מבוא לאקונומטריקה יישומית

פרק 13 - תיאוריה מולטיקוליניאריות

תוכן העניינים

מולטיקוליניאריות:

רקע:

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

נבחין בין מולטיקוליניאריות מלאה לחלקית.

מולטיקוליניאריות מלאה:

מתאם מלא בין המשתנים המסבירים במודל.

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

מכאן ש: $r_{12} = 1$.

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

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

מולטיקוליניאריות חלקית:

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

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

זיהוי מולטיקולינאריות חלקית:

1. כאשר קיימת סתירה בין התוצאה במבחן F למובהקות המודל (המודל מובהק) לבין מבחני t למובהקות השיפועים (אף אחד מן השיפועים איננו מובהק).
הסתירה נוצרת כתוצאה מהגדלת השונות של כל אחד מהשיפועים בשל המתאם הגבוה בין הב"ת, באופן שלא מאפשר לדחות את השערת האפס למובהקות השיפועים: $S^2_{\hat{\beta}_1}=\dfrac{MSE}{SSX_1\left(1-r_{12}\right)}$, $t=\dfrac{\hat{\beta}_1}{S_{\hat{\beta}_1}}$.

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

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

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

פתרונות למולטיקולינאריות חלקית:

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

2. ניתן לעיתים לאחד את שני המשתנים למשתנה אחד.

שלבי בדירת ההשערות:

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

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

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