2009

אקונומטריקה

ד"ר חמי גוטליבובסקי

סמסטר א' תש"ע

סיכום: דביר צנוע

הקדמה

הדפים שלפניכם מהווים סיכום של קורס מבוא לאקונומטריקה, אשר הועבר באוניברסיטת תל-אביב ע"י ד"ר חמי גוטליבובסקי בסמסטר א' תש"ע. הסיכום הוקלד בידיי במהלך ההרצאות, ולא אושר על ידי גורמים אקדמיים באופן כללי או ד"ר גוטליבובסקי בפרט, ויש לקחת זאת בחשבון בעת הלמידה. הסיכום הינו כלי עזר בלבד ולא מחליף למידה פעילה וניהול מחברת מסודרת במהלך הקורס.

אם מצאתם את השימוש במחברת מועיל אך זקוקים לעזרה נוספת, אני עומד לרשותכם, ומעביר **שיעורים פרטיים** בתשלום – הכנה למבחן, עזרה בשיעורי בית והשלמת פערים.

ליצירת קשר תוכלו לפנות אליי בפייסבוק, בדואר אלקטרוני dvirsmail@gmail.com או בטלפון 054-2209558.

זכרו: כלכלה זה כיף בסך הכל, אז תנו בראש.

אז בהצלחה (ואתם תזדקקו לה),

דביר צנוע

נ"ב: כמובן שאי אפשר ללמוד בלי מוסיקה, אז קבלו המלצה:
http://www.youtube.com/watch?v=IJM68s5UAis

תוכן עניינים

חזרה על חוקי סטטיסטיקה חשובים 5
הבנת התנהגות של יחידות כלכליות 6
אומד הריבועים הפחותים 6
התוחלת של $\hat{\beta}$ 8
השונות של $\hat{\beta}$ 9
הוכחת משפט גאוס מרקוב 11
שימוש באומדים 14
רווח סמך ל-β 14
מבחני השערות 15
האומד של α 17
השונות המשותפת של $\hat{\alpha}$ ו-$\hat{\beta}$ 19
מודלים שאינם לינאריים בפרמטרים 20
חיזוי 21
השפעת שינוי סקאלה על האומד 23
מדד לאיכות הרגרסיה R^2 23
הקשר בין R^2 למקדם המתאם 24
שינויים ב-R^2 25
השוואת שתי רגרסיות 25
רגרסיה מרובת משתנים 26
מולטי קוליניאריות מושלמת 30
מבחן F 31
רגרסיה עם משתנה ריבועי 34
טעויות ספציפיקציה 35
צורות שונות של משוואת הרגרסיה 37
משתנים איכותיים 37
מודל הכולל משתנים איכותיים עם משתנים כמותיים 40

ריבוי משתנים איכותיים 41
הסרת ההנחות 42
משתנה מסביר מקרי 42
תלות בין x ל-u 42
עקיבות 44
מתאם בין ההפרעות האקראיות 45
משתנה מסביר מקרי יחד עם מתאם סדרתי בהפרעות 50
שונות שונה 51
בעית האנדוגניות של המשתנה המסביר 55
אומד משתני עזר IV 56
שיטת 2SLS 59
שיטת הריבועים הפחותים העקיפה (ILS) 60
משוואות מזוהות 60

חזרה על חוקי סטטיסטיקה חשובים

$$E(x+c)=c+E(x)$$

$$E(x+y)=E(x)+E(y)$$

$$E(c\cdot x)=cE(x)$$

משתנים תלויים: צפיה באחד מהם משפרת את היכולת לנבא את השני.

x ו-y ב"ת אמ"מ $E(x\backslash y)=E(x)$

אם x ו-y ב"ת אזי cov(x,y)=0 (ההיפך אינו מתחייב)

$$var(x)=E\big(x-E(x)\big)^2=\frac{\sum(x_i-\bar{x})^2}{n}$$

זו היתה נוסחה לחישוב **השונות במדגם**. לחישוב **אומדן לשונות האוכלוסיה**, מחלקים ב-n-1 במקום ב-n.

$$var(x+c)=var(x)$$

$$var(c\cdot x)=c^2\cdot var(x)$$

$$var(x+y)=var(x)+var(y)+2cov(x,y)$$

$$var(x-y)=var(x)+var(y)-2cov(x,y)$$

$$cov(x,y)=E\Big[\big(x-E(x)\big)\big(y-E(y)\big)\Big]=\frac{\sum(x_i-\bar{x})(y_i-\bar{y})}{n}$$

$$cov(c\cdot x,y)=c\cdot cov(x,y)$$

$$cov(c\pm x,y)=cov(x,y)$$

מדד COV נותן לנו את **עצמת** המתאם בין המשתנים, אבל לא את **כיוונה**.

$$cov(x,y,z)=var(x)+var(y)+var(z)+2cov(x,y)+2cov(y,z)+2cov(x,z)$$

x ו-y ב"ת אמ"מ $E(x\cdot y)=E(x)\cdot E(y)$

t– התפלגות נורמלית כאשר השונות לא ידועה.

טרנספורמציה ליניארית של משתנה המתפלג נורמלית, מתפלגת אף היא נורמלית.

מקדם המתאם ρ:

$$-1\le\frac{cov(x,y)}{\sigma_x\sigma_y}\le 1$$

הבנת התנהגות של יחידות כלכליות

y – המשתנה הנבדק / המוסבר, x – המשתנה המסביר.

אנחנו מניחים כי קיים קשר ליניארי בין המשתנים:

$$y = \alpha + \beta x_i + \epsilon_i$$

בכלכלה קיימים שלושה סוגי מדגמים:

<u>חתך אורך</u> – בדיקה של אותו משתנה לאורך זמן. האינדקס מסומן ע"י t.

<u>חתך רוחב</u> – בדיקה של משתנים שונים ביח' זמן אחת (או בלי מימד של זמן). האינדקס מסומן ע"י i.

<u>מדגם פנל</u> – דגימה של קבוצה על פני זמן.

אנחנו תמיד נניח כי לכל האנשים יש את אותו α ו-β. כדי למצוא את α̃ ואת β̃, כלומר, האומדים ל-α ו-β האמיתיים, ניעזר בשיטה שנקראת ריבועים פחותים (OLS).

אומד הריבועים הפחותים

בשיטה זו נרצה למצוא קו העובר בין התצפיות, הממזער את ריבועי המרחק (הטעות) בין כל תצפית אל הקו. את זה נעשה באמצעות פתרון בעית מינימום:

$$min\sum\left(y_i - (\tilde{\alpha} + \tilde{\beta}x_i)\right)^2$$

$$(I)\,\frac{\partial}{\partial\tilde{\alpha}}:-2\sum\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = 0$$

$$(II)\,\frac{\partial}{\partial\tilde{\beta}}:-2\sum\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)\cdot x_i = 0$$

שתי משוואות תנאי ראשון אלה נקראות גם **המשוואות הנורמליות**. תפקיד המשוואות האלה הוא "לסחוט" מ-Y את כל המידע האפשרי, ולהעמיס אותו על השיפוע β. כל מה שלא ניתן להסביר, מועמס על החותך α. ניתן לסמן את שגיאת האומד (ההפרש בין y האמיתי ל-ŷ שנאמד ע"י α̂ ו-β̂) כך:

$$u_i = y_i - \hat{\alpha} - \hat{\beta}x_i$$

ואז ניתן להבין את משמעות המשוואות הנורמליות, כלומר, את התנאים שאנו כופים על האומד שלנו:

$$\boldsymbol{\sum u_i = 0}$$

$$\boldsymbol{\sum u_i x_i = 0}$$

משתי המשוואות הנורמליות אפשר גם להראות שאין מתאם בין u ל-x:

$$\sum(x_i - \bar{x})\left(\hat{u}_i - \bar{\hat{u}}\right) = 0 \Rightarrow \boldsymbol{I\cdot cov(x_i, \hat{u}_i) = 0}$$

נמשיך לפתח את תנאי הקיצון:

$$(I) - 2\sum\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right) = 0$$

$$\sum y_i = \sum\hat{\alpha} + \hat{\beta}\sum x_i$$

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

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

כלומר, מתנאי סדר ראשון, נובע שהרגרסיה עוברת דרך נקודת הממוצעים.נמשיך לתנאי השני:

$$(II) - 2\sum\left(y_i - \hat{\alpha} - \hat{\beta}x_i\right)\cdot x_i = 0$$

$$\sum\left(y_i - \bar{y} + \hat{\beta}\bar{x} - \hat{\beta}x_i\right)\cdot x_i = 0$$

$$\sum(y_i - \bar{y})x_i = \hat{\beta}\sum(x_i - \bar{x})x_i$$

$$\hat{\beta} = \frac{\sum(y_i - \bar{y})x_i}{\sum(x_i - \bar{x})x_i} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})(x_i - \bar{x})}$$

$$\boldsymbol{\hat{\beta} = \frac{cov(x, y)}{var(x)}}$$

אם כן, פתרון בעית ישר הריבועים הפחותים:

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\beta} = \frac{\sum(y_i - \bar{y})x_i}{\sum(x_i - \bar{x})x_i} = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sum(x_i - \bar{x})(x_i - \bar{x})} = \frac{cov(x, y)}{var(x)}$$

מכאן משתמע שהקשר בין X ל-Y תלוי בשונות המשותפת. הבעיה, שהשונות המשותפת לא מציינת את כיוון הקשר. כלומר, אנחנו לא יכולים לדעת אם X השפיע על Y או להיפך.

- *קשר שימושי:*

$$\sum(y_i - \bar{y})x_i = \sum(y_i - \bar{y})(x_i - \bar{x}) = \sum(x_i - \bar{x})y_i$$

השאלות הנשאלות:

1. האם $\hat{\beta}$ הינו אומד חסר הטיה ל-β?
2. מה השונות של $\hat{\beta}$?

התוחלת של $\hat{\beta}$

נתחיל בבדיקה של השאלה הראשונה, כלומר, האם $\mathrm{E}(\hat{\beta}) = \beta$ (חוסר הטיה).

נחזור ל-$\hat{\beta}$ ונבטא אותו כך:

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}$$

במקום y_i ניתן להציב את הנחת העבודה המקורית שלנו, הכוללת את α ו-β האמיתיים:

$$y_i = \alpha + \beta x_i + u_i$$

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})(\alpha + \beta x_i + u_i)}{\sum(x_i - \bar{x})^2}$$

נפרק את המכפלה שנוצרה לשלושה מחוברים, ונראה מה קורה.

$$\sum(x_i - \bar{x})\alpha = 0$$

(המכנה לא רלוונטי כיוון שבכל מקרה הביטוי מתאפס)

$$\frac{\sum(x_i - \bar{x})\beta x_i}{\sum(x_i - \bar{x})^2} = \beta\frac{\sum(x_i - \bar{x})(x_i - \bar{x})}{\sum(x_i - \bar{x})^2} = \beta$$

(את β הוצאנו החוצה כי היא קבועה, וחיסרנו $\bar{x}$ בהסתמך על העיקרון שהוכחנו קודם)

כתוצאה מהפיתוחים הנ"ל קיבלנו ביטוי חדש ל-$\hat{\beta}$:

$$\hat{\beta} = \beta + \frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}$$

למעשה, ההפרעות האקראיות u_i מפריעות ל-$\hat{\beta}$ להיות שווה ל-β. לכן, במדגם יחיד, בטוח ש-$\beta \neq \hat{\beta}$. נבדוק מה קורה כשעוברים ממדגם יחיד למדגמים מרובים, כלומר, מה התוחלת של $\hat{\beta}$.

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}\left[\frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2}\right] = \mathrm{E}\left[\frac{\sum(x_i - \bar{x})(\alpha + \beta x_i + u_i)}{\sum(x_i - \bar{x})^2}\right]$$

תוחלת של סכום שווה לסכום התוחלות, לכן בדיוק כמו קודם, נפרק את המכפלה למחובריה:

$$\mathrm{E}\left[\frac{\sum(x_i - \bar{x})\alpha}{\sum(x_i - \bar{x})^2}\right] = \alpha\mathrm{E}\left[\frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}\right] = 0$$

$$\mathrm{E}\left[\frac{\sum(x_i - \bar{x})\beta x_i}{\sum(x_i - \bar{x})^2}\right] = \beta\mathrm{E}\left[\frac{\sum(x_i - \bar{x})x_i}{\sum(x_i - \bar{x})^2}\right] = \beta$$

ולכן, התוחלת של $\hat{\beta}$:

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}\right]$$

תזכורת – אנו רוצים להוכיח ש-$E(\hat{\beta}) = \beta$. כדי להמשיך, עלינו להניח שתי הנחות בשלב זה:

1. $\mathbf{x_i}$ **לא מקרי**, ולכן אפשר להוציא אותו מהתוחלת. זו הנחה סבירה – X יכול להיות טמפרטורה או כל פרמטר אחר שקובע החוקר. תוצאת ההנחה הזו:

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}\right] = \beta + \frac{\sum(x_i - \bar{x})}{\sum(x_i - \bar{x})^2}E(u_i)$$

2. $\mathbf{E(u_i) = 0}$. זאת כיוון ש-α תופסת את תוחלת ההזזה הנובעת מ-u. ההנחה הזו כוללת גם הנחה סמויה, והיא שלכל בני האדם יש את אותה התוחלת להפרעה אקראית. אם מוסיפים גם את ההנחה הזו, ניתן להסיק ש-$\hat{\beta}$ חסר הטיה.

יתכן כי שתי ההנחות הללו לא מתקיימות. ניתן להניח, לחילופין, כי $\mathrm{cov}(x,u) = 0$. זו לא הנחה מובנת מאליה, ויש לבחון את המשמעות הכלכלית של המשתנים השונים. נחזור לשלב שלפני ההנחות ונמשיך משם:

$$E(\hat{\beta}) = \beta + E\left[\frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}\right] = \beta + E\left[\frac{\sum(x_i - \bar{x})(u_i - \bar{u})}{\sum(x_i - \bar{x})^2}\right] = \beta + E\left[\frac{\mathrm{cov}(x,u)}{\mathrm{var}(x)}\right] = \beta$$

ובכן, הוכחנו כי $\hat{\beta}$ חסר הטיה.

המטרה כעת היא למצוא את השונות של אומד הריבועים הפחותים, ולהוכיח את משפט גאוס מרקוב, הטוען כי אומד הריבועים הפחותים הוא האומד הטוב ביותר תחת ההנחות.

השונות של $\hat{\beta}$

אם כך, קיים לנו מודל הטוען $y_i = \alpha + \beta x_i + u_i$ ואומד ל-β :

$$\hat{\beta} = \frac{\Sigma(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \beta + \frac{\Sigma(x_i - \bar{x})u_i}{\Sigma(x_i - \bar{x})^2}$$

ברור כי $\hat{\beta}$ ישתנה ממדגם למדגם. המטרה שלנו היא למצוא את השונות של האומד, ולשאוף למינימום שונות, כמובן בהינתן ההוכחה על כך שהאומד גם חסר הטיה.

נתחיל מחישוב נוסחת השונות:

$$\mathrm{var}(\hat{\beta}) = E\left(\hat{\beta} - \underbrace{E(\hat{\beta})}_{\beta}\right)^2 = E\left(\frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}\right)^2$$

הנחנו ש-X משתנה לא מקרי ולכן אפשר להוציא אותו החוצה, כמובן בריבוע.

$$= \frac{1}{(\sum(x_i - \bar{x})^2)^2}E(\sum(x_i - \bar{x})u_i)^2$$

נטפל כעת רק בחישוב התוחלת, ולצורך הפשטות נעשה זאת עם שתי תצפיות בלבד:

$$\mathrm{E}(\sum(\mathrm{x_i}-\bar{\mathrm{x}})\mathrm{u_i})^2=\mathrm{E}\big((\mathrm{x_1}-\bar{\mathrm{x}})\mathrm{u_1}+(\mathrm{x_2}-\bar{\mathrm{x}})\mathrm{u_2}\big)^2$$
$$=\mathrm{E}\big((\mathrm{x_1}-\bar{\mathrm{x}})\mathrm{u_1}\big)^2+\mathrm{E}\big((\mathrm{x_2}-\bar{\mathrm{x}})\mathrm{u_2}\big)^2+2(\mathrm{x_1}-\bar{\mathrm{x}})(\mathrm{x_2}-\bar{\mathrm{x}})\mathrm{E}(\mathrm{u_1u_2})$$

הנחנו על התוחלת של ההפרעה האקראית שהיא שווה 0, ולכן מותר לנו לעשות את הצעד הבא:

$$\mathrm{E}(\mathrm{u_1u_2})=\mathrm{E}\big[\big(\mathrm{u_1}-\mathrm{E}(\mathrm{u_1})\big)\big(\mathrm{u_2}-\mathrm{E}(\mathrm{u_2})\big)\big]=\mathrm{cov}(\mathrm{u_1},\mathrm{u_2})$$

הנחה שלישית: אין מתאם בהפרעה של שתי תצפיות שונות. כלומר, $\mathrm{cov}(\mathrm{u_i},\mathrm{u_j})=0,\forall \mathrm{i}\neq \mathrm{j}$.

אחרי הצבה, אנחנו מקבלים שהסכום יצא מהסוגריים:

$$\mathrm{var}(\hat{\beta})=\frac{1}{(\sum(\mathrm{x_i}-\bar{\mathrm{x}})^2)^2}\mathrm{E}\Big[\sum\big((\mathrm{x_i}-\bar{\mathrm{x}})\mathrm{u_i}\big)^2\Big]$$
$$=\frac{1}{(\sum(\mathrm{x_i}-\bar{\mathrm{x}})^2)^2}\sum\left((\mathrm{x_i}-\bar{\mathrm{x}})^2\underbrace{\mathrm{E}\big(\mathrm{u_i}-\mathrm{E}(\mathrm{u_i})\big)^2}_{\sigma_\mathrm{u}^2}\right)$$

החלק האחרון בנוסחה, הוא למעשה השונות של u_i. את $E(u_i)$ יכולנו לחסר כי כזכור, הוא שווה 0.

הנחה רביעית: $\mathbf{var(u_i)=var(u)}$. להניח שהפיזור של ההפרעות של כל הפרטים הוא זהה היא הנחה מתקבלת על הדעת, אבל לא מתחייבת.

אם $\mathrm{E}\big(\mathrm{u_i}-\mathrm{E}(\mathrm{u_i})\big)^2$ הוא השונות של u, היא זהה לכולם, וניתן להוציא אותה מהסכום.

$$=\frac{\sigma_\mathrm{u}^2\sum(\mathrm{x_i}-\bar{\mathrm{x}})^2}{(\sum(\mathrm{x_i}-\bar{\mathrm{x}})^2)^2}$$

ולכן,

$$\mathbf{var}\hat{\boldsymbol{\beta}}=\frac{\boldsymbol{\sigma}_\mathbf{u}^\mathbf{2}}{\sum_{\mathbf{i=1}}^{\mathbf{I}}(\mathbf{x_i}-\bar{\mathbf{x}})^\mathbf{2}}$$

לנוסחה הזו יש שני חלקים, חלק טריוויאלי יותר וחלק טריוויאלי פחות.

1. הגדלת המדגם, מקטינה את השונות. אבל זה די טריוויאלי.
2. X אינו מקרי. אם החוקר יבחר X עם שונות קטנה, השונות של β תגדל.לכן על החוקר לבחור Xים עם שונות גדולה. זה הגיוני, כיוון שאם בוחרים Xים קרובים, יהיה מאוד קשה לוודא את הקשר בין X ל-Y בצורה אמינה.

במידה ואנו איננו יודעים את השונות של u (וזהו המצב במציאות), הנוסחה הזו בלתי פתירה. כדי לפתור בעיה זו, אנו צריכים למצוא **אומד** ל-$\mathrm{var}\hat{\beta}$, באמצעות מציאת אח"ה ל- σ_u^2 (לא נוכיח את חוסר ההטיה של הביטוי):

$$\hat{\sigma}_\mathrm{u}^2=\frac{\sum_{i=1}^{I}(\hat{u}_i)^2}{I-2}$$

כאשר

$$\hat{u}_i = y_i - (\hat{\alpha} + \hat{\beta} x)$$

כיוון שאיננו יודעים את α ואת β האמיתיים, אנחנו משתמשים בנאמדים על תוצאות המדגם כדי לחשב אומדן ל-u. המכנה I-2 מוסבר ע"י כך שאיבדנו שתי דרגות חופש במהלך הדרך, כשגזרנו את בעית המינימום המקורית, והשווינו את הנגזרות החלקיות ל-0. אחרי הצבת האומד, מקבלים:

$$\widehat{\mathbf{var}}\hat{\boldsymbol{\beta}} = \frac{\hat{\boldsymbol{\sigma}}_{\mathbf{u}}^{2}}{\sum_{\mathbf{i=1}}^{\mathbf{I}} (\mathbf{x_i} - \bar{\mathbf{x}})^2}$$

הוכחת משפט גאוס מרקוב

עד עתה הנחנו ארבע הנחות:

1. x_i לא מקרי.
2. $E(u_i) = 0$
3. $\mathrm{cov}(\mathrm{u_i}, \mathrm{u_j}) = 0$
4. $\mathrm{var}(\mathrm{u_i}) = \mathrm{var}(\mathrm{u})$

ישנה עוד תכונה של המודל שעדיין לא דיברנו עליה – y מופיע בצורה לינארית במודל $y_i = \alpha + \beta x_i + u_i$.

עבור ההוכחה שלנו, אנחנו **דורשים מודל לינארי בפרמטרים.**

מבין כל האומדים הלינאריים (ב-Y) שאינם מוטים, אומד הריבועים הפחותים הינו בעל השונות המינימלית, בכפוף לארבעת ההנחות התחיליות.

המודל שנבחר חייב להיות לינארי בפרמטרים וגם לינארי ב-Y.

נניח אומד כלשהו שהינו ליניארי ב-Y:

$$\tilde{\beta} = \sum \mathrm{k_i y_i}$$

כאשר k הוא ערך מחושב כלשהו שאינו מכיל את Y.

נביט לרגע בנוסחת האומד המקורי:

$$\hat{\beta} = \frac{\sum (\mathbf{x_i} - \bar{\mathbf{x}}) \mathrm{y_i}}{\sum (\mathbf{x_i} - \bar{\mathbf{x}})^2}$$

מתוכו, ניקח את החלק המודגש ונגזור ממנו משתנה חדש:

$$\mathrm{w_i} = \frac{\mathrm{x_i} - \bar{\mathrm{x}}}{\sum (\mathrm{x_i} - \bar{\mathrm{x}})^2}$$

ולכן כעת:

$$\hat{\beta} = \sum w_i y_i = \beta + \sum w_i u_i$$

נחזור לאומד שהנחנו קודם, ונפרק את k לשני חלקים – w, שהוא חלק מהאומד המקורי, ו-s, שישלים ל-k של האומד החדש.

$$\tilde{\beta} = \sum (s_i + w_i) y_i$$

כעת נשאל מתי $\tilde{\beta}$ הינו חסר הטיה.

$$E(\tilde{\beta}) = \underbrace{E\sum(w_i y_i)}_{=\beta} + E\sum(s_i y_i)$$

אבל $E[\sum(w_i y_i)] = \beta$. לכן כעת נותר לשאול מהם התנאים הנדרשים כך ש: $E\sum(s_i y_i) = 0$. נציב את Y:

$$E\sum\left(s_i(\alpha + \beta x_i + u_i)\right) = \sum s_i \alpha + \beta \sum s_i x_i + \sum s_i \underbrace{E(u_i)}_{=0}$$

מותר היה לנו לעשות זאת כיוון ש-S אינו מקרי.

האיבר האחרון יהיה שווה 0 כיוון שתוחלתו של U שווה 0.

כעת, כדי לאפס את שני החלקים הנותרים בביטוי, עלינו להגשים שני תנאים:

1. $\sum s_i = 0$
2. $\sum s_i x_i = 0$

כעת נכפה את התנאים הללו על $\tilde{\beta}$ כתנאי מקדים, כלומר, נבחר רק את האומדים המקיימים את התנאים הללו.

מה יצא לנו?

$$\tilde{\beta} = \sum s_i y_i + \sum w_i y_i = \beta + \sum w_i u_i + \sum s_i u_i = \beta + \sum (w_i + s_i) u_i$$

$$\boldsymbol{\tilde{\beta} = \sum(w_i + s_i)u_i + \beta}$$

כעת נחשב את שונות האומד, ונראה אם היא קטנה יותר מהשונות של אומד הריבועים הפחותים.

$$\operatorname{var}(\tilde{\beta}) = E\left(\tilde{\beta} - \underbrace{E(\tilde{\beta})}_{\beta}\right)^2 = E\left(\sum(w_i + s_i)u_i\right)^2$$

(המעבר אפשרי כיוון ש-$\tilde{\beta}$ חסר הטיה).

בדומה למה שעשינו קודם, כדי לפשט, ניקח דוגמה של שני איברים.

$$\begin{aligned} E\left(\sum(w_i + s_i)u_i\right)^2 &= E[(w_1 + s_1)u_1 + (w_2 + s_2)u_2]^2 \\ &= E(w_1 + s_1)^2 u_1^2 + E(w_2 + s_2)^2 u_2^2 + 2(s_1 + w_1)(s_2 + w_2)E(u_1 u_2) \end{aligned}$$

כבר הנחנו קודם,

$$E(u_1u_2)=\mathrm{cov}(u_1,u_2)=0$$

נמשיך:

$$=E\sum[(w_i+s_i)u_i]^2=\sum(w_i+s_i)^2\underbrace{E(u_i)^2}_{E\left(u_i-E(u_i)\right)^2=\sigma_u^2}=\sigma_u^2\sum(w_i+s_i)^2$$

$$\mathbf{var(\tilde{\beta})=\sigma_u^2\sum(w_i+s_i)^2}$$

נזכור את שונות אומד הריבועים הפחותים:

$$\mathrm{var}(\hat{\beta})=\frac{\sigma_u^2}{\sum_{i=1}^{I}(x_i-\bar{x})^2}$$

את σ_u^2 אפשר לבטל, ולכן צריך להבין מה היחס בין הגודל של האיברים שנותרו. כלומר, יש להוכיח:

$$\frac{1}{\sum(x_i-\bar{x})^2}<\sum(w_i+s_i)^2$$

הוכחה:

$$\sum(w_i+s_i)^2=\sum w_i^2+\sum s_i^2+2\sum w_is_i$$

$$\sum w_i^2=\sum\left(\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}\right)^2=\frac{\sum(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2}=\frac{1}{\sum(x_i-\bar{x})^2}$$

זה שווה בדיוק לשונות של $\hat{\beta}$. כעת נבדוק מה קורה עם האיברים האחרים – האם הם מגדילים או מקטינים את השונות.

$$\sum s_i^2>0$$

הסכום של $\sum s_i^2$ לא יכול להיות בדיוק שווה 0, כי אז האומד החדש היה שווה לאומד הריבועים הפחותים. בגלל הריבוע הוא גדול מ-0.

$$2\sum w_is_i=\frac{2\sum(x_i-\bar{x})s_i}{\sum(x_i-\bar{x})^2}$$

$$\sum(x_i-\bar{x})s_i=\sum x_is_i-\bar{x}\sum s_i=0$$

אבל עקב התנאים שהגדרנו קודם לחוסר הטיה, הביטוי הזה שווה כולו ל- 0.

בסיכומו של דבר:

$$\mathbf{var(\tilde{\beta})}=\frac{\sigma_u^2}{\sum(\mathbf{x_i}-\bar{\mathbf{x}})^2}+\sigma_u^2\sum\mathbf{s_i^2}>\frac{\sigma_u^2}{\sum(\mathbf{x_i}-\bar{\mathbf{x}})^2}=\mathbf{var(\hat{\beta})}$$

שימוש באומדים

בחלק זה, נדבר על רווח סמך, מבחני השערות ותחזיות.

$$\hat{\beta} = \beta + \frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}$$

מהנוסחה ברור שיש הפרש בין האומד לבין β האמיתית. הסטייה מ-β תלויה בעצם ב-u: זאת כיוון שב-X אני שולט.

הנחה חמישית: ההנחה היא ש-u מתפלג נורמלית!

$$u \sim N(0, \sigma_u^2)$$

ה-u במדגם בלתי תלויים, ולכן יכול להיות מצב שכל ה-u במדגם יגיעו רק מצד אחד של ההתפלגות, ובערכים קיצוניים. אבל הסיכוי בהתפלגות נורמלית לאירועים קיצוניים הוא קטן מאוד. לפיכך אנחנו יכולים לתחום איזשהו תחום הערכה ל-β, ולספק עבורו הסתברות.

רווח סמך ל- β

תחת ההנחה ש-u מתפלג נורמלית, $\hat{\beta}$ מתפלג נורמלי גם הוא – זאת כי טרנספורמציה ליניארית של משתנה המתפלג נורמלית, מתפלגת נורמלית אף היא. מכאן הגיעה דרישת הלינאריות באומד – על מנת שיתפלג נורמלית.

את השונות של $\hat{\beta}$ איננו יודעים – לכן בשיעור הקודם מצאנו לה אומד. זה אומר ש- $\hat{\beta}$ לא מתפלגת נורמלית, אלא בהתפלגות t. כדי שנוכל לעבוד, אנו זקוקים לסטנדרטיות – תוחלת 0 ושונות 1.

ראשית, נחסר את מהאומד את הפרמטר המקורי: $\hat{\beta}$-β. זאת כיוון שהתוחלת של $\hat{\beta}$ היא β, ו-$E(\beta-\hat{\beta})=0$. לאחר מכן נחלק באומד סטית התקן של $\hat{\beta}$:

$$-t \leq \frac{\hat{\beta} - \beta}{\hat{\sigma}_{\hat{\beta}}} \leq t$$

כאשר

$$\hat{\sigma}_{\hat{\beta}} = \sqrt{\widehat{var}(\hat{\beta})}$$

ולכן גם השונות תהיה 1, ונוכל לעבוד עם התפלגות t סטנדרטית.

דרגות החופש להתפלגות t נקבעות לפי מספר התצפיות, פחות מספר המשוואות הנורמליות במודל. ככל שיש פחות תצפיות, t גדל, ולכן רווח הסמך גדל, והתשובה שלנו אינה אמינה. לכן מדגם קטן יגרור קיום רווח סמך גדול.

רווח הסמך בנושא שלנו נכתב קצת אחרת ממה שעשינו בסטטיסטיקה ב'. נחזור לנוסחת האי שוויון ונעביר קצת אגפים:

$$\hat{\beta} - \hat{\sigma}_{\hat{\beta}} t \le \beta \le \hat{\beta} + \hat{\sigma}_{\hat{\beta}} t$$

ולמעשה:

$$P\left(\hat{\beta} - \hat{\sigma}_{\hat{\beta}} t \le \beta \le \hat{\beta} + \hat{\sigma}_{\hat{\beta}} t\right) = 1 - \alpha$$

כדי להקטין את רווח הסמך, יש להקטין את השונות של β (ע"י הגדלת המדגם או בחירת Xים מרוחקים), או להגדיל את α.

מבחני השערות

מבחני השערות מוצגים בצורה הבאה:

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

תוצאה של מבחן השערות לעולם אינה בטוחה ב-100% - וגם כאן יש לערב הסתברויות.

טעות מסוג ראשון – דחיתי את H_0 כאשר היא היתה נכונה.

טעות מסוג שני – קיבלתי את H_0 כאשר היא לא היתה נכונה.

בקורס אנו מדברים רק על הטעות מהסוג הראשון.

$$y_i = \alpha + \beta x_i + u_i$$

במודל שלנו, השאלה הראשונה שנשאל תמיד היא אם השיפוע שווה ל-0. כי אם השיפוע אפסי, אין קשר בין X ל-Y ואין טעם במודל.

השערת האפס היא, ש-β=0. אם זה נכון, אז:

$$-t_{\frac{\alpha}{2}} \le \frac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}} \le t_{\frac{\alpha}{2}}$$

רק אם ההשערה נכונה, המשתנה הזה מתפלג t. אם ההשערה אינה נכונה, התוחלת של הביטוי אינה שווה 0, והמשתנה לא יכול להיות סטנדרטי נורמלי.

אם **ה-t הסטטיסטי**, המחושב מתוצאות המדגם, חורג **מה-t הקריטי**, אותו מצאנו בטבלה לפי ההסתברות לטעות – אזי דוחים את ההשערה.

$\downarrow t \Leftarrow$ הסיכוי לדחות את ההשערה גדל.

פרשנות למבחני השערות ב-EVIEWS:

Coefficient = האומד

Prob = P. Value

Mean dependent var = $\bar{y}$

SD dependent var = $\hat{\sigma}_y$

Sum squared resid = $\sum (\hat{u}_i)^2$

SE of regression = $\hat{\sigma}_u$

נניח בודקים השערה אחרת:

$$H_0: \beta = 2$$

$$H_1: \beta \neq 2$$

$$-t_{\frac{\alpha}{2}} \leq \frac{\hat{\beta} - 2}{\hat{\sigma}_{\hat{\beta}}} \leq t_{\frac{\alpha}{2}}$$

מותר עדיין לחלק בפיזור של $\hat{\beta}$ כיוון שתוספת של 2 לא משנה את הפיזור של האומד.

מותר לבצע גם מבחנים חד צדדים:

$$H_0: \beta \geq 2$$

$$H_1: \beta < 2$$

במבחן שכזה, חוסמים רק מצד אחד.

$$-t_{\alpha} \leq \frac{\hat{\beta} - 2}{\hat{\sigma}_{\hat{\beta}}}$$

אותו דבר פועל גם בכיוון השני:

$$H_0: \beta \leq 2$$

$$H_1: \beta > 2$$

$$\frac{\hat{\beta} - 2}{\hat{\sigma}_{\hat{\beta}}} \leq t_{\alpha}$$

דוגמה על פונקצית ייצור

נניח שהמודל שלנו מסביר פונקצית יצור:

$$y_i = \alpha + \beta x_i + u_i$$

$$H_0: \beta = \alpha$$

$$H_1: \beta \neq \alpha$$

נגדיר משתנה חדש: $\hat{\beta}-\hat{\alpha}$. תוחלתו:

$$E(\hat{\beta} - \hat{\alpha}) = E(\hat{\beta}) - E(\hat{\alpha}) = \beta - \alpha$$

לפי ההנחה שלנו, ההפרש הזה צריך להיות 0. המשתנה המנורמל יהיה:

$$\frac{\hat{\beta} - \hat{\alpha}}{\sqrt{var(\hat{\beta} - \hat{\alpha})}}$$

השונות של המשתנה הנבדק תהיה:

$$var(\hat{\beta} - \hat{\alpha}) = var\hat{\beta} + var\hat{\alpha} - 2cov(\alpha, \beta)$$

את השונות של $\hat{\beta}$ אנחנו יודעים! חסרה לנו רק זו של $\hat{\alpha}$ והשונות המשותפת.

האומד של α

את כל הדברים שבדקנו עבור $\hat{\beta}$, כעת נבדוק עבור $\hat{\alpha}$.

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

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

$$\hat{\alpha} = \frac{1}{I}\sum y_i - \hat{\beta}\bar{x}$$

נציב את האומד של β:

$$\hat{\alpha} = \frac{1}{I}\sum y_i - \frac{\bar{x}(\sum(x_i - \bar{x})y_i)}{\sum(x_i - \bar{x})^2} = \sum\left(\frac{1}{I} - \frac{\bar{x}\left(\sum(x_i - \bar{x})\right)}{\sum(x_i - \bar{x})^2}\right)y_i$$

נשים לב שהאומד שלנו ל-α הוא לינארי ב-Y. כעת, נציב את y מהמדגם.

$$\hat{\alpha} = \sum\left(\frac{1}{I} - \frac{\bar{x}\left(\sum(x_i - \bar{x})\right)}{\sum(x_i - \bar{x})^2}\right)(\alpha + \beta x_i + u_i)$$

נפרק את האיברים אחד אחד:

$$\sum\frac{1}{I}\cdot\alpha = I\frac{1}{I}\alpha = \alpha$$

$$\frac{\sum\bar{x}(x_i-\bar{x})\alpha}{\sum(x_i-\bar{x})^2}=\bar{x}\alpha\frac{\overbrace{\sum(x_i-\bar{x})}^{=0}}{\sum(x_i-\bar{x})^2}=0$$

$$\sum\frac{1}{I}\beta x_i=\frac{1}{I}\beta\sum x_i=\frac{1}{I}\beta I\bar{x}=\beta\bar{x}$$

$$-\frac{\sum\bar{x}(x_i-\bar{x})\beta x_i}{\sum(x_i-\bar{x})^2}=-\bar{x}\beta\overbrace{\frac{\sum(x_i-\bar{x})x_i}{\sum(x_i-\bar{x})^2}}^{=1}=-\bar{x}\beta$$

עם הכפולות של U_i אין מה לעשות, ולכן נותרנו עם:

$$\boldsymbol{\hat{\alpha}=\alpha+\sum\left(\frac{1}{I}-\frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)u_i}$$

למדנו כי שני האומדים שלנו, גם α וגם β, תלויים בהפרעות של המדגם.

נסמן:

$$s_i=\frac{1}{I}-\frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}$$

ולכן:

$$\boldsymbol{\hat{\alpha}=\alpha+s_iu_i}$$

נבדוק ש-α חסר הטיה:

$$E(\hat{\alpha})=E(\alpha+\sum s_iu_i)=E(\alpha)+E(\sum s_iu_i)=\alpha+\sum s_i\underbrace{E(u_i)}_{=0}=\alpha$$

נמצא את השונות של α:

$$var(\hat{\alpha})=E\left(\hat{\alpha}-\overbrace{E(\hat{\alpha})}^{=\alpha}\right)^2=E(\sum s_iu_i)^2$$

למען הפשטות נדגים שוב עם שתי תצפיות:

$$E(s_1u_1+s_2u_2)^2=E(s_1u_1)^2+E(s_2u_2)^2+2s_1s_2E(u_1u_2)$$

$$E(u_1u_2)=E\Big(\big(u_1-E(u_1)\big)\big(u_2-E(u_2)\big)\Big)=cov(u_1,u_2)\overset{\text{הנחה}}{=}0$$

וכעת:

$$var(\hat{\alpha})=E(\sum(u_is_i)^2)=\sum s_i^2\underbrace{E\big(u_i-E(u_i)\big)^2}_{\sigma_{u_i}^2=\sigma_u^2}=\sigma_u^2\sum s_i^2$$

נמשיך לפתח את הסכום של ה-s_i^2:

$$\sum s_i^2 = \sum\left(\frac{1}{I} - \frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)^2$$

נפצל למחוברים לנוחות העבודה:

$$\sum\left(\frac{1}{I}\right)^2 = \frac{I}{I^2} = \frac{1}{I}$$

$$\sum\left(-\frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)^2 = \frac{\bar{x}^2\sum(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2} = \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}$$

$$-2\sum\left(\frac{\frac{1}{I}\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right) = -\frac{2\bar{x}}{I}\cdot\frac{\sum(x_i-\bar{x})}{\sum(x_i-\bar{x})^2} = 0$$

ולכן:

$$var(\hat{\alpha}) = \sigma_u^2\left(\frac{1}{I} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)$$

וכרגיל, מי שקובע את שונות האומד, הוא הסטיה האקראית. וכמו קודם, את השונות של הסטיה איננו יודעים, ונוכל להשתמש באומד:

$$\boldsymbol{\widehat{var}(\hat{\alpha}) = \hat{\sigma}_u^2\left(\frac{1}{I} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}\right)}$$

השונות המשותפת של $\hat{\alpha}$ ו- $\hat{\beta}$

$$cov(\hat{\alpha},\hat{\beta}) = E\left[\left(\hat{\alpha} - \overset{=\alpha}{E(\hat{\alpha})}\right)\left(\hat{\beta} - \overset{=\beta}{E(\hat{\beta})}\right)\right] = E\left[\left(\sum s_i u_i\right)\left(\sum w_i u_i\right)\right]$$

כרגיל, נדמה על שני איברים:

$$\begin{aligned} &E[(s_1u_1 + s_2u_2)(w_1u_1 + w_2u_2)] \\ &\qquad = s_1w_1E(u_1^2) + s_2w_2E(u_2^2) + s_1w_2E(u_1u_2) + s_2w_1\underbrace{E(u_1u_2)}_{cov(u_1,u_2)=0} \end{aligned}$$

כל ההצלבות נמחקות, ולכן נשארות לנו רק המכפלות עם אותו אינדקס.

$$cov(\hat{\alpha},\hat{\beta}) = E\left[\sum(s_iw_iu_i^2)\right] = \sum s_iw_iE\left(u_i - E(u_i)\right)^2 = \sigma_u^2\sum s_iw_i$$

$$\sum s_iw_i = \sum\left(\frac{1}{I} - \frac{\bar{x}(x_i-\bar{x})}{\sum(x_i-\bar{x})^2}\right)\left(\frac{x_i-\bar{x}}{\sum(x_i-\bar{x})^2}\right) = -\frac{\sum\bar{x}(x_i-\bar{x})^2}{\left(\sum(x_i-\bar{x})^2\right)^2} = -\frac{\bar{x}}{\sum(x_i-\bar{x})^2}$$

נציב חזרה בנוסחה:

$$\boldsymbol{cov(\hat{\alpha},\hat{\beta}) = -\frac{\sigma_u^2 \bar{x}}{\sum (x_i - \bar{x})^2}}$$

יש קשר ישיר במדגם בין השונות המשותפת לבין השונות של $\hat{\beta}$:

$$var(\hat{\beta}) = \frac{\sigma_u^2}{\sum (x_i - \bar{x})^2}$$

<u>נחזור לבדיקת ההשערות שהתחלנו קודם:</u>

$$\frac{\hat{\beta} - \hat{\alpha}}{\sqrt{\widehat{var}(\hat{\alpha}) + \widehat{var}(\hat{\beta}) - 2\widehat{cov}(\hat{\alpha},\hat{\beta})}}$$

נבדוק, למשל, את ההשערה הבאה:

$$\alpha = 2\beta$$

אז המשתנה הנורמלי שלנו יהיה:

$$\frac{\hat{\alpha} - 2\hat{\beta}}{\sqrt{var(\hat{\alpha} - 2\hat{\beta})}}$$

$$var(\hat{\alpha} - 2\hat{\beta}) = var(\hat{\alpha}) + 4var(\hat{\beta}) - 2 \cdot 2cov(\hat{\alpha},\hat{\beta})$$

מודלים שאינם לינאריים בפרמטרים

מודל סמי לוגריתמי

המודל הרגיל שלנו נראה כך:

$$y_i = \alpha + \beta x_i + u_i$$

יתכנו מודלים שאינם לינאריים במשתנים, כמו **מודל סמי לוגריתמי**:

$$\ln(y_i) = \alpha + \beta x_i + u_i$$

במודל שכזה,

$$\frac{\partial \ln(y_i)}{\partial x_i} = \beta$$

אבל החוקרים מעוניינים בקשר בין y_i ל-x_i, כך:

$$\frac{\partial y_i}{\partial x_i}$$

כדי להגיע לשם, נפצל את הביטוי לקשר דו שלבי.

$$\frac{\partial \ln(y_i)}{\partial x_i} = \frac{\partial y_i}{\partial x_i} \cdot \underbrace{\frac{\partial \ln(y_i)}{\partial y_i}}_{=\frac{1}{y_i}} \Rightarrow \boldsymbol{\frac{\partial y_i}{\partial x_i} = \beta y_i}$$

כלומר, כל הזזה של x, מזיזה את y בשיעור β אחוזים.

מודל לוגריתמי מלא

ניתן לבדוק גם מודל שהוא **לוגריתמי מלא**:

$$\ln(y_i) = \alpha + \beta \ln(x_i) + u_i$$

מודל כזה יכול להיווצר ממודל שאינו ליניארי בפרמטרים:

$$y_i = \alpha \cdot x_i^{\beta} \cdot e^{u_i}$$

עליו מפעילים לוגריתם ומקבלים מודל לוגריתמי מלא:

$$\ln(y_i) = \underbrace{\ln(\alpha)}_{\alpha'} + \beta \ln(x_i) + u_i$$

נפעיל את אותו כלי דו שלבי שהשתמשנו בו קודם:

$$\frac{\partial \ln(y_i)}{\partial \ln(x_i)} = \frac{\partial y_i}{\partial x_i} \cdot \underbrace{\frac{\partial \ln(y_i)}{\partial (y_i)}}_{\frac{1}{y_i}} \cdot \underbrace{\frac{\partial x_i}{\partial \ln(x_i)}}_{x_i} \Rightarrow \boldsymbol{\frac{\partial y_i}{\partial x_i} = \beta \cdot \frac{y_i}{x_i}}$$

זוהי בדיוק הגדרת הגמישות! המשמעות – הזזה של X באחוז אחד, מזיזה את Y בשיעור β אחוזים.

חיזוי

השאיפה שלנו היא לחזות במונחי t, זמן, ולא במונחי i – יותר מעניין לבצע תחזיות על אותה אוכלוסיה לאורך זמן, מאשר לפרטים שנמצאים מחוץ לאוכלוסיה. לכן המודל שלנו יראה כך:

$$y_t = \alpha + \beta x_t + u_t$$

התחזית שאנו רוצים לבצע היא:

$$y_{T+K} \backslash x_{T+K}$$

ברוב המקרים נקבע את K=1. יש לזכור כי את X "של מחר" אנחנו קובעים, כי אינו משתנה מקרי. לכן התחזית שלנו:

$$\hat{y}_{T+1} = \hat{\alpha} + \hat{\beta} x_{T+1}$$

בשלב זה אנו משמיטים את ה-U כיוון שהתוחלת שלו היא 0.

$$E(\hat{y}_{T+1}) = E\left(\hat{\alpha} + \hat{\beta} x_{T+1}\right) = \alpha + \beta x_{T+1} = E(y_{T+1})$$

כעת אנו מעוניינים לבנות רווח סמך ל-Y האמיתי, על סמך Y החזוי. הסיפור פה יהיה יותר מסובך, זאת כיוון ש-β האמיתי היה קבוע, ולא היתה לו שונות – לעומת זאת, ל-Y האמיתי בהחלט יש שונות.

$$-t_{\frac{\alpha}{2}} \le \frac{y_{T+1} - \hat{y}_{T+1}}{\sqrt{var\left(\underbrace{y_{T+1} - \hat{y}_{T+1}}_{P.E}\right)}} \le t_{\frac{\alpha}{2}}$$

$$var(y_{T+1} - \hat{y}_{T+1}) = var\left(\alpha + \beta x_{T+1} + u_{T+1} - \left(\hat{\alpha} + \hat{\beta} x_{T+1}\right)\right)$$

α,β,x אינם מקריים, לכן אינם משפיעים על השונות.

$$= var\left(u_{T+1} - \hat{\alpha} - \hat{\beta} x_{T+1}\right)$$
$$= \sigma_u^2 + var(\hat{\alpha}) + x_{T+1}^2 \cdot var\left(\hat{\beta}\right) - 2\underbrace{cov(u_{T+1}, \hat{\alpha})}_{=0}$$
$$- 2x_{T+1}\underbrace{cov\left(u_{T+1}, \hat{\beta}\right)}_{=0} - 2x_{T+1}\underbrace{cov\left(\hat{\alpha}, \hat{\beta}\right)}_{=\bar{x} var\left(\hat{\beta}\right)}$$

את השונות המשותפת של α̂ עם ההפרעה האקראית של מחר אפשר לאפס: זאת כיוון שהאומד תלוי בהפרעה האקראית של המדגם, והנחנו כי אין תלות בין ההפרעה האקראית של תצפיות שונות. כנ"ל לגבי β.

$$= \sigma_u^2 + var(\hat{\alpha}) + x_{T+1}^2 var\left(\hat{\beta}\right) - 2x_{T+1}^2 \cdot \bar{x} \cdot var\left(\hat{\beta}\right)$$

מציבים במקום α̂ ו-β̂ את השונויות שלהם ומוציאים גורם משותף:

$$\boldsymbol{var(y_{T+1} - \hat{y}_{T+1} \backslash x_{T+1}) = \hat{\sigma}_{PE}^2 = \hat{\sigma}_{\hat{u}}^2 \left(1 + \frac{1}{T} + \frac{(x_{T+1} - \bar{x})^2}{\sum (x_t - \bar{x})^2}\right)}$$

הגורמים המשפיעים על התחזית:

1. ככל ששונות ההפרעה גדלה, אי הדיוק גדל.
2. ככל שנבחר X למחר שקרוב יותר ל-X הממוצע, הדיוק גדל.
3. ככל שכמות התצפיות T גדלה, גם הדיוק גדל.

נציב חזרה ברווח הסמך ונקבל:

$$\boldsymbol{\hat{y}_{T+1} - t_{\frac{\alpha}{2}} \hat{\sigma}_{PE} \le y_{T+1} \backslash x_{T+1} \le \hat{\sigma}_{PE} t_{\frac{\alpha}{2}} + \hat{y}_{T+1}}$$

השונות של התחזית לא משתנה, כלומר, רווח הסמך לא גדל, במידה ונרצה לבצע תחזית של T+2 או רחוק מכך (בהנחה ש-α ו-β האמיתיים לא משתנים על פני זמן). זה כיוון שאנו מניחים ש-X לא מקרי: אם X כן היה מקרי, זה היה מכניס עוד מקום לטעות.

השפעת שינוי סקאלה על האומד

יש לנו מודל כזה:

$$y_i = \alpha + \beta x_i + u_i$$

ומישהו אומד מודל אחר:

$$y_i = \gamma + \delta z_i + \epsilon_i$$

וניתן הקשר:

$$z_i = 10x_i$$

מקבלים כי:

$$\hat{\delta} = \frac{\sum(z_i - \bar{z})y_i}{\sum(z_i - \bar{z})^2} = \frac{\sum(10x_i - 10\bar{x})y_i}{\sum(10x_i - 10\bar{x})^2} = \frac{1}{10} \cdot \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \frac{\hat{\beta}}{10}$$

כלומר, אם Z חזק פי 10 מ-X, אז β צריך להיות חזק יותר פי 10 מ-δ.

$$\hat{\gamma} = \bar{y} - \hat{\delta}\bar{z} = \bar{y} - \frac{1}{10} \cdot \hat{\beta} \cdot 10\bar{x} = \hat{\alpha}$$

המשמעות: הכפלה בקבוע אינה מזיזה את החותך.

נבדוק מה קורה אם השינוי יהיה הזזה לינארית:

$$z_i = x_i + 5$$

$$\hat{\delta} = \frac{\sum(z_i - \bar{z})y_i}{\sum(z_i - \bar{z})^2} = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \hat{\beta}$$

אין השפעה על β.

$$\hat{\gamma} = \bar{y} - \hat{\delta}\bar{x} = \bar{y} - \hat{\beta}(\bar{x} + 5) = \bar{y} - \hat{\beta}\bar{x} - \hat{\beta}5 = \hat{\alpha} - 5\hat{\beta}$$

מדד לאיכות הרגרסיה R^2

המדד הזה הוא מדד מסוכן, ומותר להשתמש בו רק במקרים מסויימים. ראשית נוכיח את המדד, ולאחר מכן נלמד את הלוגיקה באשר מתי מותר להשתמש בו.

$$y_i = \alpha + \beta x_i + u_i$$

$$H_0: \beta = 0$$

$$H_1: \beta = 1$$

דחינו את השערת האפס, ואנחנו בטוחים שיש קשר β בין X ל-Y. אבל זה עדיין לא נותן לנו מידע לגבי חשיבות הקשר! כלומר, יתכן שיש דברים שמשפיעים על התזוזה של Y בצורה משמעותית הרבה יותר מ-X. ככל ש-X אחראי יותר על התזוזות של Y, הרגרסיה שלנו באיכות טובה יותר.

אנחנו רוצים להסביר את התזוזות של Y מתצפית לתצפית, ולראות מתי התזוזה היתה בגלל X ומתי לא. כשאנחנו בוחנים את Y, הממוצע שלו לא מעניין – הרגרסיה תמיד עוברת בנקודת הממוצעים בכל מקרה. לכן ננטרל את הממוצעים. כדי שהפרשים שליליים לא ינטרלו חיוביים, נעלה את ההפרש בריבוע. שיטת הבדיקה הזו מתאימה רק למודלים עם חותך, כי אחרת הרגרסיה לא עוברת בנקודת הממוצעים.

$$y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{u}_i$$

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

$$\sum_{i=1}^{I} (y_i - \bar{y})^2 = \sum \left(\hat{\beta}(x_i - \bar{x}) + \hat{u}_i \right)^2 = \hat{\beta}^2 \cdot \sum (x_i - \bar{x})^2 + \sum \hat{u}_i^2 + 2\hat{\beta} \underbrace{\sum (x_i - \bar{x}) \hat{u}_i}_{=0}$$

הסיבה שאין מתאם בין U ל-X היא ש-X "סוחט" את כל ההשפעה האפשרית על Y מ-U. זו למעשה המשמעות של המשוואות הנורמליות. מה שנשאר ב-U אלה דברים שאין שום אפשרות להסביר אותם על ידי X.

נחלק את שני הצדדים של המשוואה בפיזור של Y:

$$1 = \underbrace{\frac{\overbrace{\hat{\beta}^2 \sum (x_i - \bar{x})^2}^{SSR}}{\underbrace{\sum (y_i - \bar{y})^2}_{SST}}}_{R^2} + \underbrace{\frac{\overbrace{\sum \hat{u}_i^2}^{SSE}}{\underbrace{\sum (y_i - \bar{y})^2}_{SST}}}_{1-R^2}$$

$$R^2 = \frac{SSR}{SST}$$

R^2 מבטא כמה המשתנים המסבירים תורמים לפיזור של Y. ככל ש-R^2 יותר גדול, אז $1- R^2$, כלומר, החלק של ההפרעות – יותר קטן.

הקשר בין R^2 למקדם המתאם

$$R^2 = \hat{\beta}^2 \frac{\sum (x_i - \bar{x})^2}{\sum (y_i - \bar{y})^2} = \left(\frac{cov(x,y)}{var(x)} \right)^2 \cdot \frac{var(x)}{var(y)} = \left(\frac{cov(x,y)}{\sigma_x \sigma_y} \right)^2$$

$$\boldsymbol{R^2 = \rho^2}$$

שוויון זה תקף רק אם יש משתנה מסביר אחד בלבד.

שינויים ב-R^2

אם יש רק שתי תצפיות, $R^2=1$.

אם נוסיף למדגם תצפית בה X ו-Y שווים לערכים הממוצעים, למעשה הוספנו עוד תצפית שנמצאת על נקודת הממוצעים, שלא יוצרת חריגה. גם המונה וגם המכנה לא השתנו, ולכן R^2 לא השתנה.

אם הוספנו תצפית היושבת על הרגרסיה, המכנה גדל, בצד שמאל המונה גדל, בצד ימין הוא קטן - R^2 עולה.

אם הוספנו תצפית שלא יושבת על קו הרגרסיה, α ו-β השתנו, ולא ניתן לדעת בדיוק מה יקרה ל-R^2 .

שיטת הריבועים הפחותים מבטיחה R^2 מקסימלי. זה בגלל שבבסיס השיטה הוא מזעור הטעויות בריבוע.

השוואת שתי רגרסיות

מתי אסור להשוות שתי רגרסיות באמצעות R^2?

<u>מקרה 1</u>

האם ניתן להשוות את ציוני R^2 שנקבל משני המודלים הללו?

$$min\sum\underbrace{\left(y_i-\tilde{\alpha}-\tilde{\beta}x_i\right)}_{\hat{u}_i}$$

$$min\sum\underbrace{\left(y_i-\tilde{\gamma}-\tilde{\delta}x_i-\tilde{\theta}z_i\right)}_{\hat{\epsilon}_i}$$

למעשה, תמיד הולך להתקיים המצב הבא:

$$\sum\hat{\epsilon}_i^2\leq\sum\hat{u}_i^2$$

למה? זאת כיוון שאם θ היה שווה 0, כלומר, המחשב לא היה מוצא קשר בין Z ל-Y, סכום הטעויות היה זהה. בכל מקרה אחר סכום הטעויות היה קטן יותר, כיוון שאין סיבה שהמחשב יחפש להגדיל את הטעויות. המשמעות היא ש- **R^2 גדל כשמספר המשתנים גדל – ולכן אי אפשר להשתמש בו להשוואה בין מודלים בהם מספר המשתנים שונה.**

קיים גם R^2 מתוקנן:

$$\bar{R}^2=\frac{I+K}{I-K}\left(1-R^2\right)$$

כאשר K הוא מספר המשתנים המסבירים. כלומר, כשמוסיפים עוד משתנים מסבירים, R^2 המתוקנן יקבל "מכה" כלפי מטה, וכדי שלא ירד, השונות תצטרך לרדת באופן חד יותר. עם זאת, R^2 המתוקנן יעלה לא רק אם המשתנה המסביר החדש מספיק מובהק וחשוב. כדי שהמתוקנן יעלה מספיק שיתקיים:

שהוא, כידוע, רחוק מאוד מהמובהקות הנדרשת על ידינו (1.96 עבור מובהקות 95%).

בשורה התחתונה: **גם עם תיקנון של R^2, יתכן שהמדד יעלה עם הוספת משתנה מסביר נוסף, גם אם רמת המובהקות שלו נמוכה. לכן צריך להיזהר בשימוש גם במדד הזה בהשוואה בין מודלים בעלי כמות שונה של משתנים.**

<u>מקרה 2</u>

$$y_i = \alpha + \beta x_i + u_i$$

$$\ln(y_i) = \gamma + \delta x_i + \epsilon_i$$

זאת כיוון שמשתנה באופן מהותי הפיזור של Y.

<u>מקרה 3</u>

כשבודקים את שיעור השינוי של Y לאורך זמן:

$$y_t = \alpha + \beta x_t + u_t$$

$$\frac{y_t - y_{t-1}}{y_{t-1}} = \gamma + \frac{\delta(x_t - x_{t-1})}{x_t} + \epsilon_t$$

הבעיה במודל השני, הוא שרמת השינוי היא מאוד עדינה. הגודל במונחים אבסולוטיים הוא פחות או יותר יציב, אבל ההפרש בין שינוי באחוז לבין שינוי בשני אחוז, הוא מאוד דרמטי. לכן **מדידת שיעורי שינוי מבטיחה R^2 גרוע ביותר.**

רגרסיה מרובת משתנים

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

$$\min_{\tilde{\alpha},\tilde{\beta},\tilde{\gamma}} \sum \left(y_i - \tilde{\alpha} - \tilde{\beta} x_i - \tilde{\gamma} z_i\right)^2$$

$$(I)\ \frac{\partial}{\partial \tilde{\alpha}}: -2\sum\left(y_i - \hat{\alpha} - \hat{\beta} x_i - \hat{\gamma} z_i\right) = 0$$

$$(II)\ \frac{\partial}{\partial \tilde{\beta}}: -2\sum\left(y_i - \hat{\alpha} - \hat{\beta} x_i - \hat{\gamma} z_i\right) x_i = 0$$

$$(III)\ \frac{\partial}{\partial \tilde{\gamma}}: -2\sum\left(y_i - \hat{\alpha} - \hat{\beta} x_i - \hat{\gamma} z_i\right) z_i = 0$$

$$(I)\ I\bar{y} = I\hat{\alpha} + \hat{\beta}I\bar{x} + \hat{\gamma}I\bar{z}\ \Rightarrow \bar{y} = \hat{\alpha} + \hat{\beta}\bar{x} + \hat{\gamma}\bar{z}$$

כלומר, גם ברגרסיה מסוג זה, היא תעבור דרך נקודת הממוצעים.

נציב את $\hat{\alpha}$ מהמשוואה הנ"ל במשוואות II ו-III.

$$(\mathrm{II})\sum\left(y_i - \left(\bar{y} - \hat{\beta}\bar{x} - \hat{\gamma}\bar{z}\right) - \hat{\beta}x_i - \hat{\gamma}z_i\right)x_i = 0$$

$$\sum(y_i - \bar{y})x_i = \hat{\beta}\sum(x_i - \bar{x})x_i + \hat{\gamma}\sum(z_i - \bar{z})x_i$$

$$(\mathrm{III})\sum\left(y_i - \left(\bar{y} - \hat{\beta}\bar{x} - \hat{\gamma}\bar{z}\right) - \hat{\beta}x_i - \hat{\gamma}z_i\right)z_i = 0$$

$$\sum(y_i - \bar{y})z_i = \hat{\beta}\sum(x_i - \bar{x})z_i + \hat{\gamma}\sum(z_i - \bar{z})z_i$$

כדי שיהיה לנו קל לעבוד, נלמד כתב קיצור:

$$\sum(y_i - \bar{y})(x_i - \bar{x}) = M_{yx}$$

$$\sum(x_i - \bar{x})x_i = M_{xx}$$

נציב:

$$(II)M_{yx} = \hat{\beta}M_{xx} + \hat{\gamma}M_{xz} \Rightarrow\ \hat{\beta} = \frac{M_{yx} - \hat{\gamma}M_{zx}}{M_{xx}}$$

$$(III)M_{yz} = \hat{\beta}M_{xz} + \hat{\gamma}M_{zz} \Rightarrow\ \hat{\gamma} = \frac{\left(M_{yz} - \hat{\beta}M_{xz}\right)}{M_{zz}}$$

$$\hat{\beta} = \frac{M_{yx} - \left(\frac{M_{yz} - \hat{\beta}M_{xz}}{M_{zz}}\right)\cdot M_{zx}}{M_{xx}} = \frac{M_{yx}M_{zz} - \left(M_{yz} - \hat{\beta}M_{xz}\right)\cdot M_{zx}}{M_{xx}M_{zz}}$$

$$\hat{\beta}\cdot M_{zz}M_{xx} - \hat{\beta}M_{xz}^2 = M_{yx}M_{zz} - M_{yz}M_{xz}$$

$$\boldsymbol{\hat{\beta}} = \boldsymbol{\frac{M_{yx}M_{zz} - M_{yz}M_{xz}}{M_{zz}M_{xx} - M_{xz}^2}}$$

לצורך השוואה, במודל עם שני משתנים, האומד שלנו הוא:

$$\hat{\beta} = \frac{M_{yx}}{M_{xx}}$$

אם M_{xz} היה שווה 0, היינו מקבלים את אותו האומד. מתברר, שאם אין מתאם בין X ל-Z, ניתן להשתמש גם במודל הפשוט כדי לאמוד את הקשר בין X ל-Y, גם אם המודל הכולל את Z הוא הנכון.

אם X ו-Z מתואמים, חייבים למדוד את כל המודל יחד: אחרת X יספוג את כל ההשפעה של Z!

אם אין מתאם בין X ל-Z, הייתי מקבל את אותם אומדים גם במודל של משתנה אחד וגם במודל של שניים, והשפעת המשתנה המושמט תיספג ב-u.

$$\boldsymbol{\hat{\gamma} = \frac{M_{yz}M_{xx} - M_{yx}M_{xz}}{M_{xx}M_{zz} - M_{xz}^2}}$$

<u>חוסר הטיה</u>

$$\hat{\beta} = \frac{M_{yx}M_{zz} - M_{yz}M_{xz}}{M_{zz}M_{xx} - M_{xz}^2} = \frac{\sum[(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}](y_i - \bar{y})}{M_{xx}M_{zz} - M_{xz}^2}$$
$$= \frac{\sum[(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}](\alpha + \beta x_i + \gamma z_i + u_i)}{M_{xx}M_{zz} - M_{xz}^2}$$

נפרק את המכפלה למחובריה. כופל ראשון: α

$$\sum(x_i - \bar{x})M_{zz} \cdot \alpha - \sum(z_i - \bar{z})M_{xz} \cdot \alpha = M_{zz} \cdot \alpha \cdot \underbrace{\sum(x_i - \bar{x})}_{=0} - M_{xz} \cdot \alpha \cdot \underbrace{\sum(z_i - \bar{z})}_{=0} = 0$$

כופל שני: β

$$\frac{\sum(x_i - \bar{x})M_{zz}\beta x_i - \sum(z_i - \bar{z})M_{xz}\beta x_i}{M_{zz}M_{xx} - M_{xz}^2} = \frac{\beta M_{zz}\underbrace{\sum(x_i - \bar{x})x_i}_{=M_{xx}} - \beta M_{xz}\underbrace{\sum(z_i - \bar{z})x_i}_{=M_{xz}}}{M_{zz}M_{xx} - M_{xz}^2} = \beta$$

כופל שלישי: γ

$$\sum(x_i - \bar{x})M_{zz}\gamma z_i - \sum(z_i - \bar{z})M_{zx}\gamma z_i = \gamma M_{zz}\sum(x_i - \bar{x})z_i - M_{xz}\gamma\sum(z_i - \bar{z})z_i = 0$$

כרגיל נישאר רק עם הכופל האחרון, u.

$$\boldsymbol{\hat{\beta} = \beta + \frac{\sum[(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]u_i}{M_{xx}M_{zz} - M_{xz}^2}}$$

$$E(\hat{\beta}) = \underbrace{E(\beta)}_{=\beta} + \frac{\sum[(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]}{M_{xx}M_{zz} - M_{xz}^2} \cdot \underbrace{E(u_i)}_{=0\ \text{הנחה}} = \beta$$

ולכן האומד חסר הטיה.

<u>שונות</u>

$$var(\hat{\beta}) = E\left(\hat{\beta} - \underbrace{E(\hat{\beta})}_{=\beta}\right)^2 = E\left(\frac{\sum[(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]u_i}{M_{xx}M_{zz} - M_{xz}^2}\right)^2$$

למען הנוחות, נסמן:

$$L_i = \frac{[(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]}{M_{xx}M_{zz} - M_{xz}^2}$$

$$var(\hat{\beta}) = E(\sum L_i u_i)^2$$

שוב, למען הפשטות, נעבוד עם שני איברים ראשונים:

$$E(L_1 u_1 + L_2 U_2)^2 = L_1^2 E(u_1^2) + L_2^2 E(u_2^2) + 2L_1 L_2 \underbrace{E(u_1 u_2)}_{=0}$$

לכן, שוב , הסכום יכול לצאת מהריבוע.

$$var(\hat{\beta}) = E(\sum L_i u_i)^2 = E[\sum (L_i u_i)^2] = \sum L_i^2 E(u_i^2)$$

נחזור מ-L לביטוי המקורי ונטפל רק במונה:

$$\sum [(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]^2 E(u_i^2) = \sigma_u^2 \sum [(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]^2$$

ניפרד לרגע משונות הטעות ונמשיך לעבוד עם הסכום שנותר:

$$\begin{aligned}
&\sum [(x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}]^2 \\
&\quad = \overbrace{\sum (x_i - \bar{x})^2}^{M_{xx}} M_{zz}^2 + \overbrace{\sum (z_i - \bar{z})^2}^{M_{zz}} \cdot M_{xz}^2 - 2M_{zz}M_{xz} \overbrace{\sum (x_i - \bar{x})(z_i - \bar{z})}^{M_{xz}} \\
&\quad = M_{xx}M_{zz}^2 - M_{zz}M_{xz}^2
\end{aligned}$$

נציב הכל בחזרה במקום ונקבל:

$$var(\hat{\beta}) = \frac{\sigma_u^2 (M_{xx}M_{zz}^2 - M_{zz}M_{xz}^2)}{(M_{xx}M_{zz} - M_{xz}^2)^2} = \sigma_u^2 \cdot \frac{M_{zz}}{M_{xx}M_{zz} - M_{xz}^2} = \frac{\sigma_u^2}{M_{xx} - \frac{M_{xz}^2}{M_{zz}}}$$

$$= \frac{\sigma_u^2}{M_{xx}\left(1 - \frac{M_{xz}^2}{M_{xx}M_{zz}}\right)}$$

$$\boldsymbol{var(\hat{\beta}) = \frac{\sigma_u^2}{M_{xx}(1 - \rho_{x,z}^2)}}$$

במודל הקודם, השונות היתה:

$$var(\hat{\beta}) = \frac{\sigma_u^2}{M_{xx}}$$

שתי מסקנות:

א. אם אין מתאם בין *x* ל-*z* (כלומר, $\rho_{x,z}=0$), המודל יוצא אותו מודל, והשונות היא אותה שונות.

ב. ככל שיש מתאם גדול יותר בין X ל-Z, השונות גדלה. זה בגלל שקשה יותר להבחין מי תרם יותר ל-y.
מתאם גבוה יותר בין X ל-Z, מחליש את יכולת האבחנה של המודל, ותגרור רווחי סמך גדולים יותר.
אם הם זזים לגמרי ביחד, כלומר מתאם מלא ($\rho_{X,Z}=1$), השונות תהיה אינסופית.

יש לשים לב: את שונות הטעות האמיתית, *אנחנו לא יודעים*. אומד השונות לטעות הוא:

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{I - \text{מס פרמטרים}}$$

במדגם הגדול יש פחות טעויות, ולכן הסכום במונה יהיה קטן יותר. מצד שני, יש יותר פרמטרים, ולכן גם המכנה קטן. המשמעות היא שאנחנו לא יודעים מה יקרה לאומד השונות כשעוברים למודל בעל יותר פרמטרים.

אם עובדים עם מודל שיש בו יותר משני משתנים מסבירים, לא ניתן לחשב מקדם מתאם. במקרה כזה, השונות תהיה:

$$\frac{\sigma_u^2}{M_{xx}\,(1 - R^2)}$$

כאשר את ה-R^2 יש לחשב עבור רגרסיה שתקשר **בין המשתנים המסבירים**:

$$x_i = \delta_0 + \delta_1 z_i + \delta_2 k_i + \epsilon_i$$

הסדר של הקשר בין המשתנים המסבירים אינו משנה:

$$z_i = \gamma_0 + \gamma_1 x_i + \gamma_2 k_i + \epsilon_i$$

מולטי קוליניאריות מושלמת

אם יש מולטי קוליניאריות מושלמת בין המשתנים המסבירים, מקבלים שונות אינסופית, ולמעשה לא ניתן אפילו לקבל אומדים.

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

$$(I) \sum (y_i - \alpha - \beta x_i - \gamma z_i) = 0$$

$$(II) \sum (y_i - \alpha - \beta x_i - \gamma z_i) x_i = 0$$

$$(III) \sum (y_i - \alpha - \beta x_i - \gamma z_i) z_i = 0$$

נניח שיש קשר ליניארי מושלם בין X ל-Z:

$$x_i = A + c z_i$$

$$(II) \sum (y_i - \alpha - \beta x_i - \gamma z_i)(A + c z_i) = 0$$

$$A \cdot \sum (y_i - \alpha - \beta x_i - \gamma z_i) + c \cdot \sum (y_i - \alpha - \beta x_i - \gamma z_i) z_i = 0$$

אבל החלק הראשון התאפס בגלל I, והחלק השני התאפס בגלל III: ומשוואה II כולה מתאפסת. אז נותר לנו לפתור שתי משוואות עם שלושה פרמטרים, מצב בו יש אינסוף פתרונות – ולכן לא ניתן לאמוד את הפרמטרים. אין שום יכולת להפריד בין ההשפעה של X להשפעה של Z.

למשל, מודל של שכר:

$$\ln(w_i) = \alpha + \beta\,\text{השכלה}_i + \gamma\,\text{גיל}_i + \delta\,\text{ותק}_i + \epsilon_i$$

$$\text{ותק}_i = \text{גיל}_i + \text{השכלה}_i - 6 - 3$$

מודל כזה כופה קוליניאריות מושלמת על המודל. במקרה כזה, חלק מהתוכנות לא יחזירו פתרון כלל, ותוכנות אחרות מניחות שאחד הפרמטרים הוא 0, ואומדות את האחרים.

אפשר להסתכל על המולטיקוליניאריות בדרך נוספת:

$$x_i = A + cz_i$$

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

$$y_i = \alpha + \beta(A + cz_i) + \gamma z_i + u_i$$

$$y_i = \underbrace{\alpha + \beta A}_{\text{חותך}} + \underbrace{(\beta c + \gamma)}_{\text{שיפוע}} z_i + u_i$$

כלומר, המקדם של Z כולל רכיב של X, ולא ניתן לדעת כמה כל אחד מהם תרם באמת לשינוי ב-y.

המולטי קוליניאריות לא חייבת להיות מושלמת כדי שתעמוד בפנינו בעיה: מספיק שהיא תהיה מספיק גבוהה, כדי שנקבל שונות אסטרונומית, איתה לא נוכל לעבוד. לא נוכל לדחות אף השערה!

$$-t \le \frac{\hat{\beta} - \text{השערה}}{\hat{\sigma}_{\hat{\beta}}^{\,2}} \le t$$

אף רווח סמך לא יוכל לדחות את השערת האפס שהפרמטר $\beta=0$ או $\gamma=0$ בוודאות מספיקה. אף חוקר לא יוכל להתחייב על קשר בין אף אחד מהפרמטרים לתזוזה ב-y. אבל y זז!

כדי שנוכל בכל זאת לבדוק את ההשפעה של הפרמטרים על המשתנה התלוי, ננטש את מבחן t ונעבור למבחן f, שיודע לבדוק יותר מהשערה אחת בו זמנית.

$$H_0: \beta = \gamma = 0$$

$$H_1: \beta \neq 0 \lor \gamma \neq 0$$

מבחן כזה יוכל לומר לנו אם יש קשר בין β ו-γ ל-y – אבל לא יוכל לומר לנו **מי** הוא זה שהשפיע. כלומר, לא נוכל לדעת את ההשפעה בנפרד, אבל לפחות נוכל להסיק מסקנה לגבי ההשפעה המשותפת.

מבחן F

במבחן F מריצים מספר רגרסיות.

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

זהו המודל הלא מוגבל, unrestricted, ואותו מריצים ראשון.

לאחר מכן מריצים מודל מוגבל (במקרה שלנו, מוגבל ע"י כך ש- β=γ=0):

$$y_i = \alpha + \epsilon_i$$

המבחן עובד כך:

$$\frac{SSE^R - SSE^{UN}}{\text{מספר המגבלות}}$$

התוצאה תהיה, בכמה הטעויות גדלו עבור כל מגבלה.

כעת ננרמל את התוצאה, כדי שנקבל את השינוי באחוזים; ומחלקים במספר התצפיות פחות מספר הפרמטרים, על מנת לישר את הסקאלה (את סדר הגודל).

$$\frac{\frac{SSE^R - SSE^{UN}}{\text{מספר המגבלות}}}{\frac{\mathrm{SSE^{UN}}}{\mathrm{I} - \text{מספר פרמטרים במודל הלא מוגבל}}} = \frac{SSE^R - SSE^{UN}}{\text{מספר המגבלות} \cdot \hat{\sigma}_u^2} \sim F_{\alpha}^{\text{מונה, מכנה}}$$

α – רמת הוודאות.

מונה – מספר המגבלות.

מכנה – מספר התצפיות פחות הפרמטרים במודל הלא מוגבל.

התוצאה של F תמיד תהיה חיובית (מבחן ריבועי).

אפשר לבדוק, למשל, את ההשערה:

$$H_0: \beta = \gamma = 5$$

$$H_1: \beta \neq 5 \vee \gamma \neq 5$$

במקרה כזה מריצים את המודל הלא מוגבל, ואת המודל המוגבל:

$$y_i = \alpha + 5x_i + 5y_i + \epsilon_i$$

אך המחשב לא יודע איך להריץ רגרסיה כזו, ולכן מיצרים משתנה חדש ועליו מריצים רגרסיה:

$$k_i = \alpha + \epsilon_i$$

$$k_i = y_i - 5x_i - 5z_i$$

מבחן F עם מגבלה אחת

אנו רוצים להוכיח כי כאשר בודקים רק מגבלה אחת עם F, מקבלים את אותה תוצאה כמו במבחן t.

$$F \stackrel{?}{=} t^2$$

$$\frac{\dfrac{SSE^R - SSE^{UN}}{\text{מספר המגבלות}}}{\dfrac{\mathrm{SSE^{UN}}}{\mathrm{I} - \text{מספר פרמטרים במודל הלא מוגבל}}} = F \stackrel{?}{=} t^2 = \frac{\hat{\beta}^2}{\hat{\sigma}^2_{\hat{\beta}}} = \frac{\hat{\beta}^2}{\dfrac{\dfrac{\sum \hat{u}_i^2}{I-2}}{\sum (x_i - \bar{x})^2}}$$

$$y_i = \alpha + \beta x_i + u_i$$

$$H_0: \beta = 0$$

$$H_1: \beta \neq 0$$

הרגרסיה המוגבלת:

$$y_i = \alpha + \epsilon_i$$

$$\min_{\tilde{\alpha}} \sum (y_i - \tilde{\alpha})^2$$

$$\sum (y_i - \hat{\alpha}) = 0 \Rightarrow \ \hat{\alpha} = \bar{y}$$

$$SSE^R = \sum (y_i - \hat{y}_i)^2 = \sum \left(y_i - \underbrace{\bar{y}}_{\hat{\alpha}} \right)^2 = \sum \left(\hat{\alpha} + \hat{\beta} x_i + \hat{u}_i - \left(\hat{\alpha} + \hat{\beta} \bar{x} \right) \right)^2$$
$$= \hat{\beta}^2 \sum (x_i - \bar{x})^2 + \underbrace{\sum \hat{u}_i^2}_{SSE^{UN}} + 2\hat{\beta} \underbrace{\sum (x_i - \bar{x}) \hat{u}_i}_{=0}$$

נציב חזרה במשתנה F:

$$F = \frac{\hat{\beta}^2 \dfrac{\sum (x_i - \bar{x})^2}{1}}{\dfrac{SSE^{UN}}{I - \text{מספר פרמטרים}}} = \widehat{\beta^2} \frac{\sum (x_i - \bar{x})^2}{\left(\dfrac{\sum u_i^2}{I-2} \right)} = t^2$$

והגענו חזרה להתחלה, כלומר, הוכחנו את הנדרש.

עם זאת, הוכחנו את הנדרש עבור מבחן בו יש רק α ו-β! זה בגלל שבמודל המכיל פרמטרים נוספים, השונות של β משתנה, וההוכחה שהצגנו כבר לא רלוונטית. לעשות הוכחה מפורטת על מודלים בעלי פרמטרים מרובים תהיה מורכבת, וננסה לחשוב על זה בצורה לוגית.

למשל, זהו המודל המלא שלנו:

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

נבנה משתנה חדש:

$$k_i = y_i - \hat{\gamma} z_i$$

ועכשיו שוב יש לנו מודל בו יש רק β:

$$k_i = \alpha + \beta x_i + u_i$$

ולכן כל המסקנות הקודמות שלנו תקפות גם למודלים מרובי משתנים!

בדיקת כל הפרמטרים בו זמנית

ישנו מקרה פרטי בו כל הפרמטרים שווים ל-0 מלבד החותך. במקרה כזה לא חייבים להריץ שתי רגרסיות - נפתח קיצור:

$$y_i = \alpha + \epsilon_i$$

$$\hat{\alpha} = \bar{y}$$

$$SSE^R = \sum (y_i - \bar{y})^2 = SST$$

$$SST - SSE^{UN} = SSR^{UN}$$

$$\frac{\frac{SSE^R - SSE^{UN}}{\text{מגבלות}}}{\frac{\mathrm{SSE^{UN}}}{\mathrm{I} - \text{פרמטרים}}} = \frac{\frac{SSR^{UN}}{\text{מגבלות}}}{\frac{SSE^{UN}}{I - \text{פרמטרים}}} \overset{\cdot\frac{\frac{1}{SST}}{\frac{1}{SST}}}{=} \frac{\frac{R^2}{\text{מגבלות}}}{\frac{1 - R^2}{I - \text{פרמטרים}}} \sim F$$

רגרסיה עם משתנה ריבועי

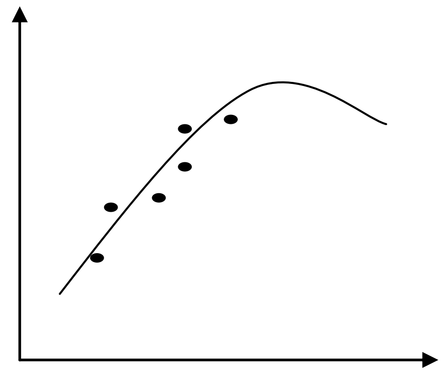
ברגרסיה שיש בה משתנה מסביר אחד, הקו יכול להיות לינארי עולה ויורד בלבד. אבל זה לא תמיד מתאר את המציאות – למשל, בשכר כפונקציה של גיל, ישנה עליה, ובנקודה מסוימת – ירידה. במקרה כזה כדאי להשתמש ברגרסיה עם מודל כזה (X – גיל, Y – שכר):

$$y_i = \alpha + \beta x_i + \gamma x_i^2 + u_i$$

ואז נוצר שיפוע שתלוי בגיל:

$$\frac{\partial y_i}{\partial x_i} : \beta + 2\gamma x_i$$

במקרה של מדגם שבו אין ירידה בשכר, אך נתנו מודל ריבועי, המחשב עדיין ייתן קו שיורד – אחרי התצפית האחרונה. יש להיזהר עם הסקת מסקנות מרגרסיות כאלה, ולבצע תחזיות רק על גילאים באיזור שבו היו תצפיות.



טעויות ספציפיקציה

טעויות ספציפיקציה – טעויות ניסוח, מודל שאינו נכון. אנו נתעסק בשני מקרים:

1. שכחנו להכניס משתנה חשוב- השמטת משתנה רלוונטי.
2. הכנסנו משתנה שאינו חשוב – הכנסת משתנה לא רלוונטי.

<u>השמטת משתנה רלוונטי</u>

נניח שהמודל הנכון הוא:

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

אבל בטעות אמדו את המודל הבא:

$$y_i = \delta_0 + \delta_1 x_i + \epsilon_i$$

השאלה הנשאלת היא, האם δ_1 עדיין נכונה, גם לאור המודל המקורי? כלומר, נשאל האם:

$$E(\hat{\delta}_1) = \beta$$

$$\hat{\delta}_1 = \frac{\sum(x_i-\bar{x})y_i}{\sum(x_i-\bar{x})^2} = \frac{\sum(x_i-\bar{x})(\alpha+\beta x_i+\gamma z_i+u_i)}{\sum(x_i-\bar{x})^2}$$
$$= \beta + \gamma\frac{\sum(x_i-\bar{x})z_i}{\sum(x_i-\bar{x})^2} + \frac{\sum(x_i-\bar{x})u_i}{\sum(x_i-\bar{x})^2}$$

$$\gamma\frac{\sum(x_i-\bar{x})z_i}{\sum(x_i-\bar{x})^2} = \frac{\gamma cov(x_i,z_i)}{var(x_i)}$$

ולכן:

$$E(\hat{\delta}_1) = \beta + \frac{\gamma cov(x_i,z_i)}{var(x_i)}$$

לכן, כפי שהסקנו בעבר, אם אין קשר בין X ל-Z, אין הטיה והפרמטר נשאר זהה. אם יש קשר, הפרמטר זז, וספג לתוכו גם השפעה של הפרמטר שהושמט.

<u>הכנסת משתנה לא רלוונטי</u>

המודל הנכון הוא:

$$y_i = \delta_0 + \delta_1 x_i + \epsilon_i$$

והמודל הנאמד הוא:

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

$$\hat{\beta} = \frac{\sum\big((x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}\big)y_i}{M_{xx}M_{zz} - M_{xz}^2} = \frac{\sum\big((x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}\big)(\delta_0 + \delta_1 x_i + \epsilon_i)}{M_{xx}M_{zz} - M_{xz}^2} = \delta_1 + \frac{\sum\big((x_i - \bar{x})M_{zz} - (z_i - \bar{z})M_{xz}\big)\epsilon_i}{M_{xx}M_{zz} - M_{xz}^2}$$

לכן, כשמוסיפים משתנים אקסטרה, המשתנים שגם ככה היו אמורים להיות במודל – הם חסרי הטיה. המחיר במקרה זה מגיע בשונות. במקום לקבל שונות "רגילה" של:

$$var\big(\hat{\delta}_1\big) = \frac{\sigma_u^2}{\sum(x_i - \bar{x})^2}$$

מקבלים שונות גדולה יותר של:

$$var\big(\hat{\beta}\big) = \frac{\sigma_u^2}{M_{xx}\big(1 - \rho_{x,z}^2\big)}$$

המסקנה היא שתוספת משתנים מיותרים גורמת למבחני השערות לא יעילים.

<u>הקשר בין שני המודלים</u>

$$y_i = \delta_0 + \delta_1 x_i + \epsilon_i$$

$$y_i = \alpha + \beta x_i + \gamma z_i + u_i$$

האם יש קשר מספרי מחייב בין המשתנים בשני המודלים? נציב את האומדים זה בזה:

$$\hat{\delta}_1 = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \frac{\sum(x_i - \bar{x})\big(\hat{\alpha} + \hat{\beta}x_i + \hat{\gamma}z_i + \hat{u}_i\big)}{\sum(x_i - \bar{x})^2} = \hat{\beta} + \hat{\gamma}\frac{\sum(x_i - \bar{x})z_i}{\sum(x_i - \bar{x})^2} + \frac{\overbrace{\sum(x_i - \bar{x})\hat{u}_i}^{cov(x,u)=0}}{\sum(x_i - \bar{x})^2}$$

$$\hat{\boldsymbol{\delta}}_1 = \hat{\boldsymbol{\beta}} + \hat{\boldsymbol{\gamma}} \frac{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})\boldsymbol{z_i}}{\sum(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}$$

צורות שונות של משוואת הרגרסיה

$$y_i = \alpha x_i^{\beta} k_i^{\gamma} + e^{u_i}$$

$$\ln(y_i) = \ln(\alpha) + \beta\ln(x_i) + \gamma\ln(k_i) + u_i$$

אם β=2, אז כל תזוזה של אחוז ב-X, Y תזיז את Y בשני אחוזים.

$$\ln(y_i) = \alpha + \beta x_i + \gamma z_i + u_i$$

במקרה זה, β מתאר את אחוז השינוי ב-Y עבור כל יחידת שינוי ב-X.

בכל אחד מהמודלים האלה, ההשפעה של X על Y אינה ליניארית.

נניח שאנחנו רוצים לבדוק את השפעת ההשכלה והגיל על השכר. אך אנו מאמינים שהשפעת ההשכלה *תלויה בגיל* (נניח, מגיל 30 עד 65 יש השפעה גדולה של השכר על גובה ההכנסה). **יתכן שההשפעה של משתנה מסוים תהיה תלויה במשתנים אחרים!**

$$y_i = \alpha + \beta x_i + \gamma z_i + \delta x_i z_i + u_i$$

$$\frac{\partial y_i}{\partial x_i} = \beta + \delta z_i$$

לפי הגזירה ברור שההשפעה של X תלויה ב-Z. אם נדבר במונחי הדוגמה, ההשפעה של ההשכלה, תלויה באיזה גיל אתה. באופן דומה:

$$\frac{\partial y_i}{\partial z_i} = \gamma + \delta x_i$$

גם במודל ה-lnי מתקיים הקשר הזה: הפרמטר β מבטא אחוזי שינוי ב-y, אבל ה-y בעצמו נקבע על ידי כל המשתנים המסבירים.

משתנים איכותיים

נניח y מתאר הוצאה כספית חודשית של משפחה.

$$y_i = \alpha + u_i$$

לפי המודל הזה, אוכל רק לתאר את ממוצע ההוצאה.

כעת נניח כי קיימים שלושה סוגי משפחה – רווקים, נשואים וגרושים. אנו מאמינים כי דפוסי ההוצאות של כל אחד מסוגי המשפחה שונה.

הפתרון לפיו מייצרים X המקבל 1 עבור רווקים, 2 לנושאים ל-3 לגרושים לא נכון. אנו כופים כך סדר פנימי (הנשואים תמיד יהיו בין הרווקים לגרושים) ועוצמת שינוי (בחירת מספר גדול יותר לאינדקס תגרור שינוי משמעותי במספרים). לכן, עבור כל מצב משפחתי אנו מייצרים משתנה מסביר, אותו נסמן בד"כ ב- d (dummy).

$$D_{1i}, D_{2i}, D_{3i}$$

המשתנה D_{1i} יקבל 1 עבור רווקים, ו-0 אחרת. כנ"ל לשני המשתנים האחרים, עבור נשואים וגרושים. בשיטה הזו, זה לא משנה אם נרשום 8 במקום 1: זהו פשוט שינוי סקאלה, ונצטרך לחלק את האומד ב-8.

$$y_i = \alpha_1 + \alpha_2 D_{1i} + \alpha_3 D_{2i} + \alpha_4 D_{3i} + u_i$$

את המודל הזה, כפי שהוא כתוב כרגע, לא ניתן לאמוד!

$$min\sum(y_i - \tilde{\alpha}_1 - \tilde{\alpha}_2 D_{1i} - \tilde{\alpha}_3 D_{2i} - \tilde{\alpha}_4 D_{3i})^2$$

$$(I)\frac{\partial}{\partial\tilde{\alpha}_1}: -2\sum\hat{u}_i = 0$$

$$(II)\frac{\partial}{\partial\tilde{\alpha}_2}: -2\sum\hat{u}_i D_{1i} = 0$$

$$(III)\frac{\partial}{\partial\tilde{\alpha}_3}: -2\sum\hat{u}_i D_{2i} = 0$$

$$(IV)\frac{\partial}{\partial\tilde{\alpha}_4}: -2\sum\hat{u}_i D_{3i} = 0$$

$$(II) + (III) + (IV) = \sum\hat{u}_i \underbrace{(D_{1i} + D_{2i} + D_{3i})}_{\text{קבוע } 1} = 0$$

$$const\sum\hat{u}_i = 0 = (I)$$

קיבלנו מולטיקולינאריות מושלמת. מדוע זהו המצב?

$$\text{רווק} = \alpha_1 + \alpha_2$$

$$\text{נשוי} = \alpha_1 + \alpha_3$$

$$\text{גרוש} = \alpha_1 + \alpha_4$$

אם α_1 מופיע בכולם, כיצד נוכל להפריד ולומר מה ההשפעה של החלק הייחודי לכל קבוצה? אם כולם שונים מהנורמה, אז מהי הנורמה?

לכן, נוכל לאמוד שלושה פרמטרים בלבד, ולא ארבעה. ישנן שלוש שיטות לעשות זאת.

שיטה 1

$$y_i = \alpha_1 D_{i1} + \alpha_2 D_{2i} + \alpha_3 D_{3i} + u_i$$

$$min\sum(\hat{u}_i)^2$$

$$(I)\sum(y_i - \hat{\alpha}_1 D_{1i} - \hat{\alpha}_2 D_{2i} - \hat{\alpha}_3 D_{3i})D_{1i} = 0$$

$$(II)\sum(y_i - \hat{\alpha}_1 D_{1i} - \hat{\alpha}_2 D_{2i} - \hat{\alpha}_3 D_{3i})D_{2i} = 0$$

$$(III)\sum(y_i - \hat{\alpha}_1 D_{1i} - \hat{\alpha}_2 D_{2i} - \hat{\alpha}_3 D_{3i})D_{3i} = 0$$

כל משוואה נורמלית מייצגת מצב משפחתי אחד- כי המצבים המשפחתיים האחרים מתאפסים. כך, אם הנבדק בתצפית שלנו הוא רווק, נקבל:

$$\sum(y_i - \hat{\alpha}_1) = 0 \Rightarrow \hat{\alpha}_1 = \bar{y}_{\text{רווקים}}$$

וכך, כל פרמטר מייצג את ממוצע ההכנסה של כל אוכלוסיה.

שיטה 2

נשמיט את אחת מהקבוצות מהרגרסיה:

$$y_i = \beta_1 + \beta_2 D_{1i} + \beta_3 D_{2i} + u_i$$

בשיטה הזו, הקבוצה ה"נורמטיבית" היא הגרושים – אליה מתייחסות הקבוצות הנותרות.

$$\hat{\beta}_1 = \bar{y}_{\text{גרושים}}, \qquad \hat{\beta}_1 + \hat{\beta}_2 = \bar{y}_{\text{רווקים}}, \qquad \hat{\beta}_1 + \hat{\beta}_3 = \bar{y}_{\text{נשואים}}$$

התוצאה המתקבלת בשתי השיטות היא זהה לחלוטין! ההבדל הוא, בנוחיות העבודה בבדיקת השערות. בשיטה הראשונה, זו תהיה השערת המחקר שלנו:

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

הבעיה במקרה זה היא שלא נוכל להשתמש בבדיקה המקוצרת שלמדנו על מבחני F, במקרה בו מאפסים את כל הפרמטרים – וזה כי אין לנו חותך.

עם זאת, בשיטה השניה:

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

במקרה זה, למרבה האושר, כן נוכל להשתמש בבדיקת F המקוצרת באמצעות R^2! הידד!

במידה ובחינה זו נכשלה, נרצה לבדוק אם נוכל "לחסל" פרמטרים מהמודל המקורי: כלומר, יתכן ואין בכלל הבדל בהכנסה בין רווקים לנשואים, כך נוכל להיפטר מאחד הפרמטרים ולהקטין את שונות האומדים. כדי לעשות זאת נשתמש במבחן t רגיל, על כל אחד מהפרמטרים. בשיטה השניה, עושים זאת בקלות:

$$-t \leq \frac{\hat{\beta}_2 - 0}{\hat{\sigma}_{\hat{\beta}_2}} \leq t$$

אבל, בשיטה הראשונה והמבאסת, זה יהיה הסטטיסטי שלנו:

$$\frac{\hat{\alpha}_1 - \hat{\alpha}_2}{\sqrt{\widehat{var}(\hat{\alpha}_1 - \hat{\alpha}_2)}}$$

שהוא, כמובן, הרבה יותר קשה לחישוב.

בשיטה השניה, לקבוצה המושמטת נקרא "קבוצת ההתייחסות" (Reference Group).

שיטה 3

בשיטה זו, ניתן פשוט להריץ רגרסיות נפרדות על כל חלק של המדגם המייצג קבוצה נבדקת. התוצאה תצא זהה לחלוטין לזו של השיטות הקודמות, אבל נקבל שונות אחרת. בשיטות 1 ו-2, זו תהיה שונות הטעות:

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{I - 3}$$

אבל בשיטה הזו, נקבל שונות גדולה יותר:

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_i^2}{\text{רווקים} - 1}$$

אם השונויות של הקבוצות במדגם שווה, אפשר לחשב שונות אחת לכולם יחד, וכך להימנע מההפחתה ברמת הדיוק. עם זאת, אם השונות אינה שווה בין הקבוצות, חייבים לפצל את חישוב השונות.

מה היה קורה לו היינו משמיטים שני משתנים איכותיים?

$$y_i = \beta_1 + \beta_2 D_{2i} + u_i$$

במקרה כזה, β_1 יהווה הממוצע של שני המשתנים שהושמטו, ו-β_2 יהיה ההבדל מהממוצע. באופן עקרוני, במקרים רבים נצטרך להשמיט משתנים כאלה כדי למנוע מהרגרסיה להתנפח לגדלים עצומים.

מודל הכולל משתנים איכותיים עם משתנים כמותיים

נניח היה לנו המודל ההתחלתי,

$$y_i = \alpha_0 + \alpha_1 x_1 + \epsilon_i$$

כעת נניח שאנו סבורים שהשפעת משתנה איכותי, למשל, מצב משפחתי, משנה את השפעת המשתנים הכמותיים, למשל גיל.

המודל בשיטה 1:

$$y_i = \beta_1 D_{1i} + \beta_2 D_{2i} + \beta_3 D_{3i} + \beta_4 D_{1i} x_i + \beta_5 D_{2i} x_i + \beta_6 D_{3i} x_i + u_i$$

כל מה שמסומן ב-D_{1i} בא לידי ביטוי רק במקרה של רווקים: ולכן החותכים והשיפועים שיחושבו באמצעות β_1 ו-β_4 יהיו רלוונטים רק למקרה זה.

המודל בשיטה 2:

$$y_i = \gamma_1 + \gamma_2 D_{2i} + \gamma_3 D_{3i} + \gamma_4 x_i + \gamma_5 D_{2i} x_i + \gamma_6 D_{3i} x_i + u_i$$

בשיטה השניה, γ_2 ו-γ_5 מבטאים את ההפרש בין נשוי לרווק (כלומר, ההפרד ביניהם לבין γ_1 ו-γ_4 בהתאמה). אין חובה שהקבוצה המושמטת תהיה זהה בחותך ובשיפוע. התוצאות בין המודלים של שתי השיטות יהיו כמובן זהות:

$$\hat{\beta}_2 = \hat{\gamma}_1 + \hat{\gamma}_2, \hat{\beta}_5 = \hat{\gamma}_4 + \hat{\gamma}_5$$

וכן הלאה.

אחרי שנאמוד את המודל, נרצה לבדוק:

$$H_0: \gamma_2 = \gamma_3 = \gamma_5 = \gamma_6 = 0$$

בדיקה זו מוודאת שאכן יש הבדל בין הקבוצות. המודל המוגבל שלנו יראה כך:

$$y_i^R = \alpha_0 + \alpha_1 x_i + \epsilon_i$$

ריבוי משתנים איכותיים

יכול להיות מצב שבו ארצה להפריד את הרגרסיה שלי, למצב שבו יש הבדל בין גבר ואישה, וגם הבדל בין מי שגר בעיר למי שגר בכפר. במקרה כזה, יהיו שתי קטגוריות של משתני Dummy: אחת למין, והשניה למגורים. נגדיר:

$$D_{1i} = 1 \text{ אישה}, 0 \text{ אחרת}; D_{2i} = 1 \text{ גבר}, 0 \text{ אחרת}$$

$$F_{1i} = 1 \text{ עיר}, 0 \text{ אחרת}; F_{2i} = 1 \text{ כפר}, 0 \text{ אחרת}$$

$$y_i = \underbrace{\beta_1 D_{1i} F_{1i}}_{\text{אישה בעיר}} + \underbrace{\beta_2 D_{1i}(1 - F_{1i})}_{\text{אישה בכפר}} + \underbrace{\beta_3 (1 - D_{1i}) F_{1i}}_{\text{גבר בעיר}} + \underbrace{\beta_4 (1 - D_{1i})(1 - F_{1i})}_{\text{גבר בכפר}} + \epsilon_i$$

ניתן לכתוב את המודל גם בשיטת קבוצת ההתייחסות:

$$y_i = \gamma_1 + \gamma_2 D_{1i} + \gamma_3 F_{1i} + \gamma_4 D_{1i} F_{1i} + \epsilon_1$$

גבר בכפר – γ_1

גבר בעיר – $\gamma_1+\gamma_3$

אישה בכפר – $\gamma_1+\gamma_2$

אישה בעיר – $\gamma_1+\gamma_2+\gamma_3+\gamma_4$

איך בודקים השערות באמצעות מודל זה?

ההוצאה לאישה ולגבר בכפר שונה. כדי לבדוק השערה זו, נבדוק אם $\gamma_2=0$.

במעבר מהכפר לעיר, השינוי בהוצאה של גברים ונשים שווה. כדי לבדוק השערה זו, נבדוק אם $\gamma_4=0$.

הסרת ההנחות

משתנה מסביר מקרי

$$y_i = \alpha + \beta x_i + u_i$$

עד עכשיו הנחנו ש-X אינו משתנה מקרי. כעת נסיר את ההנחה הזו. המשמעות: אם נדגום את אותו אדם בין ימים שונים, נקבל X אחר בכל פעם. בשיטה הקודמת שמרנו את X קבוע בין המדגמים, ולכן אם משהו הזיז את y, זה היה ה-u. כעת הכל זז והדברים מסתבכים.

ההבדל בין X מקרי ללא מקרי בא לידי ביטוי כשעוסקים בנושא התוחלת – הרי היא זו שמתעסקת במה שקורה *בין מדגמים*.

$$\hat{\beta} = \frac{\sum(x_i - \bar{x})y_i}{\sum(x_i - \bar{x})^2} = \beta + \frac{\sum(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}$$

את המעבר שעשינו כשההנחה היתה בפועל, במסגרתה הוצאנו את X מחוץ לתוחלת, כבר לא ניתן לעשות.

$$E(\hat{\beta}) = \beta + \sum\left[E\left(\frac{(x_i - \bar{x})u_i}{(x_i - \bar{x})^2}\right)\right] \overset{x,\ u\ \text{בלתי תלויים}}{=} \beta + \sum\left[E\left(\frac{x_i - \bar{x}}{\sum(x_i - \bar{x})^2}\right) \cdot \underbrace{E(u_i)}_{0}\right]$$

אפשר לראות שאם X ו-u אכן בלתי תלויים, מקבלים שוב אומד חסר הטיה. אבל אם X ו-u משתנים תלויים, $\hat{\beta}$ הינו אומד מוטה ל-β. כש-X היה לא מקרי, זה הבטיח לנו שהם היו בלתי תלויים ובלתי מתואמים – לצערנו זה כבר לא המצב.

מעתה נעבור לסמן את המדגמים ב-t במקום ב-i- הסיבה היא שהמודלים בהם X מקרי הם בדרך כלל בחתך אורך.

אם כך, ראינו כי בתוחלת אין בעיה. אך, מה קורה בשונות?

$$var(\hat{\beta}) = E\left(\hat{\beta} - \underbrace{E(\hat{\beta})}_{=\beta}\right) = E\left[\frac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}\right]^2 = E\left[\sum\left(\frac{(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}\right)^2\right]$$

$$= E\left[\frac{\sum(x_t - \bar{x})^2 u_t^2}{(\sum(x_t - \bar{x})^2)^2}\right] \overset{x,\ u\ \text{בלתי תלויים}}{=} E\left[\frac{\sum(x_t - \bar{x})^2}{(\sum(x_t - \bar{x})^2)^2}\right] \cdot E[u_t^2]$$

$$var(\hat{\beta}) = \frac{\sigma_u^2}{\sum(x_t - \bar{x})^2}$$

במצב הנוכחי לא נוכל לדעת מה השונות של β, כיוון שאין לנו את השונות של X. לכן במקום זאת, מחשבים **שונות מותנה:** זוהי לא שונות אמיתית כיוון שהיא תהיה שונה בכל מדגם.

תלות בין x ל-u

ישנם שני סוגים של תלות בין X ל-u:

1. תלות בין X_t ל-u_t (תלות בו זמנית)

2. תלות בין u_t ל-x_{t-j} (תלות בין דורית)

התלות מסוג (1) (תלות במונה) חמורה הרבה יותר מהתלות מסוג (2).

במקרה של תלות מכל אחד מהסוגים, שיטת אומד הריבועים הפחותים, לצערנו, תמיד תוביל אותנו לאומד מוטה. אנו צריכים למצוא שיטה כדי שהאומד לא יהיה מוטה.

מודל ההתאמה החלקית

כדי להבין את מודל זה, נשתמש במודל שהוא קצת מחוץ לעולם הכלכלה: מודל חקלאי.

באמצעות y_t^* נסמן מה היבול הרצוי.

$$y_t^* = \alpha + \beta x_t + u_t$$

היבול שנראה בפועל, יהיה סוג של שקלול בין היבול הרצוי ליבול שנתקבל בתקופה הקודמת.

$$y_t = \delta y_t^* + (1-\delta)y_{t-1} = \delta(\alpha + \beta x_t + u_t) + (1-\delta)y_{t-1}$$

במודל שנתקבל, y_{t-1} הוא משתנה מקרי שכולל הפרעה אקראית (שכן הוא תלוי ב-${y^*}_{t-1}$). לא נוכל להפריד כאן בין המשתנה המסביר להפרעות האקראיות, ונקבל אומד מוטה.

נפשט ונעבוד עם המודל הבא:

$$y_t = \beta y_{t-1} + u_t$$

למודל זה אנו קוראים **מודל דינאמי** – יש בו התפתחות על ציר הזמן. נראה מה קורה כשמנסים לפתור אותו באמצעות OLS:

$$\hat{\beta} = \frac{\sum y_{t-1} y_t}{\sum_{t=2}^{T} y_{t-1}^2} = \frac{\sum y_{t-1}(\beta y_{t-1} + u_t)}{\sum y_{t-1}^2} = \beta + \frac{\sum y_{t-1} u_t}{\sum y_{t-1}^2}$$

$$E(\hat{\beta}) = \beta + \sum \left[E\left(\frac{\sum y_{t-1} u_t}{\sum y_{t-1}^2} \right) \right]$$

במונה, אין בעיה: אין שום תלות בין y_{t-1} ל-u_t. זאת כיוון שאנו מניחים שאין מתאם בין ההפרעה האקראית בין הימים. הבעיה מופיעה בין המונה למכנה.

נפתח את המכנה:

$$y_1^2 + y_2^2 + y_3^2 + \cdots = (\beta y_0 + u_1)^2 + (\beta y_1 + u_2)^2 + (\beta y_2 + u_3)^2 + \cdots$$

בעצם, במכנה מופיעים כל ה-u! לכן ברור כי יש מתאם בין המונה למכנה, *והאומד מוטה*. מעתה ועד סוף הקורס, אם יש אומד מוטה, לא נוכל למצוא אומד חסר הטיה. נוכל רק לנסות למצוא אומד מוטה פחות מה- OLS.

כל אחד מה-y המופיע במכנה, כולל בתוכו את כל ה-u לתקופות שקדמו לו, כל אחד מוכפל ב-β בחזקה המתאימה למספר התקופות שחוזרים אחורה. אם |β|<1, ככל שמוסיפים תצפיות, כל u יקבל עוצמה הולכת ופוחתת, והמתאם בין המונה והמכנה יהיה קטן – ההטיה הולכת ושואפת לאפס.

אנו מאמינים שהכלכלה היא סטציונרית – השינויים בה נעשים בהדרגה, ולכן, השונות נשארת קבועה לאורך התקופות. תחת הנחה זו, נוכיח כי בהכרח 0<|β|<1.

<u>דרך 1</u>

$$var(y_t) = var(\beta y_{t-1} + u_t) = \beta^2 var(y_{t-1}) + \sigma_u^2 + \overset{cov(y_{t-1},u_t)=0}{0}$$

אנו רוצים שהשונות תהיה זהה בכל התקופות. לכן נסיר את סימון התקופה t.

$$var(y) = \frac{\sigma_u^2}{1-\beta^2} > 0 \Rightarrow |\beta| < 1$$

תכונה זו גורמת לכך שאירוע חד פעמי דועך עם הזמן ולא גדל.

<u>דרך 2</u>

$$y_t = \beta y_{t-1} + u_t$$

$$y_{t-1} = \beta y_{t-2} + u_{t-1}$$

$$y_t = \beta^2 y_{t-2} + \beta u_{t-1} + u_t$$

$$y_t = \beta^3 y_{t-3} + \beta^2 y_{t-2} + \beta u_{t-1} + u_t$$

$$y_t = \beta^n y_{t-n} + \beta^{n-1} u_{t-(n-1)} + \cdots + \beta^2 u_{t-2} + \beta u_{t-1} + u_t$$

כלומר, ה-y של היום נקבע לפי y של פעם, בתוספת ההפרעות האקראיות. אם β בערך מוחלט לא יהיה קטן מאחד, ההשפעה של העבר הרחוק תהיה הרבה יותר דרמטית על ההווה מאשר אירועים אחרונים.

בכל מקרה, ההוכחה הזו מביאה לנו תועלת כאשר המתאם הוא בין המונה למכנה, ואנו עדיין מניחים כי אין תלות בין y_1 ל-u_2.

עקיבות

נחלק את האומדים המוטים לשני סוגים:

1. מוטה ועקיב (consistent)
2. מוטה ולא עקיב

אומד עקיב – אומד, כשאשר המדגם שואף לאינסוף, הוא שואף לגודל האמיתי.

$$P\left[\left|\tilde{\theta}_n - \theta\right| > \epsilon\right] = \underset{n\to\infty}{0}$$

עקיבות קשורה להתנהגות האומד כשמוסיפים תצפיות למדגם, ולא כאשר מוסיפים מדגמים (כי זו תהיה ההגדרה של חוסר הטיה). תכונת העקיבות גוררת גם את התכונה הבאה:

$$var(\tilde{\theta}_n) = \underset{n \to \infty}{0}$$

ניתן לבדוק עקיבות גם אצל אומד שאינו מוטה.

$$y_i = \alpha + \beta x_i + u_i$$

האם α ו-β, אשר אינם מוטים, הם גם עקיבים? מתוך חוסר ההטיה, אנו יודעים שבין מדגמים, האומד שווה בממוצע לגודל האמיתי. מתוך עקיבות, אנו יודעים שלא קיימת שונות. חיבור שתי התכונות האלה גורר שהאומד יהיה שווה לגודל האמיתי, גם אם נבדק רק מדגם אחד (שגודלו שואף לאינסוף).

$$var(\hat{\beta}) = \frac{\sigma_u^2}{\sum(x_i - \bar{x})^2}$$

ככל שמוסיפים יותר תצפיות (שאינן שוות לממוצע), השונות הולכת וקטנה. מכאן, ה- OLS הינו גם חסר הטיה, וגם עקיב.

יתכן גם אומד חסר הטיה שאינו עקיב, לדוגמה:

$$y_i = \beta x_i + u_i$$

$$\tilde{\beta} = \frac{y_T - y_1}{x_T - x_1} = \frac{\beta x_T + u_T - \beta x_i - u_1}{x_T - x_1} = \beta + \frac{u_T - u_1}{x_T - x_1}$$

$$var(\tilde{\beta}) = E\left[\frac{u_T - u_1}{x_T - x_1}\right]^2 = \frac{1}{(x_T - x_1)^2} \cdot E[u_T - u_1]^2 = \frac{2\sigma_u^2}{(x_T - x_1)^2}$$

האומד $\tilde{\beta}$ הינו חסר הטיה אך אינו עקיב.

מסקנה: אין קשר ישיר בין עקיבות לחוסר הטיה – אחד עוסק בהוספת מדגמים, השני בהוספת תצפיות.

מתאם בין ההפרעות האקראיות

נחזיר את X להיות משתנה שאינו מקרי, ונסיר את ההנחה כי אין קשר בין ההפרעות האקראיות.

$$y_t = \alpha + \beta x_t + u_t$$

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

כעת כבר לא נוכל להניח שתוחלת u היא אפס – אך את ההנחות הישנות על u, נוכל להניח כעת על ϵ.

$$E(\epsilon_t) = 0, \qquad cov(\epsilon_t, \epsilon_{t-j}) = 0, \qquad var(\epsilon_t) = \sigma_\epsilon^2, \qquad \epsilon_t \sim N$$

ראשית נבדוק אם האומד שלנו מוטה:

$$\hat{\beta} = \frac{\sum(x_t - \bar{x})y_t}{\sum(x_t - \bar{x})^2} = \frac{\sum(x_t - \bar{x})(\alpha + \beta x_t + u_t)}{\sum(x_t - \bar{x})^2} = \beta + \frac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}$$

$$E(\hat{\beta}) = \beta + \frac{\sum(x_t - \bar{x}) \cdot E(u_t)}{\sum(x_t - \bar{x})^2}$$

$$u_t = \rho u_{t-1} + \epsilon_t = \rho^2 u_{t-2} + \rho\epsilon_{t-1} + \epsilon_t = \rho^3 u_{t-3} + \rho^2\epsilon_{t-2} + \rho\epsilon_{t-1} + \epsilon_t$$
$$= \rho^\infty u_{t-\infty} + \sum_{i=0}^{\infty-1} \rho^i \epsilon_{t-i}$$

$$E(u_t) = \overbrace{\rho^\infty E(u_{t-\infty})}^{\to 0} + \overbrace{\sum_{i=0}^{\infty-1} \rho^i \overbrace{E(\epsilon_{t-i})}^{=0}}^{=0} = 0$$

אנו מאמינים שהעבר פחות דומיננטי בקביעת ההווה. לכן $|\rho|<1$, והאיבר השמאלי שואף לאפס. בסופו של דבר, $\hat{\beta}$ נשאר חסר הטיה.

לתוחלת כפי שתיארנו כאן, קוראים **תוחלת לא מותנה**. עם זאת, נוכל גם לשאול מה ה**תוחלת המותנה** של u:

$$E(u_t)|u_{t-1}$$

זאת כיוון, שכשיש לנו את ה-u של אתמול, יש לנו הרבה יותר אינפורמציה לגבי הגובה של u של היום – לכן אין סיכוי שהתוחלת שלו תהיה 0!

כעת נבדוק מה קרה לשונות, וכדי לעשות זאת, נעשה קצת עבודת הכנה:

שונות של u ומתאם בין תקופות שונות

$$var(u_t) = E\big(u_t - E(u_t)\big)^2 = E\left(\sum_{i=0}^{\infty-1} \rho^i \epsilon_{t-i}\right)^2$$

למען הפשטות, נפתח את הביטוי עבור שני גלגולים:

$$E(\rho^1\epsilon_{t-1} + \rho^2\epsilon_{t-2})^2 = \cdots + 2\rho^1\rho^2 \overbrace{E(\epsilon_{t-1}\epsilon_{t-2})}^{=0}$$

$$var(u_t) = \sum_{i=0}^{\infty-1} \rho^{2i} \cdot \underbrace{E(\epsilon_{t-i}^2)}_{=E\left(\epsilon_{t-1} - E(\epsilon_{t-1})\right)^2 = \sigma_\epsilon^2} = \sigma_\epsilon^2 \sum_{i=0}^{\infty-1} \rho^{2i} = \frac{\sigma_\epsilon^2}{1-\rho^2} = var(u)$$

קיבלנו כי יש שונות קבועה עבור כל ה-u שאינה משתנה עם התצפיות.

ניתן להוכיח זאת בדרך נוספת:

$$var(u_t) = var\left(\sum_{i=0}^{\infty-1} \rho^i \epsilon_{t-i}\right) = \sum_{i=0}^{\infty-1} \rho^{2i} \cdot var(\epsilon_{t-i}) = \sigma_\epsilon^2 \sum_{i=0}^{\infty-1} \rho^{2i} = \frac{\sigma_\epsilon^2}{1-\rho^2} = var(u)$$

אם כך, יש ל-u שונות קבועה, שאינה תלויה ב-t, והיא סופית.

כעת נחשב את השונות המשותפת של ההפרעה בין שתי תקופות:

$$\begin{aligned} cov(u_t, u_{t-1}) &= E\big[\big(u_t - E(u_t)\big)\big(u_{t-1} - E(u_{t-1})\big)\big] \\ &= E\left[\left(\rho u_{t-1} + \overbrace{\epsilon_t}^{\text{אין מתאם}} - \overbrace{E(u_t)}^{=E(u_{t-1})}\right)\big(u_{t-1} - E(u_{t-1})\big)\right] \\ &= \rho \cdot cov(u_{t-1}, u_{t-1}) = \boldsymbol{\rho \cdot var(u)} \end{aligned}$$

המסקנה – המתאם בין u של שתי תקופות עוקבות, חלש יותר מהפיזור של u.

$$cov(u_t, u_{t-2}) = cov(\rho^2 u_{t-2} + \rho\epsilon_{t-1} + \epsilon_t, u_{t-2}) = \rho^2 cov(u_{t-2}, u_{t-2}) = \rho^2 var(u)$$

והמסקנה הכללית הנגזרת:

$$cov\big(u_t, u_{t-j}\big) = \rho^j\, var(u)$$

ככל ששני איברים רחוקים יותר זה מזה, כך גם הקשר ביניהם חלש יותר.

נחשב את מקדם המתאם:

$$\text{מקדם המתאם}(u_t, u_{t-1}) = \frac{cov(u_t, u_{t-1})}{\sqrt{var(u_t) var(u_{t-1})}} = \frac{\rho \cdot var(u)}{var(u)} = \rho$$

לכן גם הקשר ביניהם מסומן מלכתחילה ב-ρ! ובאופן כללי:

$$\text{מקדם המתאם}(u_t, u_{t-j}) = \rho^j$$

שונות של $\hat{\beta}$

$$var(\hat\beta) = E\left(\hat\beta - E(\hat\beta)\right)^2 = E\left(\frac{\sum(x_t - \bar{x})u_t}{\sum(x_t - \bar{x})^2}\right)^2$$

נפתח את המונה באמצעות שלושה איברים:

$$\begin{aligned} &E[(x_1 - \bar{x})u_1 + (x_2 - \bar{x})u_2 + (x_3 - \bar{x})]^2 \\ &= \dots + 2(x_1 - \bar{x})(x_2 - \bar{x})\,\overbrace{E\big[\big(u_1 - E(u_1)\big)\big(u_2 - E(u_2)\big)\big]}^{\rho\sigma_u^2} \\ &+ 2(x_1 - \bar{x})(x_3 - \bar{x})\,\overbrace{E\big[\big(u_1 - E(u_1)\big)\big(u_3 - E(u_3)\big)\big]}^{\rho^2\sigma_u^2} \\ &+ 2(x_2 - \bar{x})(x_3 - \bar{x})E\big[\big(u_2 - E(u_2)\big)\big(u_3 - E(u_3)\big)\big] \end{aligned}$$

יש לשים לב כי ככל שמשתמשים ביותר איברים, נמצא איברים צולבים ביניהם יש מרחק גדול יותר, כלומר, חזקות גבוהות יותר של ρ. נחזיר את המכנה, ונרכז איברים דומים:

$$var(\hat{\beta}) = \overbrace{\frac{\sigma_u^2}{\sum(x_t-\bar{x})^2}}^{var(\hat{\beta})\text{ ללא מתאם}} + \frac{2\rho\sigma_u^2\cdot\sum_{t=2}^{T}(x_t-\bar{x})(x_{t-1}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} + \frac{2\rho^2\sigma_u^2\cdot\sum_{t=3}^{T}(x_t-\bar{x})(x_{t-2}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} + \frac{2\rho^3\sigma_u^2\cdot\sum_{t=4}^{T}(x_t-\bar{x})(x_{t-3}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} + \cdots + \frac{2\rho^{T-1}\sigma_u^2\cdot(x_1-\bar{x})(x_T-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2}$$

אפשר להוציא את השונות כגורם משותף:

$$var(\hat{\beta}) = \sigma_u^2\cdot\left[\frac{1}{\sum(x_t-\bar{x})^2} + \frac{2\rho\cdot\sum_{t=2}^{T}(x_t-\bar{x})(x_{t-1}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} + \frac{2\rho^2\cdot\sum_{t=3}^{T}(x_t-\bar{x})(x_{t-2}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} + \frac{2\rho^3\cdot\sum_{t=4}^{T}(x_t-\bar{x})(x_{t-3}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} + \cdots + \frac{2\rho^{T-1}\cdot(x_1-\bar{x})(x_T-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2}\right]$$

עד עכשיו, היינו סופרים בשונות רק את האיבר הראשון במשוואה הנ"ל. למעשה, במידה ובמציאות היה קשר בין ה-u, היינו שוכחים כמות גדולה מאוד של איברים, ומקבלים שונות לא נכונה! עם זאת, האיברים הולכים ונחלשים (עקב החזקה ההולכת ועולה של ρ). כעת היינו רוצים לדעת לאיזה כיוון מתכנסים האיברים הנוספים: האם השונות המקורית שחישבנו, ה"מוטעית", גדולה מהשונות האמיתית או לא? זו שאלה קריטית – היא משפיעה על הקביעה האם מדובר באומד שונות מינימלית; האם מדובר באומד מוטה; היא גם משפיעה על גודל רווחי הסמך שהשתמשנו בהם.

ניקח את האיבר הראשון:

$$\frac{2\rho\cdot\sum_{t=2}^{T}(x_t-\bar{x})(x_{t-1}-\bar{x})}{\sum(x_t-\bar{x})^2\cdot\sum(x_t-\bar{x})^2} = 2\rho\cdot\frac{1}{\sum(x_t-\bar{x})^2}\cdot\overbrace{\frac{\sum_{t=2}^{T}(x_t-\bar{x})(x_{t-1}-\bar{x})}{\sqrt{\sum(x_t-\bar{x})^2}\sqrt{\sum(x_t-\bar{x})^2}}}^{x_{t-1},x_t\text{ מקדם המתאם של}}$$

מה שקובע את **הכיוון** של הביטוי הזה, הוא המכפלה של ρ עם מקדם המתאם של ה-Xים.

אם נדע את ρ, נוכל לחשב נכון את השונות. אבל - **משפט גאוס מרקוב לא חל על האומד הזה!** השונות תהיה **נכונה** (לא מוטית) אך **לא מינימלית!** כדי שמשפט גאוס מרקוב יתפוס, ה-uים חייבים להיות בלתי תלויים.

כעת, נמצא אומד חסר הטיה **אחר**, במקום ה-OLS, שיהיה בעל שונות מינימלית. נלמד גם את מבחן דארבין ווטסון שירמז לנו אם עלינו להשתמש ב-OLS או באומד האחר.

- <u>הערה:</u>

עד עתה ידענו כי:

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_t^2}{T-2}$$

אך אומד זה, כאשר יש מתאם סדרתי, הוא אומד מוטה! אם במציאות קיים מתאם סדרתי אך אמדנו מודל רגיל, הוספנו טעות על טעות: גם השמטנו איברים מנוסחת השונות, וגם השונות הנאמדת של u היא מוטה.

אמידת המודל בעל מתאם סדרתי

כדי להגיע לשונות מינימלית, נתקן את **המודל** עצמו. המודל המתוקן הומצא על ידי **קורקרן ואורקוט**, שאמרו את הדבר הבא: קחו את המודל, והציבו לתוכו את ההפרעה האקראית u_t.

$$y_t = \alpha + \beta x_t + u_t$$

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

$$y_t = \alpha + \beta x_t + \rho u_{t-1} + \epsilon_t$$

$$u_{t-1} = y_{t-1} - \alpha - \beta x_{t-1}$$

$$y_t = \alpha + \beta x_t + \rho(y_{t-1} - \alpha - \beta x_{t-1}) + \epsilon_t$$

$$y_t = \alpha(1-\rho) + \beta x_t - \beta\rho x_{t-1} + \rho y_{t-1} + \epsilon_t$$

במודל הזה נעלמה ההפרעה הסדרתית, וקיבלנו מודל עקיב. עם זאת, את המודל הזה לא ניתן למדוד באמצעות *OLS* – ה-*OLS* חייב להיות ליניארי בפרמטרים, ואילו לנו יש כפל בין β ל-ρ. ניתן לסמן $\beta\rho=\gamma$, אך בכך אנו למעשה מסירים מגבלה מהמודל. γ חייבת להיות המכפלה של β ו-ρ, אך ברגע שאיחדנו אותם, הסרנו מהמחשב את המגבלה. למעשה, מתוך הרגרסיה לעולם לא נקבל שמכפלת האומדים של β ו-ρ שווה לאומד של γ. לכן, המודל הזה, כפי שכתוב – הינו **אסור לאמידה באמצעות *OLS***.

השיטה למציאת האומדים הללו היא פשוט ניסוי וטעייה, עד שמקבלים פרמטרים המביאים למינימום את ריבועי הסטיות. זו בעיה – ישנן אינסוף שלשות אפשריות, ויתכן שהסיפור בלתי פתיר. לכן בתוך המחשב מוצב גבול, הוא בוחר מספרים באופן אקראי ומשוטט מסביבים, וברגע שהוא לא מצליח לשפר את ריבועי הסטיות על ידי שינויים קלים הוא עוצר.

קורקרן ואורקוט פיתחו שיטה לפתור את המודל. הם אומרים להתחיל בבחירת גודל $\tilde{\rho}$, הצבתו במודל, והרצת *OLS*. $\tilde{\alpha}$ ו-$\tilde{\beta}$ המתקבלים הם לא הנכונים, אלא רק אלו המתקבלים בהינתן $\tilde{\rho}$. כעת לוקחים את $\tilde{\alpha}$ ואת $\tilde{\beta}$, מציבים במודל, ואומדים את ρ. את $\tilde{\rho}$ הנאמד מציבים בחזרה, ושוב אומדים את α ו-β.

$$\hat{\rho}_1 \mid \hat{\alpha}_0, \hat{\beta}_0$$

$$\hat{\alpha}_2, \hat{\beta}_2 \mid \hat{\rho}_1$$

$$\hat{\rho}_3 \mid \hat{\alpha}_2, \hat{\beta}_2$$

$$\hat{\alpha}_4, \hat{\beta}_4 \mid \hat{\rho}_3$$

בסופו של דבר, אם ההבדל בין $\hat{\alpha_0},\hat{\beta_0}$ ל-$\hat{\alpha_4},\hat{\beta_4}$ נעשה קטן מאוד, עוצרים ומניחים ש-α ו-β כבר לא תלויים ב-ρ, ולכן הם האמיתיים.

התוצאה של התהליך הזה תהיה אומדים בעלי שונות מינימלית, בהסתמך על משפט גאוס מרקוב.

ב-*E-VIEWS*, מריצים רגרסיה על *Y*, *X* ו-*AR(1)*.

דרבין ווטסון

ישנו משתנה סטטיסטי בשם דרבין ווטסון. המשתנה הסטטיסטי הזה מודד את ρ, כשמריצים רגרסיה *רגילה*.

$$DW = \frac{\sum_{t=2}^{T}(\hat{u}_t - \hat{u}_{t-1})^2}{\sum_{t=1}^{T}\hat{u}_t^2} = \overbrace{\frac{\sum_{t=2}^{T}\hat{u}_t^2}{\sum_{t=1}^{T}\hat{u}_t^2}}^{\approx 1} + \overbrace{\frac{\sum_{t=2}^{T}\hat{u}_{t-1}^2}{\sum_{t=1}^{T}\hat{u}_t^2}}^{\approx 1} - \overbrace{\frac{2\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sqrt{\sum_{t=1}^{T}\hat{u}_t^2}\sqrt{\sum_{t=1}^{T}\hat{u}_t^2}}}^{\approx \rho}$$

הסטטיסטי הזה יהיה בקירוב בין 0 ל-4. האיבר הזה הוא למעשה תוצאת הרגרסיה:

$$\hat{u}_t = \gamma\hat{u}_{t-1} + \epsilon_t$$

$$\hat{\gamma} = \frac{\sum_{t=2}^{T}\hat{u}_t\hat{u}_{t-1}}{\sum_{t=2}^{T}\hat{u}_t^2}$$

ככל שהסטטיסטי קרוב ל-2, זו עדות לכך שאין מתאם. ככל שהוא קרוב ל-4 או 0, זו עדות למתאם. את התוצאה יש לבדוק מול טבלת התפלגות דרבין ווטסון. טבלה זו היא בעייתית: ישנם תחומים מספריים בהם לא נוכל לתת הכרעה חד משמעית לכאן או לכאן ברמת ביטחון מסויימת.

- שיטה זו בודקת מתאם מסדר ראשון בלבד. לא ניתן להיעזר בשיטה זו מודלים הכוללים מתאם מסדר גבוה יותר, כמו $u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \epsilon_t$.

נשאלת השאלה, מדוע לא להשתמש תמיד בשיטת קורקרן ואורקוט. למעשה בשיטה זו אנו כופים על המדגם מתאם סדרתי, ועלולים לקבל תוצאות מעוותות. לכן תמיד כדאי להתחיל עם בדיקת דרבין ווטסון.

משתנה מסביר מקרי יחד עם מתאם סדרתי בהפרעות

$$y_t = \beta y_{t-1} + u_t$$

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

כעת, גם המשתנה המסביר מקרי, וגם קיים מתאם סדרתי בהפרעות.

$$\hat{\beta} = \frac{\sum_{t=2}^{I} y_{t-1}\cdot y_t}{\sum_{t=2}^{T} y_{t-1}^2} = \frac{\sum_{t=2}^{I} y_{t-1}\cdot(\beta y_{t-1} + u_t)}{\sum_{t=2}^{T} y_{t-1}^2} = \beta + \frac{\sum_{t=2}^{I} y_{t-1}\cdot u_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

כבר ראינו את זה קודם: קיים מתאם בין המונה למכנה. אך במקרה הקודם שבו ראינו את זה (משתנה מסביר לא מקרי), לא היה מתאם בין האיברים במונה. במצב החדש, יש מתאם גם בין המונה למכנה, וגם בין האיברים במונה. הבעיה הזו גורמת לכך שגם אם נוסיף תצפיות, לא נוכל להיעזר בתכונת העקיבות. האומד הזה מוטה בכל מקרה. גם אם יהיו אינסוף מדגמים, לא נוכל לגלות את β. זה בגלל שאנחנו לא יודעים מי השפיע באמת על השינוי של היום, ולא נוכל להפריד בין ההשפעה של y (שהוא נצפה) להשפעה של u (שהינו בלתי נצפה).

$$= \beta + \frac{\sum_{t=2}^{I} y_{t-1} \cdot (\rho u_{t-1} + \epsilon_t)}{\sum_{t=2}^{T} y_{t-1}^2} = \beta + \frac{\sum_{t=2}^{I} y_{t-1} \cdot (\rho(y_{t-1} - \beta y_{t-2}) + \epsilon_t)}{\sum_{t=2}^{T} y_{t-1}^2}$$

$$= \beta + \overbrace{\frac{\sum_{t=2}^{I} y_{t-1} \cdot \rho y_{t-1}}{\sum_{t=2}^{T} y_{t-1}^2}}^{=\rho} - \frac{\sum_{t=2}^{I} y_{t-1} \cdot \rho\beta y_{t-2}}{\sum_{t=2}^{T} y_{t-1}^2} + \overbrace{\frac{\sum_{t=2}^{I} y_{t-1} \cdot \epsilon_t}{\sum_{t=2}^{T} y_{t-1}^2}}^{\to 0 \mid T\to\infty}$$

נביט על האיבר האמצעי:

$$\frac{\sum_{t=2}^{I} y_{t-1} \cdot \rho\beta y_{t-2}}{\sum_{t=2}^{T} y_{t-1}^2}$$

הוא די דומה ל:

$$\hat{\beta} = \frac{\sum_{t=2}^{I} y_{t-1} \cdot y_t}{\sum_{t=2}^{T} y_{t-1}^2}$$

בייחוד כאשר אנו עוסקים בכמות אינסופית של תצפיות. נסכם מה קורה באינסוף:

$$plim\ \hat{\beta} = \beta + \rho - plim\ \hat{\beta} \cdot \rho \cdot \beta$$

$$plim\ \hat{\beta} = \frac{\beta + \rho}{1 + \beta\rho}$$

אם ρ היה 0, $\hat{\beta}=\beta$. אך אם $\rho\neq 0$, המשתנה מוטה ואף אינו עקיב. אנחנו מקבלים שהקשר בין ה-y (β) תלוי בקשר בין ה-u (ρ).

שונות שונה

כלומר, מצבי רוח של אנשים שונים יכולים לנוע ממרחב מדגם אחר. יתכן גם שהשונות תגדל עם הזמן, למשל, בשינוי שערי מטבע (שקל יותר או שקל פחות לא מהווים שונות גדולה, אך כשערך השקל משתנה, השונות משתנה יחד איתו).

אנחנו חוזרים להנחות הקודמות: x לא מקרי, אין מתאם בין ה-u, u מתפלג נורמלית.

הנחת ההסרה של השונות השווה לא מפריעה לחוסר ההטיה של β. עם זאת, בתצפיות בהן השונות גדולה, המחשב "מתאמץ" יותר למזער את הריבועים, וכך נותן יותר משקל לתצפיות אלה. המטרה שלנו תהיה לנטרל את השפעת השונות המוגדלת, ולכן נצטרך להריץ מבחן שיסייע לנו להבחין בין שונות מנופחת לשונות הגדולה באופן מקרי.

$$E(u_t) = 0$$

$$u_t \sim N(0, \sigma_{ut}^2)$$

נבדוק את ההשפעה על $\hat{\beta}$ (אנו משמיטים את החותך על מנת לפשט את העבודה):

$$y = \beta x_t + u_t$$

$$\hat{\beta} = \frac{\sum x_t \cdot y_t}{\sum x_t^2} = \frac{\sum x_t \cdot (\beta x_t + u_t)}{\sum x_t^2} = \beta + \frac{\sum x_t \cdot u_t}{\sum x_t^2}$$

$$E(\hat{\beta}) = \beta$$

$$var(\hat{\beta}) = E\left(\hat{\beta} - E(\hat{\beta})\right)^2 = E\left(\frac{\sum x_t \cdot u_t}{\sum x_t^2}\right)^2 = E\sum\left(\frac{x_t \cdot u_t}{\sum x_t^2}\right)^2 = \frac{\sum x_t^2 \cdot E(u_t)^2}{(\sum x_t^2)^2}$$

$$\boldsymbol{var(\hat{\beta}_{OLS}) = \frac{\sum x_t^2 \cdot \sigma_{ut}^2}{\left(\sum x_t^2\right)^2}}$$

תיקון בדיעבד: חישוב המודל, ואז תיקון רטרואקטיבי של השונות בכל תצפית.

תיקון מראש: התאמה של הסקאלה לפני האמידה.

השונות שקיבלנו רלוונטית אך ורק לאמידה באמצעות המודל הקלאסי, ללא תיקון מראש. כעת נלמד שיטה אחרת, שלה תהיה שונות קטנה יותר. זה אפשרי כיוון שמשפט גאוס מרקוב לא חל על ה-OLS כשהוא לא מקיים את ההנחות הקלאסיות.

שיטת WLS (Weighted Least Squares)

בקורס זה אנו מניחים שאנו יודעים מה גורם לשונות השונה – למשל, אינפלציה או מין. לדוגמה:

$$\sigma_{ut}^2 = \sigma_u^2 \cdot z_t^2$$

במקרה כזה, נאמוד את המודל:

$$\frac{y_t}{z_t} = \frac{\beta x_t}{z_t} + \frac{u_t}{z_t}$$

$$\frac{y_t}{z_t} = \frac{\beta x_t}{z_t} + v_t$$

$$\boldsymbol{\hat{\beta}_{WLS}} = \frac{\sum\left(\frac{x_t}{z_t}\right)\cdot\left(\frac{y_t}{z_t}\right)}{\sum\left(\frac{x_t}{z_t}\right)^2} = \frac{\sum\left(\frac{x_t}{z_t}\right)\cdot\left(\frac{\beta x_t + u_t}{z_t}\right)}{\sum\left(\frac{x_t}{z_t}\right)^2} = \boldsymbol{\beta + \frac{\sum\left(\frac{x_t}{z_t}\right)\cdot\left(\frac{u_t}{z_t}\right)}{\sum\left(\frac{x_t}{z_t}\right)^2}}$$

$$var(\hat{\beta}_{WLS}) = E\left(\hat{\beta}_{WLS} - E(\hat{\beta}_{WLS})\right)^2 = E\left(\frac{\sum\left(\frac{x_t}{z_t}\right)\cdot\left(\frac{u_t}{z_t}\right)}{\sum\left(\frac{x_t}{z_t}\right)^2}\right)^2 = E\sum\left(\frac{\left(\frac{x_t}{z_t}\right)\cdot\left(\frac{u_t}{z_t}\right)}{\sum\left(\frac{x_t}{z_t}\right)^2}\right)^2$$

$$= \frac{\sum\left(\frac{x_t}{z_t}\right)^2\cdot\left(\frac{E(u_t^2)}{z_t^2}\right)}{\left(\sum\left(\frac{x_t}{z_t}\right)^2\right)^2} = \frac{\sigma_u^2}{\sum\left(\frac{x_t}{z_t}\right)^2}$$

$$\boldsymbol{var(\hat{\beta}_{WLS}) = \frac{\sigma_u^2}{\sum\left(\frac{x_t}{z_t}\right)^2}}$$

<u>כעת נוכיח ששונות ה-WLS קטנה משונות ה-OLS:</u>

נציב בשונות של ה-OLS את $\sigma_{ut}^2 = \sigma_u^2 \cdot z_t^2$:

$$var(\hat{\beta}) = \frac{\sigma_u^2 \cdot \sum x_t^2 \cdot z_t^2}{(\sum x_t^2)^2}$$

נחלק את השונויות זו בזו כדי להוכיח ששונות ה-WLS קטנה יותר:

$$\frac{\frac{\sigma_u^2}{\sum\left(\frac{x_t}{z_t}\right)^2}}{\frac{\sigma_u^2 \cdot \sum x_t^2 z_t^2}{(\sum x_t^2)^2}} = \frac{(\sum x_t^2)^2}{\sum\left(\frac{x_t}{z_t}\right)^2 \cdot \sum x_t^2 z_t^2}$$

נסמן: $a_t = x_t z_t,\ c_t = \frac{x_t}{z_t}$

$$= \frac{(\sum a_t c_t)^2}{\sum a^2{}_t \cdot \sum c^2{}_t} \overset{?}{<} 1$$

$$(\sum a_t c_t)^2 \overset{?}{<} \sum a_t^2 \cdot \sum c_t^2$$

$$(a_1 c_1 + a_2 c_2)^2 \overset{?}{<} (a_1^2 + a_2^2)(c_1^2 + c_2^2)$$

$$a_1^2 c_1^2 + a_2^2 c_2^2 + 2a_1 a_2 c_1 c_2 \overset{?}{<} a_1^2 c_1^2 + a_2^2 c_2^2 + a_1^2 c_2^2 + a_2^2 c_1^2$$

$$0 < (a_1 c_2 - a_2 c_1)^2 \ \blacksquare$$

ולכן, עדיף לתקן **לפני** האמידה **ולא** אחריה.

בעיה נוספת באמידת OLS:

כשאם משתמשים באומד הרגיל לשונות β, הכולל את σ_u, בדרך כלל הצבנו בו את האומד הבא:

$$\hat{\sigma}_u^2 = \frac{\sum \hat{u}_t^2}{T-1}$$

$$\hat{u}_t = y_t - \hat{\beta} x_t$$

למעשה, האומד הזה "מערבב" טעויות, **נותן לכולן את אותו המשקל**, ואומד לכל התצפיות $\hat{\sigma}_u^2$ אחד. זה מביא **לאומדן שונות מוטה** ל-$\hat{\beta}$, וזוהי **טעות**. הצבה של האומדן המוטה הזה עשויה לגרום למצב בו:

$$\widehat{var}(\hat{\beta}_{OLS}) < \widehat{var}(\hat{\beta}_{WLS})$$

לעומת זאת, כשאומדים WLS, כבר תיקנו מראש את הסטיות, ולכן האומדן הנ"ל של $\hat{\sigma}_u^2$ מותר להצבה בנוסחת השונות. מדוע?

$$var(v_t) = var\left(\frac{u_t}{z_t}\right) = \frac{\sigma_{ut}^2}{z_t^2} = \frac{\sigma^2 z_t^2}{z_t^2} = \sigma_u^2$$

איתור שונות שונה

כדי למצוא שונות שונה, צריך להחליט קודם מה החשד שלנו. מסדרים את התצפיות לפי ציר הזמן, ולוקחים 20 תצפיות גבוהות ו-20 תצפיות נמוכות. אם מגלים שהשונות שונה, מבינים שהיא משתנה לאורך הזמן (עולה או יורדת, לפי החשד). אם כמות התצפיות שלנו היא 40, אי אפשר לקחת 20 מכל כיוון – כיוון שהתצפיות 20 ו-21, שהן קרובות זו לזו, גם יקרבו את השונות של שתי הקבוצות. לכן מה שבדרך כלל עושים, זורקים 20-25% מהתצפיות שבאמצע.

אם כך, ברגע שיש 40 תצפיות:

$$\frac{\sum_{t=23}^{40}\frac{\hat{u}_t^2}{17}}{\sum_{t=1}^{18}\frac{\hat{u}_t^2}{17}} \sim F$$

(הערה: לכאורה כמות דרגות החופש שבמכנה לא חשובה כיוון שהיא מצטמצמת, אך במקרים שבהם המדגם לא יהיה באותו גודל בשני החלקים – זה לא קורה)

המונה והמכנה, כל אחד מהם מקיים התפלגות F (שהיא למעשה מקבילה ל-t^2). על מנת שגם המנה תקיים התפלגות F, חייבים להניח אי תלות בין המונה למכנה. אך בפועל, חייבת להיות תלות בין המונה למכנה, כיוון שכל הטעויות נאמדו בעזרת אותו $\hat{\beta}$ (הרי האומד עצמו כולל את כל ה-u).

לכאורה, את זה פותרים באמצעות אומדן שונה לכל מדגם, ו-$\hat{\beta}$ שונה לכל אחד. המחיר לזה הוא אובדן של תצפיות, ולכן ערכים גדולים מאוד בטבלת F (ויהיה קשה לדחות את ההשערה שאין שונות שונה).

ואם יש מדגם גדול מאוד? למדנו על $\hat{\beta}$ שהוא אומד עקיב, כלומר, ככל שהמדגם יותר גדול, אנו מתקרבים יותר ל-β האמיתי – ואז מתלכתחילה כלל אין תלות בין המונה למכנה! במקרה כזה אפשר שלא לפצל.

- הערה לגבי משתנה דאמי: במקרה של קבוצות שונות, למשל גברים ונשים, יתכן ומלכתחילה לכל אחד מהם יהיה β שונה. אבל אם גם ה-β שונה וגם השונות שונה, יתכן ומלכתחילה לא היתה שום סיבה לחבר בין שתי הקבוצות, והיה צורך לעשות להם מודלים נפרדים.

המבחן הזה הוא בעייתי: אם קיבלנו בקבוצה הראשונה נמוך ובאחרונה גבוה, אנחנו מניחים עליה לאורך כל המדגם – אבל זה לא מתחייב! יתכן ולעשר התצפיות הראשונות תהיה שונות גבוהה, אחרי זה נמוכה, אחרי זה גבוהה שוב וכו'. בפועל, יכולים להיות אינסוף צירופים כאלה. המשמעות היא שאין מבחן שיכול להגיד אם השונות שווה בלי לבדוק את כל הקומבינציות של תתי קבוצות במדגם. כלומר, אם התוצאה של מבחן F היא שאין הבדל מובהק בין המונה למכנה, הדבר היחיד שנוכל לומר הוא שאנחנו פוסלים שונות שונה בין 20 הראשונים ל-20 האחרונים – אבל בשום אופן אין להסיק שהשונות שווה לכל החלקים של המדגם!

- הערה לגבי הסקת מסקנות מהמבחן: הקבוצה לה אנו משערים יש את השונות הגבוהה יותר אמורה להיות במכנה. אבל יתכן ובטעות הפכנו את הסדר, וקיבלנו סטטיסטי קטן שבטעות יגרום לנו לקבל את השערת האפס (שהשונות שווה). לכן, אם קיבלנו סטטיסטי קטן מאוד יש להפוך את המונה והמכנה ולבדוק שוב.

מבחן WHITE

לעתים אנחנו לא בטוחים במי או במה השונות תלויה, ולא יודעים את כיוון השינוי שלה. במקרה זה, אומדים את המודל הסטנדרטי:

$$y_t = \beta x_t + u_t$$

$$\hat{u}_t \rightarrow \hat{u}_t^2$$

ואז אומדים את הרגרסיה:

$$\hat{u}_t^2 = \gamma_0 + \gamma_1 x_t + \gamma_2 x_t^2 + \gamma_3 z_t + \gamma_4 z_t^2 + \gamma_5 x_t z_t + \epsilon_t$$

כלומר, בודקים את הקשר בין X ו-Z (שהוא משתנה מקרי כלשהו בו אנחנו חושדים) ל-$\hat{u}^2$ בפולינום עד לדרגה שניה. הרגרסיה הזו תרמז לנו **עם מי מתואם השינוי** בהפרעה האקראית. מבחן זה הוא חלש. המסקנה היחידה שניתן להסיק ממודל זה היא אם קיימת שונות שונה <u>המתואמת</u> עם המשתנים הנבדקים – **אי אפשר** לעשות תיקון מראש באמצעות משתנים אלה (כי יתכן ומשתנים נוספים משפיעים שלא ידענו עליהם). גם אם נותרנו רק עם γ_0 **אי אפשר** להסיק שהשונות שווה – רק שהיא לא מתואמת עם X ו-Z.

בעית האנדוגניות של המשתנה המסביר

בדרך כלל, X לא נקבע על ידי גורם חיצוני, אלא מתוך המערכת. למשל, במודל של השכר לפי ההשכלה:

$$\ln(w_i) = \alpha + \beta s_i + u_i$$

ההשכלה היא משתנה מקרי – כל אחד בוחר כמה ללמוד על פי משתנים אישיים שלו (כמו אינטיליגנציה). היא אינה משתנה שקבע החוקר. במקרה כזה, יתכן וההפרעה האקראית לא השפיעה רק על השכר, אלא גם על ההשכלה. יתכן שאדם אינטיליגנטי משכיל יותר וגם מרוויח יותר עקב עובדה זו. היינו רוצים לבדוק מה היה מרוויח אותו אדם ללא ההשכלה, איך אין לנו אפשרות להשיג מידע כזה. המודל שלנו היה מעמיס את ההשפעה ה"אמיתית", של האינטיליגנציה, הנובעת מ-u – על ה-β. זהו אומד מוטה קלאסי. יש אף כאלה הטוענים שלהשכלה אין השפעה כלל על השכר, אלא שמעסיקים מתייחסים להשכלה כאיתות על האינטליגנציה.

כבר הנחנו ש-E(u)=0. אבל לאור הנאמר לעיל, ניתן לפצל את התוחלות של u לתוחלות מותנות בהתאם למשתנה המסביר: $E(u_i \backslash S = 10), E(u_i \backslash S = 20)$ וכן הלאה. כל רגרסיה שנעשתה על השכר ללא התיחסות לתיאום בין ההפרעה להשכלה, היא מוטית באופן ודאי.

טעות נפוצה היא בבדיקה של השפעה של איכות בית הספר על הציונים: בד"כ תלמידים גרועים הולכים לבתי"ס גרועים, ותלמידים טובים לבתי"ס טובים. כך, מי שהשפיע על הציונים היה איכות התלמידים ולא איכות בית הספר. כדי לבצע בדיקה נכונה היה צורך לפזר תלמידים בין בתי הספר באופן אקראי.

דוגמה נוספת היא במחקר יעילות של תרופה: לוקחים גליון חולים, בודקים אילו תרופות לקחו ומה תוצאת הטיפול. לאחר מכן מריצים רגרסיה ובודקים את היעילות. כאן המסקנה יכולה להתעוות כיוון שמי שהיה חולה יותר, מלכתחילה קיבל יותר תרופות. המסקנה של חוקר שבדק באופן כזה היתה להפסיק לתת תרופות... על כן, במחקר רפואי נכון, מחלקים את החולים לקבוצות באופן אקראי.

בכלכלה אנחנו בבעיה, כיוון שבגדול קשה לערוך ניסויים: אנחנו תמיד צריכים לחקור בדיעבד. מחקר לדוגמה שיכול לעלות, הוא הקשר בין מספר הילדים לגיל הפרישה. יתכן שהילדים אינם מסבירים בכלל את גיל הפרישה, אלא שגורם שלישי השפיע גם על מספר הילדים וגם על גיל הפרישה. במקרים רבים אין מה לעשות כדי לפתור את הבעיה, אך במקרה זה דווקא כן נמצא פתרון. אנו רוצים איכשהו לנתק את הקשר בין המשתנה המסביר להפרעה האקראית. החוקרים חיפשו מאפיין כלשהו שקשור למספר הילדים, אך אינו קשור להפרעה האקראית, שיוכל לשמש כמתווך. הם מצאו שהמין של הילדים משפיע על כמותם – וניתן להשתמש בו כמתווך שאינו תלוי בהפרעה האקראית.

לגבי המודל המסביר את השכר לפי ההשכלה – התברר שמי שנשאר שנה בגן, למד פחות בהמשך החיים. כלומר, מועד הלידה במהלך השנה הוא השפעה על ההשכלה ומכאן על השכר, שאינה תלוי באמביציה או אינטיליגנציה.

$$\ln(w_i) = \alpha + \beta s_i + u_i$$

$$E(\hat{\beta}) = \beta + \sum E\left(\frac{(s_i - \bar{s})u_i}{\sum(s_i - \bar{s})^2}\right)$$

היינו רוצים שהתוחלת של המכפלה תהיה תוחלת המכפלות, כי אז האומד יהיה חסר הטיה (תוחלת u היא אפס). כלומר, מצב בו $E(u_i \backslash s_i) = E(u_i)$.

- הערה: יש לזכור שעקב המשוואות הנורמליות, s לא מתואם עם û. אבל אם המודל לא מתוכנן נכון, בגלל המשוואה הנורמלית הזו, s "שאב" את כל ההשפעה של u, וכך גם ŝ וגם û מוטים.

אומד משתני עזר IV

$$y_i = \alpha + \beta x_i + u_i$$

$$cov(x_i, u_i) \neq 0$$

$$\hat{\beta} = \beta + \sum\left(\frac{(x_i - \bar{x})u_i}{\sum(x_i - \bar{x})^2}\right)$$

במקרה כזה תצפיות נוספות לא יועילו, בניגוד לשיטה שראינו בעבר: הן פשוט יגדילו את המכנה. אבל המתאם במונה עדיין יהיה קיים.

משתנה עזר – Z – צריך לקיים את שתי התכונות הבאות:

1. מתואם עם x: $cov(x_i, z_i) \neq 0$
2. לא מתואם עם u:
   a. $cov(u_i, z_i) = 0$ (אין קשר ליניארי)
   b. $E(u_i \backslash z_i) = 0$ (בדיקה חזקה יותר – אין קשר מכל סוג)
   c. $E(u_i \backslash z) = 0$ (בדיקה הכי חזקה – אין קשר בין ההפרעה האקראית ל-Z של אף תצפית, כדי שלא יהיה מתאם עם המכנה, ונוכל להשתמש ב-OLS)

(הערה: Z ו-X צריכים להיות מתואמים אך לא במתאם מלא, אחרת גם Z יהיה מתואם עם u.)

$$\hat{\beta}_{IV} = \frac{\sum(z_i - \bar{z})y_i}{\sum(z_i - \bar{z})(x_i - \bar{x})} = \frac{\sum(z_i - \bar{z})(\alpha + \beta x_i + u_i)}{\sum(z_i - \bar{z})(x_i - \bar{x})} = \beta + \frac{\sum(z_i - \bar{z})u_i}{\sum(z_i - \bar{z})(x_i - \bar{x})}$$

$$\sum E\left(\frac{(z_i-\bar{z})u_i}{\sum(z_i-\bar{z})(x_i-\bar{x})}\right)$$

התוחלת של הביטוי הזה אינה אפס, כיוון שיש X במכנה. אך התלות במקרה שלנו חלשה – כי ישנו מתאם רק בין X ו-u של אותה התצפית (בניגוד למצב במתאם סדרתי). יש לזכור כי כל X מופיע גם ב-$\bar{X}$, וזוהי עוד נקודה של תיאום עם u, אם כי גם הוא חלש וגם הולך ונחלש ככל שנוספות תצפיות.

<u>מסקנה: אומד ה-IV הינו מוטה אך עקיב.</u>

בפרקטיקה, בחירת משתנה העזר משנה דרמטית את התוצאות, בגלל מתאם שונה בין X ל-Z.

משתנה עזר דיכוטומי

מקבל ערכי 0 או 1. אם יש משתנה דיכוטומי המתואם עם X, ניתן להשתמש גם בו כמשתנה עזר. במקרה כזה,

$$\hat{\beta}_{IV}=\frac{\sum(z_i-\bar{z})y_i}{\sum(z_i-\bar{z})(x_i-\bar{x})}=\frac{\bar{y}_1-\bar{y}_0}{\bar{x}_1-\bar{x}_0}$$

כאשר $\bar{y}_1$ הוא ממוצא התצפיות בהן z=1, וכן הלאה. הוכחה:

$$\hat{\beta}_{IV}=\frac{\sum(z_i-\bar{z})y_i}{\sum(z_i-\bar{z})x_i}=\frac{\sum z_iy_i-\bar{z}\bar{y}I}{\sum z_ix_i-\bar{x}\bar{z}I}$$

נשתמש במשוואות:

$$\bar{y}=\bar{z}\bar{y}_1+(1-\bar{z})\bar{y}_0,\qquad \bar{x}=\bar{z}\bar{x}_1+(1-\bar{z})\bar{x}_0$$

נציב חזרה:

$$=\frac{\frac{\sum z_iy_i}{\bar{z}I}-\frac{\bar{z}\bar{y}I}{\bar{z}I}}{\frac{\sum z_ix_i}{\bar{z}I}-\frac{\bar{z}\bar{x}I}{\bar{z}I}}=\frac{\bar{y}_1-\bar{y}}{\bar{x}_1-\bar{x}}=\frac{\bar{y}_1-\bar{z}\bar{y}_1-(1-\bar{z})\bar{y}_0}{\bar{x}_1-\bar{z}\bar{x}_1-(1-\bar{z})\bar{x}_0}=\frac{(1-\bar{z})\bar{y}_1-(1-\bar{z})\bar{y}_0}{(1-\bar{z})\bar{x}_1-(1-\bar{z})\bar{x}_0}=\frac{\bar{y}_1-\bar{y}_0}{\bar{x}_1-\bar{x}_0}$$

טעויות במודלים כלכליים הנובעות מאנדוגניות המשתנה המסביר

לפעמים משתנה העזר "קופץ" מתוך המודל. למשל, לפי קיינס:

$$c_t=\alpha+\beta y_t+u_t$$

קיינס הניח שעל כל שקל עליה בתוצר y_t, הצריכה עולה ב-β יחידות. α מסמל צריכה מינימלית שאינה תלויה בהכנסה. לפי קיינס, המשק לעתים לא מגיע למצב של תעסוקה מלאה. הוא טוען כי תמיד תהיה אבטלה מסוימת, כי לא תמיד יהיה מספיק ביקוש לתוצר. המדינה רוצה לדעת את ה-β כדי לדעת כמה כסף להזרים למשק להגדלת התוצר. אבל, תוצר במשק מוגדר כך:

$$y_t=c_t+g_t+i_t$$

ולכן, כיוון ש-u_t כלול ב-c_t, הוא מתואם גם עם y_t, והאומד ל-β הוא מוטה.

עוד דוגמה אפשרית היא בביקוש והיצע, למשל, הקשר בין מחיר התפוזים לכמות הנמכרת. המודל (המוטעה) במקרה זה יהיה:

$$x_t = \alpha + \beta P_{x_t} + u_t$$

המודל הזה חסר משמעות, כיוון שהוא מייצג שיווי משקל המתקבל מיישום השפעות סותרות אצל יצרנים וצרכנים.

הוכחה לגבי מוטות ה-OLS:

$$(I)\ X_t^S = \alpha P_{x_t} + u_{1t}$$

$$(II)\ X_t^D = \beta P_{x_t} + u_{2t}$$

$$(III)\ X_t^D = X_t^S$$

$$\hat{\alpha} = \frac{\sum P_{x_t} \cdot x_t}{\sum P_{x_t}^2} = \alpha + \frac{\sum P_{x_t} \cdot u_{1t}}{\sum P_{x_t}^2}$$

$$\alpha P_{x_t} + u_{1t} = \beta P_{x_t} + u_{2t}$$

$$P_{x_t} = \frac{(u_{2t} - u_{1t})}{\alpha - \beta}$$

קיבלנו ש-P תלוי ישירות ב-u, ולכן האומד ל-α מוטה.

יש במודל הזה תקלה בסיסית מבחינה כלכלית: גם X וגם P הם תוצר של שיווי משקל! X ו-P זזים יחד בגלל תזוזה של העקומות (גורמים אחרים שאמורים להיתפס ב-u).

כדי שנוכל להסיק מסקנות בנוגע להיצע וביקוש, נצטרך "לקבע" את אחת העקומות (היצע או ביקוש) כדי שנוכל לאמוד אותה. כמו כן, נצטרך להוסיף את ההשפעה שמלכתחילה הזיזה את העקומה כמשתנה עזר (כיוון שתזוזת העקומה משפיעה גם על הביקוש וגם על המחיר). לדוגמה: נשתמש ב-I, הכנסה, כמשתנה עזר, (נניח I לא מתואם עם u), נקבע את משוואת ההיצע וננסה לאמוד אותה:

$$X_t^D = \beta_1 P_{x_t} + \beta_2 I_t + u_{2t}$$

$$\hat{\alpha}_{IV} = \frac{\sum I_t \cdot X_t}{\sum I_t P_{x_t}} = \alpha + \frac{\sum I_t \cdot u_{1t}}{\sum I_t P_{x_t}}$$

$$P_{x_t} = \frac{\beta_2}{\alpha - \beta_1} I_t + \frac{u_{2t} - u_{1t}}{\alpha - \beta_1}$$

כך, יש מתאם בין I ל-P רק בתצפית אחת, מתאם שהולך ודועך ככל שמוסיפים תצפיות – ולכן האומד עקיב.

שיטת 2SLS

זוהי למעשה שיטת אמידה בריבועים פחותים באמצעות שתי משוואות. בדוגמה הקודמת, לא ניתן לאמוד את הביקוש, כיוון שאין לנו מספיק מידע. באמידת ההיצע, יש לנו בעיה – כיוון ש-P מתואם עם u_1, אנו מקבלים אומד מוטה. כדי לעקוף את הבעיה, ננסה לאמוד את P:

$$P_{x_t} = \gamma I_t + v_t$$

$$\hat{P}_{x_t} = \hat{\gamma} \cdot I_t$$

$$\hat{\gamma} = \frac{\sum I_t P_{x_t}}{\sum {I_t}^2} = \gamma + \frac{\sum I_t v_t}{\sum {I_t}^2}$$

ההפרעה האקראית כאן, v_t, מייצגת את:

$$v_t = \frac{u_{2t} - u_{1t}}{\alpha - \beta_1}$$

I לא מתואם עם אף אחד מהאיברים פה, ולכן קיבלנו אומד OLS בלתי מוטה ל-P.

בשלב שני, מציבים את $\hat{P}$ במקום את ה-P האמיתי:

$$\hat{\alpha} = \frac{\sum \hat{P}_{x_t} \cdot x_t}{\sum \hat{P}_{x_t}^2}$$

כעת אין בעיה של תיאום בין $\hat{P}$ ל-u.

בפועל, האומד הזה זהה לחלוטין לאומד ה-IV שקיבלנו קודם, וגם הוא עקיב אך מוטה. נוכיח:

$$\hat{\alpha} = \frac{\sum \hat{P}_{x_t} \cdot x_t}{\sum \hat{P}_{x_t}^2} = \frac{\hat{\gamma} \sum I_t x_t}{\hat{\gamma}^2 \sum {I_t}^2} = \frac{\sum I_t x_t}{\hat{\gamma} \sum {I_t}^2} = \frac{\sum I_t x_t}{\frac{\sum I_t P_{x_t}}{\sum {I_t}^2} \sum {I_t}^2} = \frac{\sum I_t x_t}{\sum I_t P_{x_t}}$$

אנו יודעים שבאופן כללי, אומד ה-IV עקיב וחסר הטיה – ולכן גם אומד ה-2SLS שלנו. מדוע קיבלנו אומד מוטה? אם נחזור אחורה ל-V, נראה שהוא מחושב דרך u – ולכן, גם $\hat{\gamma}$ מתואם איתו, ובעקבותיו – $\hat{P}$ גם הוא מתואם עם u. ככל שמוסיפים יותר תצפיות, התלות ב-u הופכת לזניחה, ואנו מקבלים אומד עקיב.

אם שתי השיטות מניבות את אותו האומד, מדוע יש לנו צורך בשיטה השניה?

נניח שאנו מכניסים למודל הביקוש גם משתנה z כלשהו שלא תלוי במערכת עצמה – מחיר מוצר תחליפי מיובא. כעת,

$$P_{x_t} = \frac{\beta_2}{\alpha - \beta_1} I_t + \frac{\beta_3}{\alpha - \beta_1} z_t + \frac{u_{2t} - u_{1t}}{\alpha - \beta_1}$$

עכשיו אפשר לאמוד את P באמצעות I ו-z:

$$\hat{P}_{x_t} = \hat{\gamma}_1 I_t + \hat{\gamma}_2 z_t + v_t$$

הרגרסיה הזו מוצאת את הקורלציה הכי טובה בין I ו-z ל-P. למעשה, ייצרנו כעת את P̂ שיהיה משתנה העזר שלנו. במקרה הקודם, ייצרנו משתנה עזר P̂ באמצעות משתנה אחד – I, ולכן מן הסתם קיבלנו אותה תוצאה כמו ב-IV. אבל כעת יש יותר ממשתנה אחד שיכול לשמש כמסביר, ובמקרה זה הדבר הכי טוב שנוכל לעשות הוא לייצר באמצעות כולם משתנה עזר P̂ בודד ולהשתמש בו ברגרסיה.

שיטת הריבועים הפחותים העקיפה (ILS)

אם מספר המשתנים הבעייתיים ומשתני העזר שווה, האומד הזה יהיה שווה ל- IV. עם זאת, אם יש יותר משתני עזר ממשתנה בעייתי, השיטה הזו קורסת. כלומר, שיטה זו או שווה לשתיים האחרות, או שהיא פשוט לא עובדת. בעצם, אין לשיטה זו שום יתרון.

משוואות מזוהות

$$(I)\ y_{1t} = \alpha_1 + \alpha_2 y_{2t} + \alpha_3 x_{1t} + u_{1t}$$

$$(II)\ y_{2t} = \beta_1 + \beta_2 y_{1t} + \beta_3 x_{2t} + u_{2t}$$

משתנה מוחלף: משתנים במשוואה המתואמים עם ההפרעה האקראית. **משתנה מחליף:** המחליף מופיע במערכת המשוואות, אינו תלוי בהפרעה האקראית, ולא מופיע במשוואה של המוחלף.

באמצעות המשוואות האלה ניתן לאמוד גם את y_1 וגם את y_2, כי לשניהם יש מחליפים. מצב שבו מספר המחליפים במשוואה, שווה למספר המוחלפים במשוואה – נקרא **זיהוי מדויק**.

זיהוי בחסר הוא מצב בו מספר המוחלפים במשוואה גדול ממספר המחליפים ⇐ לא ניתן לאמידה עקיבה.

זיהוי ביתר הוא מצב בו מספר המחליפים במשוואה גדול ממספר המוחלפים ⇐ אמידת 2SLS.

$$(I)\ y_{1t} = \alpha_1 + \alpha_2 y_{2t} + \alpha_3 x_{1t} + u_{1t}$$

$$(II)\ y_{2t} = \beta_1 + \beta_2 y_{1t-1} + \beta_3 x_{2t} + u_{2t}$$

במקרה כזה, אם אין מתאם סדרתי במשוואה הראשונה, ניתן לאמוד אותה ב-OLS. את המשוואה השניה עדיין לא ניתן לאמוד. משוואות כאלה אינן משוואות סימולטניות - זוהי מערכת רקורסיבית.

מסיבה זו, לעתים במקרה של מערכת המשוואות הראשונה, יציבו במשוואות את y המקביל של התקופה הקודמת, במקום את זה של התקופה הנוכחית.

$$(I)\ y_{1t} = \alpha_1 + \alpha_2 y_{2t} + \alpha_3 y_{3t} + \alpha_4 x_{1t} + u_{1t}$$

$$(II)\ y_{2t} = \beta_1 + \beta_2 y_{1t} + \beta_3 x_{2t} + u_{2t}$$

$$(III)\ y_{3t} = \gamma_1 + \gamma_2 y_{2t} + \gamma_3 x_{3t} + u_{3t}$$

במקרה כזה ישנה התאמה סימולטנית שאינה ישירה, אלא נובעת מהמערכת. במשוואה הראשונה, y_2 בעייתי ויש להחליף אותו. אבל גם y_3 בעייתי, כיוון שהוא עצמו מכיל את y_2. את y_2 מחליפים ב-x_2, ואת y_3 ב-x_3, והמשוואה הזו מזוהה במדויק.

במשוואה השניה, מחליפים את y_1 ב-x_1. y_1 כולל בתוכו גם את y_3, ואותו מחליפים ב-x_3. כך, בסיכומו של דבר, יש לנו זיהוי ביתר – y_1 מוחלף ב-x_1 ו-x_3.

באותו אופן עושים את ההחלפה גם במשוואה השלישית, גם בה יהיה זיהוי ביתר.

עד עתה הנחנו בסמוי כי אין מתאם בין ההפרעות האקראיות:

$$cov(u_{1t}, u_{2t}) = cov(u_{1t}, u_{3t}) = cov(u_{2t}, u_{3t}) = 0$$

במקרה של מערכת המשוואות הראשונה שהוצגה קודם, אין השפעה לכך שיש מתאם בין ה-uים, כיוון שגם ככה יש בעיה הנובעת מהסימולטניות של המשוואות.

$$(I)\ y_{1t} = \alpha_1 + \alpha_2 y_{2t} + \alpha_3 x_{1t} + u_{1t}$$

$$(II)\ y_{2t} = \beta_1 + \beta_3 x_{2t} + u_{2t}$$

כעת נניח כי $cov(u_{1t}, u_{2t}) \neq 0$. במקרה זה במשוואה II אין בעיה. לכאורה, גם במשוואה I לא אמורה להיות בעיה – אך בתוך y_2 נמצא u_2, והוא מתואם עם u_1. כך עלולים לייחס בטעות ל"פירמה" השפעה שמקורה בכלל בזעזוע אקראי משותף.

הערה:

אם ישנן משוואות סימולטניות, אשר באחת מהן אין חותך – ניתן לקחת את החותך כמשתנה עזר (הרי הוא אינו מתואם עם ההפרעה האקראית).