000

ג. חשב את הסטטיסטי של WALD.
ד. כמה דרגות חופש יש במונה וכמה במכנה?
ה. האם דוחים או מקבלים את השערת האפס?

4) על מנת לאמוד את פונקציית התצרוכת נאספו נתונים על 42 משקי בית בשנת 2007 ונאמדה המשוואה הבאה: $C_t = \alpha + \beta_1 \cdot W_t + \beta_2 \cdot P_t + u_t$.
להלן תוצאות האמידה של המשוואה הנ"ל:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	---	---------	--------	-------	--------
Error	---	---------	52968		
C Total	---	---------			

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	-107.226	------	----------	-------
W	1	0.743	------		
P	1	0.561	-------		

Covariance of Estimates

COV	INTERCEP	W	P
INTERCEP	-----------	-----------	--------
W	-----------	0.0046	-0.0090
P	-----------	-0.0090	0.016

על מנת לבדוק את ההשערה שהנטייה השולית לצרוך מתוך ההכנסה זהה לנטייה השולית לצרוך מתוך ההון, נאמדה גם המשוואה הבאה:
$C_t = \alpha + \beta_1 \cdot Y_t + u_t$, כאשר: Y_t = סה"כ ההכנסה של משק בית t.
התקבל: $ESS = 0.4566$.
בדקו את ההשערה בשתי דרכים.

תרגיל מסכם:

5) חוקר אמד את התצרוכת של 500 משקי בית כפונקציה של הכנסה שלהן לפי המשוואה: $EXPENSE_t = \alpha + \beta \cdot INCOME_t + u_t$.

$EXPENSE_t$ - התצרוכת של משק הבית ה-t-י באלפי שקלים.

$INCOME_t$ - ההכנסה של משק הבית ה-t-י באלפי שקלים.

ההפרעות האקראיות מקיימות את כל ההנחות הקלאסיות התקבל הפלט הבא:

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	1	2013.105	2013.105	6495.745	0.0000
Error	498	154.3358	0.3099112		
C Total	499	2167.441			

Root MSE	0.556697	R-square	0.928794
Dep Mean	3.990208	Adj R-sq	0.928651
C.V.	13.95157		

Parameter Estimates

Variable	DF	Parameter Estimate	Standard Error	T for H0: Parameter=0	Prob>|T|
INTERCEP	1	0.041995	0.054951	0.764236	0.4451
INCOME	1	0.713503	0.008853	80.59618	0.0000

א. מהו Pvalue לבדיקת מובהקות המודל ע"י מבחן F?

ב. מהו אחוז השונות בתצרוכת המוסבר ע"י ההכנסה?

ג. מהו אומדן לתצרוכת ההתחלתית של משק בית?

ד. האם אומדן זה מובהק?

ה. על עוזר מחקר הטיל החוקר לבדוק את ההשערה כי על כל 1000 ₪ נוספים בהכנסה צורך הפרט 700 ₪, כנגד ההשערה כי הוא צורך יותר מ-700 ₪. נסח את השערת האפס ואת ההשערה האלטרנטיבית.

ו. מהו הסטטיסטי t לבדיקת ההשערה?

ז. מהו הסטטיסטי WALD לבדיקת ההשערה?

ח. התברר כי הייתה טעות בנתונים, וכי יש להוסיף 1000 ₪ לתצרוכת של כל משק בית:

i. ההוספה תגדיל את האומד ל- α: נכון / לא נכון / לא ניתן לדעת.

ii. בעקבות ההוספה האומד ל- α יהיה מובהק: נכון / לא נכון / לא ניתן לדעת.

iii. ההוספה תשנה את האומד ל- β: נכון / לא נכון / לא ניתן לדעת.

iv. ההוספה תשנה את R^2: נכון / לא נכון / לא ניתן לדעת.

החוקר טען כי יש להוסיף לפונקציית התצרוכת גם את השפעת העושר.
העושר של משק בית מורכב מתוכניות החסכון שלו ($SAVINGS$) ומניירות הערך שיש לו (NE). שתי סדרות הנתונים הן באלפי שקלים.
החוקר אמד את המשוואה:

$$EXPENSE_t = \alpha + \beta_1 \cdot INCOME_t + \beta_2 \cdot SAVINGS_t + \beta_3 \cdot NE_t + u_t ,$$

וקיבל כי סכום ריבועי הסטיות של הטעויות הוא 121.

ט. מהי השערת האפס לבדיקת הטענה של החוקר (שהמודל החדש נכון ולא המקורי)?

י. מהו הסטטיסטי WALD לבדיקת ההשערה?

החוקר רצה לבדוק את ההשערה כי הנש"צ מתוך ההכנסה שווה ל-0.6 וכי השפעת ניירות הערך על התצרוכת היא פי 2 מהשפעת תוכניות החסכון.

יא. מהי השערת האפס לבדיקה זו?

יב. המודל המוגבל לבדיקת ההשערה יהיה מהצורה: $Z_t = \gamma_0 + \gamma_1 W_t + v_t$.

בטא את Z_t, ו- W_t באמצעות המשתנים המקוריים.

תשובות סופיות:

1) א.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Prob>F
Model	4	646169.84	161542.46	51835.84	0.0000
Error	199	620.1683	3.1164236		
C Total	203	646790.01			

Root MSE	1.7653395	R-square	0.999041
Dep Mean	178.6645	Adj R-sq	0.999022
C.V.	0.988075		

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
X	1	0.975736	0.042711	22.84485	0.0000
Z	1	3.005385	0.008679	346.2721	0.0000
W	1	-5.029101	0.073149	-68.75141	0.0000
S	1	8.974106	0.029075	308.6485	0.0000

ב. יש עדות לכך. ג. יש עדות לכך.

2) ירד.

3) א. $H_0:\alpha=5\ ,\ \beta_s=3\beta_z$. ב. $Y_t-5=\beta_x X_t+\beta_z\left(Z_t+3S_t\right)+\beta_w W_t+u_t$.

ג. $WALD_{stat}=0.6145$. ד. מונה: 2 , מכנה: 199-.

ה. מקבלים.

4) בדיקה ע״י מבחן WALD ו-t: אין עדות לכך.

5) א. $PF=0.000$. ב. 92%. ג. $\hat{\alpha}=0.04195$. ד. לא.

ה. $H_0:\beta=0.70\ ,H_1:\beta>0.70$. ו. $t_{\hat{\beta}}=1.583$. ז. $WALD_{stat}=2.505$.

ח. i. נכון. ii. נכון. iii. לא נכון. iv. לא נכון.

ט. $H_0:\beta_2=\beta_3=0\ ,H_1:OTHERWISE$. י. $WALD_{stat}=68.32$.

יא. $H_0:\beta_3=2\cdot\beta_2\ ,\ \beta_1=0.6$.

יב. $W_t=SAVINGS_t+2\cdot NE_t\ ,\ Z_t=EXPENCE_t-0.6\cdot INCOME_t$.

מבוא לאקונומטריקה

פרק 7 - מבחן LM

תוכן העניינים

מבחן LM:

רקע:

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

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

השערות: $\begin{matrix} H_0 : \beta_3 = \beta_4 = 0 \\ H_1 : OTHERWISE \end{matrix}$.

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

- שימו לב כי:

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

שאלות:

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

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

UNRESTRICTED

Dependent variable: Y

Analysis of Variance

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

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

Parameter Estimates

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

RESTRICTED

Dependent variable: Y

Analysis of Variance

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

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

Parameter Estimates

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

רגרסיית עזר

Dependent variable :RES

Analysis of Variance

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

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

Parameter Estimates

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

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

ב. איזה מן המשתנים הנוספים רלוונטי למודל?

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

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

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

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

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

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

תשובות סופיות:

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

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

ג. i. U=R+עזר.

ii. $ESS_U = ESS_{ע}$.

iii. $ESS_R = TSS_{ע}$.

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

ד. $LM_{stat} = 43.489$.

ה. $WALD_{stat} = 26.962$.

מבוא לאקונומטריקה

פרק 8 - בעיות ספציפיקציה

תוכן העניינים

בעיות ספציפיקציה:

רקע:

טעויות ספציפיקציה הן טעויות בניסוח משוואת הרגרסיה.

1. הוספת משתנה לא רלוונטי:
למשל, המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$.
המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon$.
אם נקבל את H_0 במבחן t למובהקות β_3 נסיק כי המשתנה איננו רלוונטי ונאמוד את המודל מחדש הפעם ללא המשתנה השלישי. אולם, גם אם לא נוכל לאמוד מחדש, הימצאותו של משתנה שאיננו רלוונטי במודל הרגרסיה איננה פוגמת ברלוונטיות של המשתנים האחרים במודל ולא בתכונות החיוניות למבחני המובהקות שלהם.

2. השמטת משתנה רלוונטי:
למשל, המודל האמיתי: $Y_t = \alpha + \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon$.
המודל הנאמד (הטעותי): $Y_t = \alpha + \beta_1 x_{1t} + \varepsilon$.
בהיעדר x_2, בדיקות ההשערות לפרמטרים של המודל הטעותי אינן תקפות:

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

מבוא לאקונומטריקה

פרק 9 - תיאוריה מולטיקוליניאריות

תוכן העניינים

מולטיקוליניאריות:

רקע:

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

נבחין בין מולטיקוליניאריות מלאה לחלקית.

מולטיקוליניאריות מלאה:

מתאם מלא בין המשתנים המסבירים במודל.

הדבר קורה כאשר משתנה מסביר אחד הוא קומבינציה ליניארית מלאה של המשתנה המסביר השני: $x_1 = a + bx_2$ (x_1 הוא קומבינציה ליניארית מלאה של x_2)

מכאן ש: $r_{12} = 1$.

- שימו לב כי מדובר בטרנספורמציה ליניארית ולא בטרנספורמציה אחרת (למשל: $x_1 = x_2^2$), אז בהכרח: $r_{12} \neq 1$.

במצב של מולטיקוליניאריות מלאה אין כל השפעה של המשתנה האחד מעבר לשני.
מדוע זה בעייתי?
כיוון שלא ניתן לאמוד את המודל שכן אר״פ אינם מוגדרים.
פתרון: הורדת אחד המשתנים ואמידת המשוואה מחדש בלעדיו.

מולטיקוליניאריות חלקית:

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

מכיוון שיש מתאם גבוה בין המשתנים הב״ת לא נוכל לבודד באופן מלא את ההשפעה המדויקת של כל אחד מהם על ציוני המשתנה התלוי.
כל אחד מהמשתנים הב״ת ״יגזול״ מן ההשפעה הייחודית שיש למשתנה הב״ת השני על המשתנה התלוי, כך שבסופו של דבר, למרות שהמודל עם שני המשתנים הב״ת יהיה מובהק, התרומה הייחודית של כל משתנה ב״ת לניבוי התלוי לא תהיה מובהקת.

זיהוי מולטיקולינאריות חלקית:

1. כאשר קיימת סתירה בין התוצאה במבחן F למובהקות המודל (המודל מובהק) לבין מבחני t למובהקות השיפועים (אף אחד מן השיפועים איננו מובהק).
   הסתירה נוצרת כתוצאה מהגדלת השונות של כל אחד מהשיפועים בשל המתאם הגבוה בין הב"ת, באופן שלא מאפשר לדחות את השערת האפס למובהקות השיפועים: $S^2_{\hat{\beta}_1} = \frac{MSE}{SSX_1(1-r_{12})}$, $t = \frac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$.
2. רגישות לספציפיקציה – הורדת משתנה ב"ת שאיננו מובהק תהפוך משתנים ב"ת אחרים במודל למובהקים. אם אין בעיה של מולטיקולינאריות, הורדת משתנים ב"ת שאינם רלוונטיים מהמודל, לא אמורה להשפיע על מובהקותם של המשתנים הב"ת האחרים.
3. סימנים הפוכים – כאשר השיפועים של המשתנים הב"ת מקבלים סימנים הפוכים מכיוון ההשפעה שלהם על המשתנה התלוי. אם למשל, x_1 משפיע חיובית על Y ואילו x_2 משפיע שלילית על Y אבל הם יופיעו במשוואת הרגרסיה עם סימנים הפוכים ($\hat{\beta}_1$ שלילית ואילו $\hat{\beta}_2$ חיובית), יש לחשוד שקיימת בעיה.

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

פתרונות למולטיקולינאריות חלקית:

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

שלבי בדירת ההשערות:

1. מבצעים מבחן F לבדיקת מובהקות המודל.
2. במידה והמודל מובהק, מבצעים מבחן t למובהקות כל אחד מהשיפועים.
3. ביצוע מבחן WALD לבדיקת כל השיפועים שלא יצאו מובהקים:

א. אם מקבלים את H_0: אין סתירה בין מבחן WALD למבחני t - אין בעיה של מולטיקוליניאריות חלקית, נוריד את קבוצת המשתנים הלא רלוונטיים מהמודל.

ב. אם דוחים את H_0: יש סתירה בין מבחן WALD למבחני t - קיימת בעיה של מולטיקוליניאריות חלקית, יש להוריד מן המודל כל פעם משתנה אחד ולבצע מבחן WALD בלעדיו, עד שמזהים את המשתנה / משתנים שיש להוריד מהמודל.

מבוא לאקונומטריקה

פרק 10 - סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות

תוכן העניינים

סיכום ותרגול של בעיות ספציפיקציה ומולטיקוליניאריות:

רקע:

<table>
<tr><th>הבעיה</th><th>הגדרה</th><th>זיהוי</th><th>השלכות</th><th>פתרון</th></tr>
<tr>
<td>הוספת משתנה לא רלוונטי</td>
<td>המודל האמיתי:
$Y_t=\alpha+\beta_1x_{1t}+\varepsilon$
המודל הנאמד (הטעותי):
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$</td>
<td>קבלת H_0 במבחן t למובהקות β_2</td>
<td>ניתן לבצע בדיקת השערות
אר"פ $\left(\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2\right)$ חסרי הטיה
אומדי השונות $\left(S^2_{\hat{\alpha}},S^2_{\hat{\beta}_1},S^2_{\hat{\beta}_2}\right)$
חסרי הטיה</td>
<td>*הורדת המשתנה</td>
</tr>
<tr>
<td>השמטת משתנה רלוונטי</td>
<td>המודל האמיתי:
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$
המודל הנאמד (הטעותי):
$Y_t=\alpha+\beta_1x_{1t}+\varepsilon$</td>
<td>דחיית H_0 במבחן t למובהקות β_2</td>
<td>
<table>
<tr><th>בהיעדר x_2:</th><th>אומד ל- β_1</th><th>אומד ל- α</th><th>אומד לשונות הפרמטרים</th></tr>
<tr><td>$S_{12}=0$</td><td>חסר הטיה</td><td>מוטה
אלא אם:
$\bar{x}_2=0$</td><td>מוטה חיובית</td></tr>
<tr><td>$S_{12}\neq 0$</td><td>מוטה חיובית:
S_{12} ו- β_2 שווי סימן
מוטה שלילית:
S_{12} ו- β_2 מנוגדי סימן</td><td>מוטה</td><td>מוטה חיובית</td></tr>
</table>
לא ניתן לבצע בדיקת השערות
</td>
<td>הוספת המשתנה</td>
</tr>
<tr>
<td>מולטיקוליניאריות מלאה</td>
<td>מתאם מלא בין המשתנים המסבירים במודל
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$
כאשר: $r_{12}=\pm 1$</td>
<td>אם:
$x_1=a+bx_2$
אז: $r_{12}=1$</td>
<td>לא ניתן לבצע בדיקת השערות
אר"פ $\left(\hat{\alpha},\hat{\beta}_1,\hat{\beta}_2\right)$ בלתי מוגדרים.</td>
<td>הורדת אחד המשתנים</td>
</tr>
<tr>
<td>מולטיקוליניאריות חלקית</td>
<td>מתאם חזק בין המשתנים המסבירים במודל
$Y_t=\alpha+\beta_1x_{1t}+\beta_2x_{2t}+\varepsilon$
כאשר: $0.7<\left|r_{12}\right|<1$</td>
<td>א. סתירה בין מבחן F ל-t
ב. רגישות לספציפיקציה
ג. סימנים הפוכים</td>
<td>ניתן לבצע בדיקת השערות
אין פגיעה בתכונות אר"פ ושונותם</td>
<td>**הורדת אחד המשתנים
או
איחודם</td>
</tr>
</table>

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

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

שאלות:

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

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

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

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

כל האומדים חיוביים ומובהקים וקיים קשר שלילי בין גיל להשכלה.

א. $\hat{\beta}_1$ במשוואה (1) הוא:

i. אומד חסר הטיה.

ii. אומד מוטה שלילית.

iii. אומד מוטה חיובית.

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

ב. ניתן להשתמש במבחן t לבדיקת מובהקות השיפוע במשוואה (1). נכון/לא נכון/לא ניתן לדעת

ג. בנוסף למשתנים במשוואה השנייה, החליט החוקר להוסיף גם את משתנה הוותק, EXP_t. מכיוון שלא היו בידו נתונים על הוותק, החליט החוקר להעריכו עבור כל עובד על ידי הגיל של העובד פחות 24 שנים (מתוך ההנחה שהחיים המקצועיים מתחילים בגיל זה לערך).
להלן משוואה מס' 3:

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

חווה דעתך על המשוואה השלישית.

2) נתונות ארבע משוואות הרגרסיה הבאות (כאשר הסטיות במודל האמיתי מקיימות את הנחות הרגרסיה הקלאסיות):

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

2. $$\underset{}{Y_t = \alpha + \underset{(10.3)}{\beta_1 \cdot X_{1t}} + \underset{(19.8)}{\beta_2 \cdot X_{2t}} + U_t}.$$

3. $$Y_t = \alpha + \underset{(9.9)}{\beta_1 \cdot X_{1t}} + \underset{(17.3)}{\beta_2 \cdot X_{2t}} + \underset{(0.37)}{\beta_3 \cdot X_{3t}} + W_t.$$

4. $$Y_t = \alpha + \underset{(6.3)}{\beta_1 \cdot X_{1t}} + \Sigma_t.$$

(המספרים בסוגריים הם ערכי t של אומדני המקדמים).

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

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

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

ג. האומד של β_1 במשוואה (4) הינו חסר הטיה, אך אומד השונות של β_1 מוטה.

ד. האומדן $\hat{\beta}_1$ במשוואה (4) זהה ל-$\hat{\beta}_1$ במשוואה (2).

ה. השונות התיאורטית של האומדן $\hat{\beta}_1$ במשוואה (4) זהה לשונות התיאורטית של $\hat{\beta}_1$ במשוואה (2), אך אומדני השונויות שונים.

ו. האומד ל- α במשוואה (4) הינו חסר הטיה.

ז. האומד ל- α במשוואה (3) הינו חסר הטיה.

ח. R^2 של משוואה (2) גדול מ- R^2 של משוואה (3).

ט. $\bar{R}^2$ של משוואה (2) גדול מ- $\bar{R}^2$ של משוואה (3).

3) נתון המודל: $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t$.

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

א. בהנחה כי מתקיים: $Y_t = \alpha + \beta_1 \cdot X_{1t} + \beta_2 \cdot X_{2t} + U_t \qquad R^2 = 0.92$

(0.3) (0.5)

הערכים בסוגריים הם ערכי t.

למובהקות הבטות יש טעות במודל

כי המודל מובהק והמקדמים לא: נכון/לא נכון/לא ניתן לדעת

ב. בהנחה כי מתקיים: $X_{1t} - 2X_{2t} = 1$ לא ניתן לאמוד

את המודל בשיטת הריבועים הפחותים: נכון/לא נכון/לא ניתן לדעת

ג. בהנחה כי מתקיים: $X_{1t} = X_{2t}^2$ לא ניתן לאמוד

את המודל בשיטת הריבועים הפחותים: נכון/לא נכון/ לא ניתן לדעת

ד. הוכיחו תשובותיכם לסעיפים א' ו-ב'.

ה. בהנחה כי מתקיים: $r_{12} = 0.98$.

i. לא ניתן לאמוד את המודל בשיטת הריבועים הפחותים: נכון/לא נכון/ לא ניתן לדעת.

ii. איזו בעיה עלולה להיווצר במודל ומהן השלכותיה.

iii. בהנחה שהמודל יצא מובהק אולם הבטות אינן מובהקות וערכי t למובהקות הבטות הן כדלקמן: $t_{\hat{\beta}_2} = 1.45$, $t_{\hat{\beta}_1} = 1.31$, מה יהיה הפתרון הטוב ביותר, לדעתכם, לבעיה במודל (אליה התייחסתם בסעיף ii)?

1. להוריד את x_1.
2. להוריד את x_2.
3. להוריד את שני המשתנים.
4. להותיר את שני המשתנים.

תשובות סופיות:

1) א. ii. ב. לא נכון. ג. קיימת בעיית מולטיקוליניאריות מלאה.

2) א. לא נכון. ב. לא נכון. ג. נכון. ד. נכון. ה. נכון.
ו. לא נכון. ז. נכון. ח. לא נכון. ט. נכון.

3) א. לא נכון. ב. נכון. ג. לא נכון. ד. הוכחה. ה.i. לא נכון.
ii. מולטיקוליניאריות חלקית. iii. 4.

מבוא לאקונומטריקה

פרק 11 - משתנה דמי

תוכן העניינים

משתנה דמי:

רקע:

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

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

משתנה דמי לחותך:

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

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

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

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

משתנה דמי לשיפוע:

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

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

משתנה דמי לכל הפונקציה:

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

2 דרכים לבדיקה האם יש השפעה למשתנה האיכותי:

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

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

לבדיקת ההשערה נשתמש במבחן CHOW (הזהה למבחן WALD):

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

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

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

למרות התוצאות הזהות בשתי הדרכים, שיטת משתני הדמי עדיפה:

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

סיכום ביניים:

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

משתני דמי אם המשתנה האיכותי יכול לקבל יותר משני ערכים:

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

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

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

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

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

אם מדובר בחורף אז כל משתני הדמי יקבלו את הערך 0 ולכן החורף היא קבוצת הייחוס.
נניח שאנו רוצים לבדוק עונתיות במחירי הירקות:

$V_t=$ מדד מחירי הירקות.

$p_t=$ מדד המחירים לצרכן.

1. משתני דמי לחותך:
   הטענה: יש הבדל בין עונות השנה במחיר ההתחלתי של הירקות.
   המודל: $V_t=\alpha_0+\alpha_1 D_{1t}+\alpha_2 D_{2t}+\alpha_3 D_{3t}+\beta\cdot P_t+u_t$.
   כל עליה של יחידה אחת במדד המחירים לצרכן תעלה את מחירי הירקות ב- β. למחיר זה יתווסף α_0 בחורף, $\alpha_0+\alpha_1$ באביב, $\alpha_0+\alpha_2$ בקיץ ו- $\alpha_0+\alpha_3$ בסתיו.
   ניתן לראות כי: α_0 - החותך בקטגוריה שהושמטה, $\alpha_0+\alpha_1$ - החותך בקטגוריה i.

   בדיקת השערות:
   השערות: $H_0:\alpha_1=\alpha_2=\alpha_3=0$, $H_1:OTHERWISE$.
   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 H_0 במבחן הסטטיסטי של הסעיף הקודם, יש לבדוק מה מקור ההבדל בין החותכים על ידי מבחני t:

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

2. משתני דמי לשיפוע:

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

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

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.

בדיקת השערות:

השערות: $H_0: \beta_1 = \beta_2 = \beta_3 = 0$, $H_1: OTHERWISE$.

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

3. משתני דמי לכל הפונקציה:

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

המודל:

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

בדיקת השערות:

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

המבחן הסטטיסטי - מבחן WALD:

(U):

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

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

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

אם דוחים את H_0 יש להמשיך לבדוק באמצעות מבחני t:

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

משתני דמי עבור שני משתנים איכותיים:

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

1. הבדל בחותך ללא אינטראקציה:

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

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

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

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

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

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

3. $H_0 : \alpha_3 = 0$.

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

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

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

שאלות:

משתנה דמי לחותך:

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

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

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

המספרים בסוגריים הם טעויות התקן של מבחני המובהקות לפרמטרים.

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

ב. מה ההבדל בשכר ההתחלתי בין גברים לנשים?

ג. האם הבדל זה מובהק באוכלוסייה?

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

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

פונקציית רגרסיה המכילה משתנים איכותיים בלבד:

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

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

נתון: $S_{\hat{\alpha}_1} = 63$.

בדקו האם קיים הבדל מובהק בשכר הממוצע בין נשים וגברים?

משתנה דמי לשיפוע:

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

תוצאות האמידה נתונות להלן:

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

(68) (23) (25)

בדוק את ההשערה.

משתנה דמי לכל פונקציה:

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

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

כאשר:
NUM_t - מס' תאונות הדרכים הקטלניות בקטע כביש t בשנה.
$AVGD_t$ - נפח התנועה בקטע כביש t ליום באלפים.
$TYPE_t$ - משתנה דמי המקבל את הערך 1 כאשר הכביש מהיר, ו-0 כאשר הכביש לא מהיר.
תוצאות אמידת המשוואות מופיעות בהמשך השאלה.

א. בדקו את טענת החוקר בשתי דרכים שונות. ציינו איזה מן המשוואות רלוונטיות עבור כל דרך.
ב. חשבו את הערכים המספריים עבור אומדני משוואה (4).
ג. מהו האומדן הנקודתי למס' התאונות בכביש מהיר כאשר נפח התנועה עומד על ארבעת מכוניות ליום בקטע הכביש האמור?

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

ד. מהי השערת האפס לבדיקת הטענה (במונחי משוואה (4))?
ה. מהי הרגרסיה "תחת H_0" למבחן WALD?

משוואה (1) - כבישים מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	344
Number of Observations Used	344

Analysis of Variance

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

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

Parameter Estimates

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

משוואה (2) - כבישים לא מהירים בלבד:

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	410
Number of Observations Used	410

Analysis of Variance

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

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

Parameter Estimates

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

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

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

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

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

Parameter Estimates

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

משוואה (4):

The REG Procedure

Model: MODEL1

Dependent Variable: num num

Number of Observations Read	754
Number of Observations Used	754

Analysis of Variance

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

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

Parameter Estimates

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

משתנה איכותי עם יותר משתי קטגוריות:

5) ענה על הסעיפים הבאים:

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

i. מהי השערת האפס לבדיקת הטענה?

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

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

i. מהי השערת האפס לבדיקת הטענה?

ii. מהו המבחן הסטטיסטי המתאים? פרטו.

משתנה דמי עבור שני משתנים איכותיים:

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

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

$\ln(Y)$ - לוג השכר.

EXP - שנות ניסיון.

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

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

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

תוצאות אמידת משוואת הרגרסיה מוצגות בפלט להלן:

Analysis of Variance

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

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

Parameter Estimates

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

א. לפי המשוואה הניסיון זהה עבור שחורים ולבנים: נכון/לא נכון/ לא ניתן לדעת
ב. בדוק את הטענה כי בקרב אנשים בעלי השכלה נמוכה אין השפעה לגזע.
ג. בדוק את הטענה כי אין השפעות השכלה בקרב לבנים.
ד. מהי השערת האפס לבדיקת הטענה כי אין אינטראקציה בין גזע להשכלה?
ה. לבדיקת ההשערה של הסעיף הקודם בוצע מבחן WALD.
הרגרסיה המוגבלת תחת השערת האפס הינה:
$Z_0 = \gamma_0 + \gamma_1 Z_1 + \gamma_2 Z_2 + \gamma_3 Z_3 + \gamma_4 Z_4 + v$.
מהם ה-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').

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

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

כאשר:

S משתנה דמי: 1= עבור נשים, 0=גברים.

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

א. רשמו את הפונקציה לחישוב:

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

ii. תחזית לוג השכר ההתחלתי עבור נשים משכילות.

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

ב. רשמו את השערות האפס המתאימות לבדיקת הטענות הבאות:

i. אין השפעה של מגדר והשכלה על השכר.

ii. השפעת ההשכלה אינה תלויה במגדר.

iii. אין השפעות השכלה אצל גברים.

iv. אין הבדל בשיעורי התשואה לניסיון, בקרב הנשים.

תשובות סופיות:

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

2) יש עדות לכך.

3) יש עדות לכך.

4) א. יש עדות לכך, מבחן CHOW : 1 , 2 ו-3, משתנה דמי : 3 ו-4.
ב. $\hat{\alpha} = 0.14978$, $\hat{\beta}_1 = 1.40311$, $\hat{\beta}_2 = 0.002877$, $\hat{\beta}_3 = -0.008$.
ג. $NUM_t = 1.532398$. ד. $H_0 : \beta_2 + \beta_3 = 2 \cdot \beta_2$, $H_0 : \beta_3 = \beta_2$.
ה. $NUM_t = \alpha + \beta_1 \cdot TYPE_t + \beta_3 \cdot (AVGD_t + AVGD \cdot TYPE)_t + U_t$.

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

6) א. נכון. ב. יש עדות לכך. ג. יש עדות לכך.
ד. $H_0 : \alpha_2 = \alpha_1 - \alpha_3$ או $H_0 : \alpha_3 = \alpha_1 - \alpha_2$.
ה. $Z_0 = \ln(Y)_t$, $Z_1 = D_1 + D_3$, $Z_2 = D_2 - D_3$, $Z_3 = EXP_t$, $Z_4 = EXP_t^2$
ו. אין עדות לכך. ז. $\lambda_0 = \alpha_0$, $\lambda_1 = \alpha_2$, $\lambda_2 = \alpha_3$, $\lambda_3 = \alpha_1 - \alpha_2 - \alpha_3$.
ח. לא.

7) א.i. $\hat{\ln}(MWAGE) = \hat{\alpha}_0 + \hat{\beta}_0 \cdot 10$.
ii. $\hat{\ln}(MWAGE) = \hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2 + \hat{\alpha}_3$. iii. $EXP_t = \dfrac{-(\alpha_1 + \alpha_3)}{\beta_1 + \beta_3}$.
ב.i. $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \beta_1 = \beta_2 = \beta_3 = 0$. ii. $H_0 : \alpha_3 = \beta_3 = 0$.
iii. $H_0 : \alpha_2 = \beta_2 = 0$. iv. $H_0 : \beta_2 + \beta_3 = 0$.

מבוא לאקונומטריקה

פרק 12 - תיאוריה הפרת ההנחות קלאסיות

תוכן העניינים

הפרת ההנחות הקלאסיות:

רקע:

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

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

בכל אחד מן הנושאים נלמד:

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

מבוא לאקונומטריקה

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: $V(u_t)=\sigma_t^2$.

השלכות 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: $F_{stat}=\dfrac{ESS_1/T_1-K-1}{ESS_2/T_2-K-1}$ (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: $W_t = \dfrac{1}{\sqrt{Z_t}}$.

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

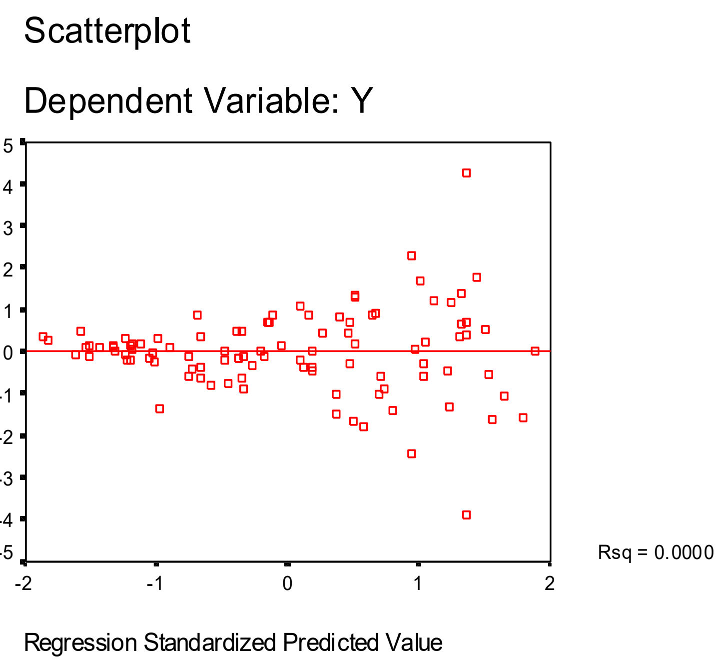
המשוואה המקורית:
$$Y_t=\alpha+\beta X_t+u_t$$
$$Y_tW_t=\alpha\cdot W_t+\beta\left(X_tW_t\right)+u_tW_t.$$

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: $VAR(U_t) = \dfrac{\sigma^2}{Z_t^2}$ (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.
$WHITE(LM)$.
RES^2.
$X1\_SQUARE$, $X1\_SQUARE^2$, $X1\_SQUARE * X2\_RENT$,
$PAPERS$, $X2\_RENT * PAPERS$, $X2\_RENT^2$, $X2\_RENT$.
$LM_{stat} = 1.73$.
אין עדות לכך.

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

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

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

מבוא לאקונומטריקה

פרק 14 - מתאם סדרתי

תוכן העניינים

מתאם סדרתי:

רקע:

מתאם סדרתי עוסק במצב שבו מופרת ההנחה (מס' 6) של אי תלות בין הטעויות: $\operatorname{cov}(u_t,u_s)=0$ ונוצרת 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. בנתוני סדרות עיתיות, מאחר ומדובר באותו הפרט הנמדד בזמנים שונים סביר שהטעויות בניבוי שלו תהיינה תלויות אחת בשנייה.

השלכות על אומדי הריבועים הפחותים (OLS):

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

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

מבנה המתאם הסדרתי:

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

כך ש:

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

4. ε_t מקיים את ההנחות הקלאסיות מאחר ומהווה סטייה מקרית לחלוטין (בניגוד ל- u_t) כך ש:

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

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

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

מלבד α ו- β נרצה לאמוד גם את ρ.

מתאם סדרתי מסדר שני:

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

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

מתאם סדרתי מסדר P:

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

u_t מושפע מתקופות שונות בעבר.

תכונות המתאם הסדרתי:

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

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

בנוסף לכך, התוחלת, השונות והשונות המשותפת של הטעויות:

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

מבחנים לזיהוי מתאם סדרתי:

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

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

$$0 - \rho > 0 - d_L - d_U - \rho = 0 - 4 - d_U - 4 - d_L - \rho < 0 - 4$$

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

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

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

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

למבחן DW יש שתי בעיות עיקריות:

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

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

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

מבחן LM:

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

המודל הלא מוגבל:

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

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

השלבים לביצוע המבחן:

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

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

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

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

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

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

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

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

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

החסרת משוואה (2) ממשוואה (1):

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

כדי לאמוד את הפרמטרים של רגרסיית ההפרשים נגדיר:

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

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

ε_t מקיים את כל ההנחות הקלאסיות ולכן שונות הרגרסיה וכן שונות הפרמטרים הנאמדים לא תהיה תלויה במקדם המתאם הסדרתי.
המשוואה "המתוקנת" אותה נאמוד: $Y^* = \alpha^* + \beta^* X_t^* + \varepsilon_t$.

לאחר אמידת משוואה זו ניתן לחלץ את האומדים של הפרמטרים המקוריים: $\hat{\beta}$, $\hat{\alpha}$.
מאחר ש- ρ איננו ידוע יש צורך לאמוד אותו.

אמידת ρ בשיטת קוקרן אורקוט:
שיטת קוקרן אורקוט לאמידת ρ היא שיטה איטראטיבית – מבוססת על חזרות של תהליך מסוים עד להתכנסות.
התהליך הממוחשב נקרא אוטורגרסיה (AUTOREGRESION) מסדר ראשון, שני, שלישי וכו' (תלוי בסדר המתאם הסדרתי). התיקון למתאם הסדרתי יתבצע על ידי הרצת רגרסיה עם משתנה AR(1) (אוטו רגרסיה מסדר ראשון), AR(1) ו-AR(2) (אם מניחים קיום אוטורגרסיה מסדר שני) וכו'. אם משתנה AR מובהק זו אינדיקציה שפתרנו את הבעיה של המודל המקורי.

שאלות:

תכונות המתאם הסדרתי:

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

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

מצאו את :

א. המתאם בין u_t ל- u_{t-1}.

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

ג. השונות σ_u^2.

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

מבחן DW:

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

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

כאשר :

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

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

תוצאות האמידה שהתקבלו :

Dependent variable: CLOSE

Analysis of Variance

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

Parameter Estimates

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

Durbin-Watson D 0.150

האם קיים מתאם סדרתי?

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

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

$\alpha = .05$

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

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

u_t - הפרעה מקרית קלאסית.

וידוע כי $T = 100$:

$u_t = 0.9u_{t-1}$

$d_L = 1.57$

$d_U = 1.65$

האם קיים מתאם סדרתי ברמת מובהקות של 5%?

מבחן LM:

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

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

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

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

תוצאות האמידה שהתקבלו:

Dependent variable: RES

Analysis of Variance

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

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

Parameter Estimates

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

א. האם קיים מתאם סדרתי?

ב. מהו ערכו של המתאם הסדרתי הנאמד?

ג. מהו כיוונו של המתאם הסדרתי באוכלוסייה?

תיקון המתאם הסדרתי:

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

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

Parameter Estimates

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

Durbin-Watson D 2.235

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

תרגול מסכם:

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

Depended Variable: RES

Parameter Estimates

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

Durbin-Watson D 1.954

נתון בנוסף כי: $R^2 = 0.479$.

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

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

Depended Variable:C

Parameter Estimates

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

ב. רשמו את המשוואה המתוקנת המשמשת לעריכת תחזיות.

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

א. נסח את ההשערה ובחן אותה בר״מ של 0.05.

החוקר מניח שיש מתאם סדרתי מסדר ראשון.
לאחר תיקון Cochrane-Orcutt התקבל: $\ln\hat{Y}_t = 7.3 - 0.2 \ln P_t + 0.4 \ln P_{t-1}$, $\hat{\rho} = 0.2$.
הניחו שהשבוע ובשבוע שעבר מחיר ממוצע של כרטיס היה 500$.
השבוע נמכרו 6,185 כרטיסים. בשבוע הבא צפוי מחיר של 400$.

ב. כמה כרטיסים יימכרו?

החוקר גם מנסה לקבוע האם בנתונים אלה קיים מתאם מסדר שני.

ג. רשמו את המשוואה הנוספת שעליו לאמוד.

במשוואה הנוספת התקבל מתאם מרובה השווה ל-0.12.

ד. מהי המסקנה בר״מ של 0.05?

תשובות סופיות:

1) א. $r_{(u_t,u_{t-1})}=0.9$. ב. $r_{(u_t,u_{t-4})}=0.6561$. ג. $\sigma_u^2=5.263$.
ד. $\sigma_u^2=1.19$, $r_{(u_t,u_{t-4})}=0.0256$, $r_{(u_t,u_{t-1})}=0.4$.

2) יש עדות לכך.

3) יש עדות לכך.

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

5) לא נכונה.

6) ראו סרטון.

7) א. קיומו של מתאם סדרתי מסדר שלישי בנתונים.
LM.
$$\begin{aligned}&H_0:\rho_1=\rho_2=\rho_3=0\\&H_1:OTHERWISE\end{aligned}.$$
$LM_{stat}=22.99$.
יש עדות לכך.
ב. $\hat{C}_t=-2103.53+0.719\cdot Y_t+0.707\cdot\hat{u}_{t-1}-0.0064\cdot\hat{u}_{t-2}-0.328\cdot\hat{u}_{t-3}$.

8) א. $\begin{aligned}&H_0:\rho=0\\&H_1:\rho\neq 0\end{aligned}$, יש עדות לכך. ב. $Y_{t+1}=5,568$.
ג. $u_t=\gamma_0+\gamma_1\ln P_t+\gamma_2\ln P_{t-1}+\gamma_3u_{t-1}+\gamma_4u_{t-2}+\omega_t$. ד. אין עדות לכך.

מבוא לאקונומטריקה

פרק 15 - סיכום מתאם סדרתי והטרוקדסטיות

תוכן העניינים

סיכום מתאם סדרתי והטרוקדסטיות:

רקע:

	מתאם סדרתי	הטרוסקדסטיות
המשוואה העיקרית של המודל	למשל: $Y_t = \alpha + \beta X_t + u_t$	
ההנחה הקלאסית המופרת	$\operatorname{cov}(u_t, u_{t-s}) = 0$	$V(u_t) = \sigma^2$
המצב לאחר ההפרה	$\operatorname{cov}(u_t, u_{t-s}) \neq 0$	$V(u_t) = \sigma_t^2$
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

מבוא לאקונומטריקה

פרק 16 - משוואות סימולטניות

תוכן העניינים

משוואות סימולטניות:

רקע:

עוסקות בהפרת ההנחה של אי תלות בין המב"ת לטעויות בניבוי: $\text{cov}(x,u)=0$.

ה- X ים במשוואה נחשבו משתנים אקסוגניים – משפיעים על Y אך לא מושפעים ממנו בחזרה לעומת זאת משתנים אנדוגניים – משפיעים על Y אך גם מושפעים ממנו בחזרה. מאחר ומשתנים אלו הם גם מסבירים וגם מוסברים, הם נחשבים כמשתנים מקריים, המתואמים עם הטעויות במודל: $\text{cov}(x,u)\neq 0$.

משוואות המבנה (משוואות סימולטניות):

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

השלכות על אר"פ:

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

הצורה המצומצמת של מודל עם משוואות סימולטניות:

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

תכונות המשוואות מהצורה המצומצמת :

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

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

בתהליך החילוץ של הפרמטרים המבניים ייתכנו 3 מצבים :

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

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

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

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

שיטות לפתרון משוואות סימולטניות:

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

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

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

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

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

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

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

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

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

סיכום תוצאות אמידה של משוואות סימולטניות:

מסי 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 ומסי 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: $\operatorname{cov}(Y_{t-1}, u_t \neq 0)$, אז Y_{t-1} אנדוגני.
אם אין מתאם סדרתי: $\operatorname{cov}(Y_{t-1}, u_t = 0)$, אז Y_{t-1} אקסוגני.

מבחנים סטטיסטיים לבחינת אנדוגניות ולחוזק משתנה עזר:

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

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

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

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

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

שאלות:

זיהוי משוואות המבנה:

1) חוקר רצה לאמוד את פונקציית הביקוש ואת פונקציית ההיצע לתות שדה.
הוא אסף נתונים עבור 30 תקופות:
Pt - מחיר קופסא בש"ח בתקופה t.
Qt - כמות נקנית בק"ג בתקופה t.
Zt - מחיר פרי תחליפי ב-₪ בתקופה t.
INCOMEt - הכנסת הצרכנים באלפי ₪ בתקופה t.
Lt - מחיר שעת עבודה ב-₪ בתקופה t.

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

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

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

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

להלן הפלטים:

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

Dependent variable: l_Q

Instruments: l_L

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

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 2.67757

Standard error of residuals = 0.32091

Unadjusted R-squared = 0.222881

Adjusted R-squared = 0.133214

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

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

Dependent variable: l_Q

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

Mean of dependent variable = 2.8776

Standard deviation of dep. var. = 0.300322

Sum of squared residuals = 0.834357

Standard error of residuals = 0.179139

Unadjusted R-squared = 0.681008

Adjusted R-squared = 0.644201

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

Log-likelihood = 11.1662

(Log-likelihood for Q = -75.1618)

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

Dependent variable: l_P

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

Mean of dependent variable = 2.99427

Standard deviation of dep. var. = 0.314761

Sum of squared residuals = 0.369832

Standard error of residuals = 0.119266

Unadjusted R-squared = 0.871281

Adjusted R-squared = 0.856428

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

Log-likelihood = 23.3704

(Log-likelihood for P = -66.4576)

שיטת ILS:

2) נניח שאנו מתכוונים לאמוד את המשוואות:

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

כאשר:

C_t - הוצאות לתצרוכת פרטית.

Y_t - הכנסה לאומית.

u_t - הפרעה אקראית.

א. מהי הבעיה באמידת המשוואות בשיטת הריבועים הפחותים? מהן תכונות אר"פ?
ב. האם המשוואות מזוהות?
ג. אמדו את מערכת המשוואות בצורתה המצומצמת באופן ידני.
ד. מהו הפתרון של המשוואות המצומצמות בשיטת ILS?

להלן תוצאות אמידת מערכת המשוואות בצורה המצומצמת:

Dependent Variable: C

Parameter Estimates

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

Dependent Variable: Y

Parameter Estimates

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

ה. חשבו את האומדים המבניים.

שיטת 2SLS:

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

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

כאשר:

C_t - הוצאות לתצרוכת פרטית.

Y_t - הכנסה לאומית.

u_t - הפרעה אקראית.

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

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

להלן תוצאות האמידה בשיטת 2 השלבים:

Dependent variable: C

Parameter Estimates

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

Dependent variable: Y

Parameter Estimates

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

4) לפניך המודל הסימולטני הבא:

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

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

P_t - מחיר המוצר בתקופה t.

Q_t^D - כמות מבוקשת בתקופה t.

Q_t^S - כמות מוצעת בתקופה t.

Z_t - מחיר המוצר התחלפי בתקופה t.

Z_t הוא משתנה אקסוגני.

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

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

$$\hat{Q}_t = 2 + 3Z_t$$
$$\hat{P}_t = 1 + 4Z_t$$.

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

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

שיטת IV:

5) נתונות המשוואות הבאות:

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

2. $X_t = \beta_0 + \beta_1 \cdot Y_t + \beta_2 \cdot Z_{1t} + \omega_t$.

נתון כי: T_t , X_t משתנים אנדוגניים ו- Z_{2t} , Z_{1t} משתנים אקסגוניים.

חוו דעתכם על כל אחת מהטענות הבאות, והסבירו:

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

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

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

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

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

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

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

מבחן האוזמן:

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

מבחן לחוזק IV:

7) נתונה מערכת המשוואות הסימולטניות הבאות:

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

כאשר: X_1, X_2, X_3 הינם משתנים 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?

תרגילים מסכמים:

1) נתונות המשוואות הבאות:

1. $Y_t = \alpha_0 + \alpha_1 X_t + \alpha_2 Z_{1t} + \alpha_3 Z_{2t} + \alpha_4 Z_{3t} + \alpha_5 Z_{4t} + u_t$.

2. $X_t = \beta_0 + \beta_1 Y_t + \beta_2 Z_{1t} + \beta_3 Z_{2t} + \beta_4 Z_{5t} + v_t$.

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

א. אמידת כל אחת מהמשוואות תניב אומדים:

i. מוטים — נכון/לא נכון/לא ניתן לדעת.

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

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

משוואה 2 — מזוהה בדיוק/ מזוהה ביתר/ בלתי מזוהה

ג. חווה דעתך על הטענות הבאות:

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

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

ד. משוואות הצורה המצומצמת הן:

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

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

נכון/לא נכון/לא ניתן לדעת.

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

ו. להלן רשימה של משתני עזר פוטנציאליים:

i. Z_5.

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

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

iv. $Z_3 + Z_4$.

v. $3Z_3 + 4Z_4$.

vi. $3Z_3 + 3Z_4$.

vii. Z_1.

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

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

i. i ו-ii.

ii. iv ו-vi.

iii. v ו-vi.

iv. iv ו-v.

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

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

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

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

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

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

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

כאשר:

$HOURS$ - היצע העבודה בשעות.

$WAGE$ - שכר לשעה.

$EDUC$ - מספר שנות הלימוד.

AGE - גיל.

$KIDSL6$ - מספר הילדים בבית מתחת לגיל 6.

$KIDS618$ - מספר הילדים בגיל 6-18.

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

א. מהם הסימנים שתצפו לקבל בכל אחד מהמקדמים?

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

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

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

3) נתונה מערכת המשוואות הסימולטניות הבאה:

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

כאשר X_1 , X_2 , X_3 הינם משתנים אקסוגניים.

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

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

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

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

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

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

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

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

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

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

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

תשובות סופיות:

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

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

$Q_t^D = Q_t^S$

משוואת הביקוש מזוהה במדויק.

משוואת ההיצע מזוהה ביתר.

ב. פונקציית הביקוש, התוצאות מתיישבות.

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

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

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

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

3) א. מוטים אך עקיבים ויעילים במדגמים גדולים.

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

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

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

ד. 0.75.

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

ו. נכון. ז. לא נכון.

6) ראו סרטון.

7) ראו סרטון.

תרגילים מסכמים:

1) א.i. נכון. ii. לא נכון.

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

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

ו.i. 1. ii. 1. iii. 2. iv. 2. v. 2.

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

ח. כן. ט. אקסוגניים. י. לא. יא. כן.

יב. משוואה 1 מזוהה בדיוק ומשוואה 2 מזוהה ביתר.

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

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

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

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

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

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

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

ה. ראו סרטון.

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

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