Lennart Oelschläger (lennart.oelschlaeger@uni-bielefeld.de)
310214 Praktische Übung zu "Einführung in die Mikroökonometrie"
Wintersemester 2023/2024

Aufgabenblatt 3

Lösungen

Überblick

Wir haben mit den *fixen (fixed, FE)* und *zufälligen (random, RE) Effekten (effects)* zwei neue Panelmethoden kennengelernt, zusätzlich zu der *erste Differenz (first difference, FD)* Methode. Die Motivation ist die gleiche: Wir vermuten, dass unbeobachtete Einflussfaktoren existieren, die entweder konstant über die Zeit sind (die so genannten fixen Effekte a_i) oder über die Zeit variieren (die so genannten *idiosynkratischen* Fehler u_{it}). Wenn wir nun ein Modell mittels pooled OLS anpassen, müssten wir annehmen, dass beide Typen unkorreliert mit den beobachteten Regressoren sind. Zumindest für die a_i ist das oft unrealistisch und würde dann zu einer verzerrten Schätzung führen. Hier setzen die Panelmethoden an, die (unter unterschiedlichen Voraussetzungen) effiziente Schätzung ermöglichen.

Aufgaben

1. Warum geben wir uns mit der FD Methode noch nicht zufrieden?

Damit der FD-Schätzer optimal ist, muss insbesondere die Annahme $\text{Cov}(\Delta u_{i,t}, \Delta u_{i,s} \mid X_i) = 0$ gelten. Das ist eine sehr spezielle Annahme, die oft gar nicht erfüllt ist. Wenn die Annahme verletzt ist, gibt es andere Schätzer, die eine kleinere Varianz besitzen.

2. Was ist die Idee der *Within Transformation* und wie funktioniert sie?

Wir subtrahieren von den Regressoren $X_{i,t}$ und der abhängigen Variable $y_{i,t}$ jeweils das zeitliche Mittel. Dadurch werden die fixen Effekte a_i eliminiert und wir können die bewährte KQ-Schätzung durchführen. Dieses Verfahren heißt *within transformation*, oder auch *fixed effects estimator*, oder *least-squares dummy variable (LSDV) estimator*.

3. Bitte vergleichen Sie die Methoden FE und FD miteinander; wann bevorzugen Sie welche?

Bei $T = 2$ sind beide Methoden äquivalent. Auch sind ihre ersten vier Annahmen FD.1 - 4 und FE.1 - 4 jeweils identisch; genau dann, wenn die eine Methode unverzerrt ist, ist also auch die andere Methode unverzerrt. Die anderen Voraussetzungen zeigen aber an, in welcher Situation welche Methode effizienter ist. Wenn der Fehler $u_{i,t}$ unkorreliert über die Zeit ist, ist FD.6 verletzt und FE effizienter. Umgekehrt, wenn $u_{i,t}$ zum Beispiel ein random walk ist, dann ist $\Delta u_{i,t}$ unkorreliert über die Zeit unkorreliert und FD ist der BLUS.

4. Uns liegen mit der Datei `unfaelle.csv` (siehe Moodle) Daten über die Anzahl der bei Autounfällen tödlich Verunglückten in den USA vor. Wir werden im Folgenden die Anzahl der Todesfälle pro 10.000 Einwohner untersuchen, das ist die Variable `tote_p10k` im Datensatz. Bitte lesen Sie die Daten ein und verschaffen Sie sich einen Überblick.

```
R> unfaelle <- read.csv("unfaelle.csv")
R> str(unfaelle)
```

```
# 'data.frame': 336 obs. of  7 variables:
#  $ staat        : chr  "AL" "AL" "AL" "AL" ...
#  $ jahr         : int  1982 1983 1984 1985 1986 1987 1988 1982 1983 1984 ...
#  $ tote_p10k    : num  2.13 2.35 2.34 2.19 2.67 ...
#  $ biersteuer   : num  1.54 1.79 1.71 1.65 1.61 ...
#  $ alo          : num  14.4 13.7 11.1 8.9 9.8 ...
#  $ lpkopf_einkom: num  9.26 9.28 9.32 9.34 9.36 ...
#  $ alk_alter    : num  19 19 19 19.7 21 ...
```

Es handelt sich um einen *balanzierten* Paneldatensatz, bei dem die Zeitdimension T für jeden Merkmalsträger (hier die Staaten) gleich groß ist.

5. Schätzen Sie das Modell

$$\texttt{tote\_p10k} = \beta_0 + \beta_1 \texttt{biersteuer} + u \tag{1}$$

für die Jahre 1982 und 1988, einmal mit pooled OLS und einmal mit dem FD Schätzer.[1] Interpretieren Sie jeweils $\hat{\beta}_1$ und die auftretenden Unterschiede.

```
R> pooled <- lm(
+   formula = tote_p10k ~ biersteuer,
+   data = unfaelle,
+   subset = (jahr %in% c(1982, 1988))
+ )
R> summary(pooled)
```

```
#
# Call:
# lm(formula = tote_p10k ~ biersteuer, data = unfaelle, subset = (jahr %in%
#     c(1982, 1988)))
#
# Residuals:
#      Min       1Q   Median       3Q      Max
# -0.88990 -0.40554 -0.09281  0.27935  2.20921
#
# Coefficients:
#             Estimate Std. Error t value Pr(>|t|)
# (Intercept)   1.9438     0.0872  22.290   <2e-16 ***
```

[1] *Tipp:* Die FD Schätzung kann mittels der `lm()` Funktion berechnet werden. Dafür müssen wir aber vorher händisch die Datendifferenzen bilden. Die Funktion `plm::plm()` tut dies automatisch für uns und kann auch unsere anderen Panelmethoden FE und RE bearbeiten (`plm` steht für linear models for panel data).

```
# biersteuer    0.2685     0.1258   2.134   0.0355 *
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Residual standard error: 0.5851 on 94 degrees of freedom
# Multiple R-squared:  0.04619, Adjusted R-squared:  0.03604
# F-statistic: 4.552 on 1 and 94 DF,  p-value: 0.03548
```

```
R> # install.packages("plm")
R> library("plm")
R>
R> fd <- plm(
+   formula = tote_p10k ~ biersteuer,
+   model = "fd",
+   data = unfaelle,
+   subset = (jahr %in% c(1982, 1988))
+ )
R> summary(fd)
```

```
# Oneway (individual) effect First-Difference Model
#
# Call:
# plm(formula = tote_p10k ~ biersteuer, data = unfaelle, subset = (jahr %in%
#     c(1982, 1988)), model = "fd")
#
# Balanced Panel: n = 48, T = 2, N = 96
# Observations used in estimation: 48
#
# Residuals:
#      Min.   1st Qu.    Median   3rd Qu.      Max.
# -1.227155 -0.096189  0.092121  0.222901  0.677450
#
# Coefficients:
#              Estimate Std. Error t-value Pr(>|t|)
# (Intercept) -0.072037   0.060644 -1.1879  0.24098
# biersteuer  -1.040973   0.417228 -2.4950  0.01625 *
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Total Sum of Squares:    8.1082
# Residual Sum of Squares: 7.1418
# R-Squared:      0.11919
# Adj. R-Squared: 0.10005
# F-statistic: 6.2249 on 1 and 46 DF, p-value: 0.016248
```

Der geschätzte Parameter $\hat{\beta}_1$ erklärt, wie sich die durchschnittliche Anzahl Verkehrstoter ändert, wenn die Biersteuer um 1 USD erhöht wird. Überraschenderweise ist der Schätzwert im gepoolten Modell positiv. Über die Bildung der Differenzen haben wir den Effekt der un-

beobachteten Heterogenität (zum Beispiel soziale Akzeptanz von Fahren unter Alkohol) kontrolliert. Damit verschwindet die Verzerrung des KQ-Schätzers, die auf die Nichtberücksichtigung dieses Effekts zurückgeht. Im FD Modell ist $\hat{\beta}_1$ (wie erwartet) negativ.

6. Wir erweitern das Modell (1) um die Arbeitslosenquote des Staates `alo` in Prozentpunkten, das logarithmierte durchschnittliche Pro-Kopf-Einkommen `lpkopf_einkom` in USD sowie eine Variable für das Alter, ab dem Alkohol legal konsumiert werden darf, `alk_alter`. Bitte führen Sie die FE-Schätzung durch, die alle Jahre 1982 bis 1988 berücksichtigt.

```
R> within <- plm(
+   formula = tote_p10k ~ biersteuer + alo + lpkopf_einkom + alk_alter,
+   model = "within",
+   data = unfaelle
+ )
R> summary(within)
```

```
# Oneway (individual) effect Within Model
#
# Call:
# plm(formula = tote_p10k ~ biersteuer + alo + lpkopf_einkom +
#     alk_alter, data = unfaelle, model = "within")
#
# Balanced Panel: n = 48, T = 7, N = 336
#
# Residuals:
#        Min.     1st Qu.      Median     3rd Qu.        Max.
# -0.52983468 -0.07514464  0.00067225  0.07880929  0.84570756
#
# Coefficients:
#                Estimate Std. Error t-value Pr(>|t|)
# biersteuer    -0.373345   0.190890 -1.9558  0.05147 .
# alo           -0.018801   0.011462 -1.6402  0.10207
# lpkopf_einkom  0.642873   0.370267  1.7362  0.08361 .
# alk_alter     -0.037392   0.019735 -1.8946  0.05915 .
# ---
# Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#
# Total Sum of Squares:    10.785
# Residual Sum of Squares: 9.4329
# R-Squared:      0.12536
# Adj. R-Squared: -0.031708
# F-statistic: 10.1761 on 4 and 284 DF, p-value: 1.033e-07
```

7. Welche Vorteile hat RE gegenüber FE, und wo liegen die Nachteile?

Der RE Schätzer erlaubt Regressoren, die konstant über die Zeit sind. Das war bei FD und FE nicht möglich. Für die RE Schätzung ist die zentrale Annahme, dass $\text{Cov}(X_{i,t}, a_i) = 0$ gilt; nur in diesem Fall führt die Methode zu einer unverzerrten Schätzung. Wenn diese Annahme nicht

erfüllt ist (und das ist kein unwahrscheinlicher Fall), dann sollte auf FD oder FE zurückgegriffen werden. Zentral bei der RE Methode ist die Schätzung der Konstanten

$$\theta = 1 - \frac{\sigma_u}{\sqrt{\sigma_u^2 + T\sigma_a^2}} \in [0, 1],$$

wobei σ_u die Standardabweichung des Fehlers, σ_a die Standardabweichung des fixen Effekts und T die Paneldimension bezeichnet. Wie genau θ geschätzt werden kann, wird in der Vorlesung beschrieben. In einem nächsten Schritt wird (wie bei der FE Methode) von den Daten das zeitliche Mittel subtrahiert, welches (und das ist der entscheidende Unterschied) vorher mit θ gewichtet wird. Dann sind die Fehler unkorreliert und es kann OLS verwendet werden. Es gibt zwei Spezialfälle:

- Wenn $\theta \approx 0$, dann entspricht RE dem pooled OLS. Das ist zum Beispiel der Fall, wenn σ_u viel größer als σ_a ist.
- Wenn $\theta \approx 1$, dann entspricht RE dem FE. Das ist zum Beispiel der Fall, wenn σ_a viel größer als σ_u ist oder wenn $T \to \infty$.

In der Vorlesung wird ein Testverfahren beschrieben, mit dem man zwischen fixen, zufälligen und nicht-existenten unbeobachteten Effekten unterscheiden kann (siehe dazu auch die Funktion `plm::plmtest()`).

8. Schätzen Sie das Modell aus Aufgabe 6 erneut mit RE.

```
R> random <- plm(
+   formula = tote_p10k ~ biersteuer + alo + lpkopf_einkom + alk_alter,
+   model = "random",
+   data = unfaelle
+ )
R> summary(random)

# Oneway (individual) effect Random Effect Model
#    (Swamy-Arora's transformation)
#
# Call:
# plm(formula = tote_p10k ~ biersteuer + alo + lpkopf_einkom +
#     alk_alter, data = unfaelle, model = "random")
#
# Balanced Panel: n = 48, T = 7, N = 336
#
# Effects:
#                    var std.dev share
# idiosyncratic 0.03321 0.18225 0.143
# individual    0.19836 0.44538 0.857
# theta: 0.8472
#
# Residuals:
#      Min.   1st Qu.    Median   3rd Qu.      Max.
# -0.476767 -0.113298 -0.022599  0.087989  0.913623
```

```
#
# Coefficients:
#                Estimate Std. Error z-value  Pr(>|z|)
# (Intercept)    5.352702   3.109154  1.7216 0.0851430 .
# biersteuer     0.055806   0.122355  0.4561 0.6483165
# alo           -0.041166   0.010789 -3.8157 0.0001358 ***
# lpkopf_einkom -0.283414   0.325582 -0.8705 0.3840362
# alk_alter     -0.016562   0.019968 -0.8294 0.4068678
# ---
# Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
#
# Total Sum of Squares:    13.077
# Residual Sum of Squares: 12.162
# R-Squared:      0.070004
# Adj. R-Squared: 0.058766
# Chisq: 24.9156 on 4 DF, p-value: 5.2314e-05
```