

Themen

Inhaltsverzeichnis

1 Das Random-Coefficient Modell

1.1 Allgemeine Grundlagen

Two-Stage-Specification

Bei der „Two-Stage-Specification“ wird das Random-Coefficient Modell in mehreren Gleichungen spezifiziert. Zunächst das Level-1-Modell:

$$y_{it} = b_{0i} + b_{1i}x_{it} + \epsilon_{it} \quad (1)$$

Es gibt hier also je eine Konstante und je ein Koeffizient für jede Ausprägung der Level-1-Variable (d.h. für jede Person beim Panel, und jedes Land beim Ländervergleich).

b_{0i} und b_{1i} sind die abhängigen Variablen der Level-2-Modelle:

$$b_{0i} = c_0 + c_1z_i + \zeta_i \quad (2)$$

$$b_{1i} = d_0 + d_1z_i + \xi_i \quad (3)$$

Eine empirische Illustration

Mit herkömmlichen Methoden lässt sich die Idee des “Random-Coefficient-Modells” so illustrieren:

```
. use data/eqls_4, clear
. statsby _b, by(s_centry) saving(x, replace): reg q31 hhinc4
. collapse (mean) gdppcap1, by(s_centry)
. merge s_centry using x, sort
. reg _b_cons gdppcap1
. reg _b_hhinc4 gdppcap1
```

Beachte: Im Gegensatz zu (21) kann sich $\text{Var}(\epsilon_{it})$ zwischen den verschiedenen Level-1-Modellen unterscheiden.

Reduzierte Form

- (22) und (23) in (21):

$$y_{it} = \underbrace{c_0 + c_1z_i + \zeta_i}_{b_{0i}} + \underbrace{(d_0 + d_1z_i + \xi_i)}_{b_{1i}}x_{it} + \epsilon_{it} \quad (4)$$

- Ausmultiplizieren:

$$y_{it} = c_0 + c_1z_i + \zeta_id_0x_{it} + d_1z_ix_{it} + \xi_ix_{it} + \epsilon_{it} \quad (5)$$

- Umorganisieren:

$$y_{it} = c_0 + c_1z_i + d_0x_{it} + d_1z_ix_{it} + \zeta_i + \epsilon_{it} + \xi_ix_{it} \quad (6)$$

- Umbenennen:

$$y_{it} = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + \zeta_i + \epsilon_{it} + \xi_ix_2 \quad (7)$$

Zusammenfassung

- Unterschied von „Multilevel-Modell“ und herkömmlicher Regression nur im komplizierteren Fehlerterm ($\zeta_i + \epsilon_{it} + \xi_i x_2$).
- Multilevel Retorik: „Y = Individuell + Kontext“ kann nicht nur in „Multilevel-Modellen“ modelliert werden.
- Fraglich ist, welche inhaltliche Bedeutung dem Fehlerterm zukommt.

1.2 Multilevel vs. Standard-Regression

Cross-Level Einflüsse

$$y_{ij} = a_j + b \times \text{Einkommen} + \epsilon_{ij}$$
$$a_j = c_0 + c_1 \times \text{GDP} + \zeta_j$$

- Reduced Form

$$y_i = b_0 + b_1 \times \text{Einkommen} + b_2 \times \text{GDP} + \zeta_j + \epsilon_{ij}$$

- Ignorieren der Autokorrelation)

```
. use data/eqls_4, clear  
. reg q31 hhinc4 gdppcap1
```

- Berücksichtigen des Fehlerterms

```
. xtreg q31 hhinc4 gdppcap1, i(s_centry) mle  
. xtmixed q31 hhinc4 gdppcap1 || s_centry: , mle
```

Beachte: Unterschiede hauptsächlich bei t -Werten.

X auf Y abhängig vom Kontext

$$y_{ij} = a_j + b_j \times \text{Einkommen} + \epsilon_{ij}$$
$$a_j = c_0 + c_1 \times \text{GDP} + \zeta_j$$
$$b_j = d_0 + d_1 \times \text{GDP} + \xi_j$$

- Reduced Form

$$y_i = b_0 + b_1 \times \text{Einkommen} + b_2 \times \text{GDP} + b_3 \times \text{GDP} \times \text{Einkommen} + \zeta_i + \epsilon_{it} + \xi_i \times \text{Einkommen}$$

```

. gen ia = hhinc4 * gdppcap1
. reg q31 hhinc4 gdppcap1 ia
. xtreg q31 hhinc4 gdppcap1 ia, i(s_cntry) mle
. xtmixed q31 hhinc4 gdppcap1 ia || s_cntry: , mle
. xtmixed q31 hhinc4 gdppcap1 ia || s_cntry: gdppcap1

```

“Exogenisieren der Varianz des Koeffizienten”

$$y_{ij} = a_j + b_j \times \text{Einkommen} + \epsilon_{ij}$$

$$a_j = c_0 + c_1 \times \text{GDP} + \zeta_j$$

$$b_j = d_0 + d_1 \times \text{GDP} + d_2 \times \text{Popdens} + \xi_j$$

- Reduced Form

$$y_i = b_0 + b_1 \times \text{Einkommen} + b_2 \times \text{GDP} + b_3 \times \text{GDP} \times \text{Einkommen} \\ + b_4 \times \text{Popdens} \times \text{Einkommen} + \zeta_i + \epsilon_{it} + \xi_i \times \text{Einkommen}$$

```

. gen ia2 = hhinc4 * popudens
. reg q31 hhinc4 gdppcap1 ia ia2
. xtreg q31 hhinc4 gdppcap1 ia ia2, i(s_cntry) mle
. xtmixed q31 hhinc4 gdppcap1 ia ia2 || s_cntry: gdppcap1

```

Gründe für Fehlerterm

- Signifikanztest Kontextvariablen (aber: Allgemeine Voraussetzungen der Interpretation des Signifikanztests sind zu beachten)
- Aussagen über Varianz der Koeffizienten (aber: bei ausreichenden Daten kann diese Varianz empirisch bestimmt werden)
- Kontrolle unbeobachteter Heterogenität (aber: Fixed effects modell ist wirkungsvoller)

Also see [beamerbibpoetter97Pötter and Rohwer, 1997]

1.3 Anwendungsbeispiel Paneldaten

Exploration

- Steigt das Einkommen von Berufsanfängern mit dem Alter?

```

. use data/sitz07_8_panel, clear
. gen age = welle-gebjahr
. keep if inrange(age,25,35)
. by persnr (unemp), sort: keep if unemp[_N] < 1
. sort persnr age
. line income age if inrange(age,25,35), c(L) by(sex)

```

Random Intercept Model

```
. xtmixed income age || persnr:
```

- Alters-Koeffizienten sind signifikant positiv
- Einkommen der Berufsanfänger variieren

Random Coefficient Model

```
. sum age  
. replace age = age-r(mean)  
. xtmixed income age || persnr: age, cov(unstr) mle
```

- Modellfit besser: Log-Likelihood steigt um ca. 200
- Einkommen der Berufsanfänger (im Alter 0) variieren. Kovariate Geschlecht?
- Anstieg variiert zwischen Personen
- Im Alter 0: Je höher das Einkommen, desto höher der Zuwachs

Kovariatenkontrolle

```
. gen men = sex==1  
. xtmixed income age men || persnr: age, cov(unstr) mle  
. xtmixed income age men whours || persnr: age, cov(unstr) mle
```

- Log-Likelihood steigt bei Berücksichtigung von der Arbeitsstunden
- Arbeitsstunden/Geschlecht erklären nicht die Personenunterschiede
- Es gilt weiter: Je höher das Einkommen, desto höher der Zuwachs

Interaktionseffekte

```
. gen menage = men * age  
. xtmixed income age men whours menage || persnr: age, cov(unstr) mle
```

- Log-Likelihood bleibt nahezu identisch
- Interaktion nicht signifikant
- Einkommensanstieg der Männer nicht schneller als bei Frauen

Literatur

- [Achen, 1986] Achen, C. H. (1986). *The Statistical Analysis of Quasi-Experiments*. University of California Press, Berkeley.
- [Allison, 1994] Allison, P. D. (1994). Using Panel Data to Estimate the Effects of Events. *Sociological Methods and Research*, 23:174–199.
- [Baum, 2006] Baum, C. F. (2006). *An Introduction to modern Econometrics Using Stata*. Stata Press, College Station.
- [King et al., 1994] King, G., Keohane, R. O., and Verba, S. (1994). *Designing Social Inquiry*. Princeton University Press, Princeton.
- [Pötter and Rohwer, 1997] Pötter, U. and Rohwer, G. (1997). Regression Models for Clustered and Interdependent Observations. Paper prepared for the LWBB-Workshop on “Multilevel Analysis in Longitudinal Data”, Potsdam, November 29–29, 1997. <http://steinhaus.stat.ruhr-uni-bochum.de/papers/mlev.ps>.
- [Rabe-Hesketh and Skrondal, 2005] Rabe-Hesketh, S. and Skrondal, A. (2005). *Multilevel and Longitudinal Modeling Using Stata*. Stata-Press, College Station.
- [Reichardt, 1979] Reichardt, C. S. (1979). The Statistical Analysis of Data From Nonequivalent Group Designs. In Cook, T. D. and Campbell, D. T., editors, *Quasi-Experimentation: Design and Analysis Issues for Field Settings*, pages 147–206. Rand McNally, Chicago.
- [Wooldridge, 2002] Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. Massachusetts Institute of Technology.