

Aufgabenblatt 3

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?
2. Was ist die Idee der *Within Transformation* und wie funktioniert sie?
3. Bitte vergleichen Sie die Methoden FE und FD miteinander; wann bevorzugen Sie welche?
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.
5. Schätzen Sie das Modell

$$\text{tote_p10k} = \beta_0 + \beta_1 \text{biersteuer} + u \quad (1)$$

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.

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.
7. Welche Vorteile hat RE gegenüber FE, und wo liegen die Nachteile?
8. Schätzen Sie das Modell aus Aufgabe 6 erneut mit RE.

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).