

## Weiterführende Fragen der Ökonometrie

### Übungsaufgaben – Blatt 2

#### Aufgabe 1

In der medizinischen Forschung wurde die Grundgesamtheit der Männer mittleren Alters (45 bis 65 Jahre alt) als Risikogruppe für einen Herzinfarkt identifiziert. Aus Erfahrungsberichten von Patienten wird vermutet, dass die regelmäßige Einnahme einer geringen Menge Aspirin das Risiko eines Herzinfarktes senken kann. Sie sind Teil einer Forschergruppe, die diese Hypothese untersuchen möchte, und beraten diese. Ihre Kollegen aus der Medizin spezifizieren die Behandlung  $C$  und das Ergebnis  $Y$  präziser:

$$C_i = \begin{cases} 1 & \text{Individuum } i \text{ beginnt ab Zeitpunkt } T \text{ mit der täglichen Einnahme einer Aspirin} \\ & \text{Tablette mit 500 mg Wirkstoff,} \\ 0 & \text{Individuum } i \text{ beginnt ab Zeitpunkt } T \text{ mit der täglichen Einnahme einer Placebo-} \\ & \text{Tablette ohne Wirkstoff} \end{cases}$$

und

$$Y_i = \begin{cases} 1 & \text{Individuum } i \text{ wurde im Zeitraum } T \text{ bis } T+365 \text{ Tage mindestens einmal wegen eines} \\ & \text{Herzinfarktes in ein Krankenhaus eingeliefert oder ist in diesem Zeitraum an} \\ & \text{einem Herzinfarkt gestorben,} \\ 0 & \text{Individuum } i \text{ wurde im Zeitraum } T \text{ bis } T+365 \text{ Tage kein einziges Mal wegen eines} \\ & \text{Herzinfarktes in ein Krankenhaus eingeliefert und ist in diesem Zeitraum nicht} \\ & \text{an einem Herzinfarkt gestorben.} \end{cases}$$

- (i) (1 Punkte) Erläutern Sie die potentiellen Ergebnisse für ein Individuum  $i$ . Wie ist der individuelle Effekt der Behandlung definiert?
- (ii) (1 Punkt) Ihnen steht eine Stichprobe der Größe  $N = 22071$  aus der Grundgesamtheit aller (in Deutschland lebenden) Männer mittleren Alters (ohne Vorerkrankungen) zur Verfügung. 11037 Individuen werden per Zufallsgenerator der Behandlungsgruppe zugewiesen ( $C = 1$ ), die anderen dienen als Kontrollgruppe ( $C = 0$ ). Das Ergebnis  $Y = 1$  wird in der Behandlungsgruppe 104 mal beobachtet, in der Kontrollgruppe 189 mal. Berechnen Sie mit diesen Daten eine Schätzung für den mittleren kausalen Effekt der Behandlung.
- (iii) (1 Punkt) Ein Kollege schlägt vor, dass bei der Berechnung des mittleren kausalen Effekts andere Beobachtungsmerkmale wie etwa das Alter der Individuen berücksichtigt werden sollten. Was halten Sie davon?

- (iv) (1 Punkt) Nach Veröffentlichung Ihrer Studienergebnisse lesen Sie in der Zeitung, dass es gesund sei, jeden Tag eine Aspirin zu schlucken. Welche Einwände lassen sich vorbringen gegen diese Interpretation der Forschungsergebnisse?

## Aufgabe 2

Es wird ein kostenloses und freiwilliges Mathematikrepetitorium für Erstsemester an einer Universität angeboten. Diese Veranstaltung hat den Zweck die Mathematikkenntnisse der Studenten aufzufrischen bzw. nicht vorhandene Kenntnisse nachzuholen. Sie steht allen immatrikulierten Studenten offen. Sie wollen den Erfolg dieser Zusatzveranstaltung evaluieren, z.B. im Hinblick auf die Abbrecherquoten in einzelnen Studiengängen, der Abschlussnote und der Studiendauer. Sei im Folgenden  $y_i$  eine Variable, die die Abschlussnote des Studenten  $i$  in einem mathematiklastigen Studiengang angibt, und  $x_i$  eine Dummyvariable, die den Wert 1 annimmt, wenn Student  $i$  an der Veranstaltung teilgenommen hat.

- (i) (1 Punkt) Erläutern Sie die potentiellen Ergebnisse für ein Individuum  $i$ . Ist Annahme **E1** erfüllt bzw. gelten Gleichungen (2.10a) und (2.10b) aus dem Skript? Begründen Sie Ihre Aussage ausführlich.
- (ii) (1 Punkt) Welche Auswirkungen auf die Gültigkeit von Annahme **E1** bzw. den Gleichungen (2.10a) und (2.10b) aus dem Skript haben folgende Szenarien:
- A Das Repetitorium wird vorwiegend von Studenten besucht, die in der Schule schwach in Mathematik gewesen sind.
- B Das Repetitorium wird vorwiegend von Studenten besucht, die sehr engagiert sind (sog. 'Streber').
- (iii) (2 Punkte) Liefert eine Regression von  $y_i$  auf (eine Konstante und)  $x_i$  eine unverzerrte Schätzung des durchschnittlichen kausalen Effekts in Szenario A? In welche Richtung ist die Schätzung verzerrt, wenn das Repetitorium kausal im Durchschnitt zu einer Verbesserung der Durchschnittsnote führt? Was erwarten Sie in Szenario B?
- (iv) (1 Punkt) Sie nehmen an, dass Szenario A vorliegt. Durch die Aufnahme weiterer Variablen in die Regressionsgleichung könnte Annahme **E1** durch Annahme **E3** ersetzt werden. Überlegen Sie, welche Variablen in  $\mathbf{x}_i$  enthalten sein sollten.

### Aufgabe 3

Betrachten Sie die Differenz-von-Differenzen Regression

$$y_i = \beta_1 + \beta_2 DT_i + \beta_3 D2_i + \beta_4 DT_i \cdot D2_i + \eta_i, \quad (1)$$

wobei  $DT_i$  den Wert 1 annimmt, wenn Individuum  $i$  zur Teilnehmergruppe gehört, sonst 0, und  $D2_i$  den Wert 1 annimmt, wenn Individuum  $i$  in der zweiten Periode beobachtet wurde, sonst 0. Um die Handhabung zu erleichtern definieren wir folgende Größen:

Für die Anzahl der Beobachtungen in einer Untergruppe:

	$DT_i = 0$	$DT_i = 1$
$D2_i = 0$	$N_1$	$N_2$
$D2_i = 1$	$N_3$	$N_4$

Beispiel: In unserer Stichprobe befinden sich  $N_1$  Beobachtungen für die Kontrollgruppe ( $DT_i = 0$ ) vor der Behandlung ( $D2_i = 0$ ).

Analog dazu definieren wir die Summen von  $y_i$  für diese Untergruppen:

	$DT_i = 0$	$DT_i = 1$
$D2_i = 0$	$S_1$	$S_2$
$D2_i = 1$	$S_3$	$S_4$

Beispiel: Die Summe von  $y_i$  für alle Beobachtungen der Stichprobe für die gilt  $DT_i = 0$  und  $D2_i = 0$  ist  $S_1$ .

- (i) (2 Punkte) Zeigen Sie, dass  $\mathbf{X}'\mathbf{X}$  folgende Form hat:

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} A & B & C & D \\ B & B & D & D \\ C & D & C & D \\ D & D & D & D \end{pmatrix} \quad (2)$$

Geben Sie auch A, B, C, und D an. (Hinweis: A, B, C und D lassen sich als Linearkombinationen von  $N_1$  bis  $N_4$  darstellen.)

- (ii) (3 Punkte) Ermitteln sie die Inverse von  $\mathbf{X}'\mathbf{X}$ , d.h.  $(\mathbf{X}'\mathbf{X})^{-1}$ . Dabei sollten Sie die Hilfe eines Computeralgebraprogramms in Anspruch nehmen, etwa die kostenfreie 'Suchmaschine' **WolframAlpha**. Um die Inverse der Matrix in Gleichung (2) zu erhalten, geben Sie

`inv {{A,B,C,D}, {B,B,D,D}, {C,D,C,D}, {D,D,D,D}}`

in das Suchfeld ein. Ein noch schöneres Ergebnis erhalten Sie, wenn sie A, B, C und D durch die Werte in  $\mathbf{X}'\mathbf{X}$  ersetzen.

- (iii) (1 Punkt) Geben Sie  $\mathbf{X}'\mathbf{y}$  an. (Hinweis: Die Elemente dieses Vektors sind Linearkombinationen von  $S_1$  bis  $S_4$ .)

- (iv) (2 Punkte) Berechnen Sie  $\hat{\beta}_1$ ,  $\hat{\beta}_2$  und  $\hat{\beta}_3$ .

- (v) (2 Punkte) Berechnen Sie  $\hat{\beta}_4$  und zeigen Sie, dass dieser Schätzer als Differenz zweier Differenzen dargestellt werden kann.

#### Aufgabe 4

In Example 13.4 in Wooldridge (2009) wird anhand des Datensatzes `injury.txt` untersucht wie lange ein verletzter Arbeiter eine Lohnausgleichszahlung in Kentucky erhält. Durch eine Reform in 1980 herrscht dabei ein Ungleichgewicht, da sie dazu führte, dass es für Gutverdiener weniger kostspielig ist länger eine Lohnausgleichszahlung zu beziehen als für Geringverdiener (bei denen sich durch die Reform nichts veränderte). Daher formen in diesem Beispiel die Geringverdiener die Kontrollgruppe. Das Treatment ist die Reformeinführung. Beachten Sie, dass der Datensatz `injury.txt` mehr Beobachtungen enthält als in Wooldridge (2009) angegeben. Ihre Ergebnisse unterscheiden sich daher von denen aus Example 13.4.

- (i) (2 Punkte) Ermitteln Sie (in der Notation der vorherigen Aufgabe) die Gruppenmittelwerte  $\frac{1}{N_1}S_1$  bis  $\frac{1}{N_4}S_4$ . Führen Sie außerdem die Differenz-von-Differenzen Regression zu Gleichung 13.12 in Wooldridge (2009)

$$\widehat{\log(durat)} = 1.126 + 0.077 afchnge + 0.256 highearn + 0.191 afchnge * highearn$$

durch. Zeigen Sie, dass der Vergleich von Mittelwerten der Untergruppen zu denselben Ergebnissen führt wie die Regression.

- (ii) (2 Punkte) Interpretieren Sie die Parameterschätzungen. Beachten Sie dabei, dass der Regressand in logarithmierter Form vorliegt und alle Regressoren diskreter Natur sind.
- (iii) (2 Punkte) Veranschaulichen Sie (händisch) die Differenz-von-Differenzen Schätzung anhand einer Graphik analog zu Figure 5.2.1 in Angrist & Pischke (2009, S.231).