Lennart Oelschläger (lennart.oelschlaeger@uni-bielefeld.de)
310214 Praktische Übung zu "Einführung in die Mikroökonometrie"
Wintersemester 2024/2025

Aufgabenblatt 4

Lösungen

Überblick

Wir wissen bereits, dass die KQ-Schätzung verzerrt ist, sofern wichtige Einflussgrößen in der Modellgleichung nicht bedacht werden. Für Panel Daten haben wir zuletzt mit dem FD- und dem FE-Schätzer zwei Verfahren für unbeobachtete, zeitkonstante Effekte kennengelernt. Beide Methoden helfen uns aber nicht, wenn nur Querschnittsdaten vorliegen oder wir uns für den Effekt einer zeitkonstanten Variable interessieren. Die *Instrumental Variables Schätzung* ist ein anderes Verfahren, um diesem Problem zu begegnen.

Und dann lernen wir mit *binären Wahlmodellen* eine neue Modellklasse kennen, bei der die abhängige Variable nicht mehr stetig, sondern binär ist.

Aufgaben

1. Welches Problem entsteht, falls ein Regressor mit dem Fehler korreliert?

In der Regressionsgleichung $y = \beta_0 + \beta_1 x + u$, falls $\text{Cov}(x, u) \neq 0$ gilt, ist MLR.4 (also die Annahme $\text{E}(u \mid x) = 0$) verletzt und der KQ-Schätzer ist verzerrt ($\text{E}(\hat{\beta}_1) \neq \beta_1$).

2. Was ist eine Instrumentenvariable? Fällt Ihnen ein Beispiel ein?

Statt x wollen wir eine andere Variable z verwenden (eine sogenannte Instrumentenvariable), die das gleiche Merkmal wie x misst, allerdings mit u unkorreliert ist. Zwei Voraussetzungen stellen wir an z:

1. Exogenität: $\text{Cov}(z, u) = 0$ (das kann im Allgemeinen nicht getestet werden, man verlässt sich hier zum Beispiel auf ökonomisches Fachwissen)
2. Relevanz: $\text{Cov}(z, x) \neq 0$ (das kann sehr einfach getestet werden, siehe unten)

Ein Beispiel: Wollen wir das Gehalt durch die Bildung erklären, können wir uns überlegen, dass Talent wohl auch einen Einfluss hat, aber nicht beobachtet wird und deshalb im Fehler verborgen ist. Falls nun Bildung und Talent korrelieren (wahrscheinlich ist das so), so wäre die Schätzung verzerrt. Stattdessen können wir uns eine Instrumentenvariable überlegen, die

1. exogen ist (unkorreliert mit Talent) und
2. relevant ist (korreliert mit Bildung).

Was könnte eine solche Variable sein?

- "Kreditkartennummer" ist exogen, aber nicht relevant

- "Intelligenzquotient" ist relevant, aber nicht exogen
- "Bildung der Eltern" könnte beides erfüllen

3. Betrachten Sie das lineare Modell

$$y = \beta_0 + \beta_1 x + u. \tag{1}$$

Angenommen, z ist eine geeignete Instrumentenvariable. Leiten Sie bitte β_1^{IV} her.

Es gilt

$$\text{Cov}(z, y) = \text{Cov}(z, \beta_0 + \beta_1 x + u) = \beta_1 \text{Cov}(z, x) + \text{Cov}(z, u).$$

Wenn nun $\text{Cov}(z, x) \neq 0$ (entspricht der obigen Annahme 2) und $\text{Cov}(z, u) = 0$ (entspricht Annahme 1), dann ist $\beta_1 = \text{Cov}(z, y)/\text{Cov}(z, x)$. Der Schätzer β_1^{IV} ergibt sich, in dem man die Populationskovarianzen und die Stichprobenkovarianzen ersetzt:

$$\beta_1^{\text{IV}} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})(y_i - \bar{y})}{\frac{1}{n-1}\sum_{i=1}^{n}(z_i - \bar{z})(x_i - \bar{x})} = \frac{\sum_{i=1}^{n}(z_i - \bar{z})y_i}{\sum_{i=1}^{n}(z_i - \bar{z})x_i}$$

4. Simulieren Sie bitte Daten wie nachfolgend angegeben und schätzen Sie dann Modell (1), einmal per KQ-Methode und einmal mit der IV-Methode.

```
set.seed(1)
Sigma <- matrix(c(1, 0.7, 0.5, 0.7, 1, 0, 0.5, 0, 1), nrow = 3)
data <- MASS::mvrnorm(n = 100, mu = c(0, 0, 0), Sigma = Sigma)
data <- as.data.frame(data)
colnames(data) <- c("x", "z", "u")
beta_0 <- 1
beta_1 <- 2
data$y <- beta_0 + data$x * beta_1 + data$u
```

```
R> kq <- lm(y ~ x, data = data)
R> coef(kq)
```

```
# (Intercept)           x
#   0.9491985   2.4688889
```

```
R> (beta_1_iv <- cov(data$z, data$y) / cov(data$z, data$x))
```

```
# [1] 1.936954
```

```
R> (beta_0_iv <- mean(data$y) - beta_1_iv * mean(data$x))
```

```
# [1] 0.8975107
```

5. Schätzen Sie Modell (1) auch mit 2SLS (Two Stage Least Squares) und reproduzieren Sie Ihr Ergebnis mit `AER::ivreg`.

```
R> ### 2SLS
R> mod_first_stage <- lm(x ~ z, data = data)
R> summary(mod_first_stage) # z ist ein relevantes Instrument
```

```
#
# Call:
# lm(formula = x ~ z, data = data)
#
# Residuals:
#     Min      1Q  Median      3Q     Max
# -1.6329 -0.3963 -0.0849  0.4677  2.1067
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept) -0.05135    0.06771  -0.758     0.45
# z            0.65583    0.07364   8.906 2.86e-14 ***
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.6752 on 98 degrees of freedom
# Multiple R-squared:  0.4473,  Adjusted R-squared:  0.4417
# F-statistic: 79.31 on 1 and 98 DF,  p-value: 2.861e-14
```

```
R> fitted_x <- fitted(mod_first_stage)
R> mod_second_stage <- lm(y ~ fitted_x, data = data)
R> coef(mod_second_stage)
```

```
# (Intercept)    fitted_x
#   0.8975107   1.9369539
```

```
R> ### mit der Paketfunktion
R> mod_iv <- AER::ivreg(y ~ x | z, data = data)
R> coef(mod_iv)
```

```
# (Intercept)           x
#   0.8975107   1.9369539
```

6. Angenommen, y in Modell (1) ist nun binär. Welche der Annahmen im klassischen linearen Regressionsmodells gelten dann nicht mehr?

- MLR.1-4 können weiterhin erfüllt sein
- MLR.5 und MLR.6 gelten nicht mehr (siehe Vorlesung)

7. Welche neue Interpretation bekäme $\hat{\beta}_1$?

Verändert sich x, so verändert sich nicht der (relative) Wert von y, sondern die **Wahrscheinlichkeit**, dass $y = 1$ eintritt. Jedoch sind die Änderungen nicht mehr linear, sodass eine gewohnte ceteris paribus nicht möglich ist und wir andere Methoden (APE, PEA, etc.) benötigen.