

Einführung in die Methode der Panelregression

RatSWD – Nachwuchsworkshop

Längsschnittanalysen auf der Basis amtlicher Sozial- und
Wirtschaftsdaten

Axel Werwatz
Stefan Mangelsdorf

$$\{y_{it}, \mathbf{x}_{it}\}_{i=1, \dots, N} \quad t=1, \dots, T$$

- N ist groß (N=75240 in AFiD)
- T ist klein (T=12 in AFiD)
- Formeln für balanced panel (jedes i in jeder Periode; unrealistisch aber hilfreich hier)
- zeitkonstante und zeitveränderliche Variablen

Wozu Paneldaten?

Josef Brüderl:

1. zur Schätzung kausaler Effekte trotz unbeobachteter Heterogenität (Marriage-Premium for Men?)
2. zur Analyse individueller Dynamik
3. zur präziseren Schätzung (mehr Beobachtungen)

zu 3: „Auf Grund der inzwischen oft anzutreffenden Stichprobengrößen in praktischen Anwendungen sind mögliche Effizienzgewinne eher von rein theoretischem Interesse“ Lechner (Asta 2002)

Beispiel zu 1): kausale Effekte

Forschungsfrage: Sind exportierende Betriebe produktiver als nichtexportierende?

$$\ln ap_{it} = \beta_0 + \beta_1 exp_{it} + \mathbf{w}_{it} \boldsymbol{\gamma} + v_{it}$$

Diagram illustrating the components of the equation:

- $\ln ap_{it}$: ln(Arbeitsproduktivität)
- exp_{it} : Exporteur in t -Dummy
- $\mathbf{w}_{it} \boldsymbol{\gamma}$: „Kontrollvariablen“
„beobachtete Heterogenität“
- v_{it} : Fehlerterm

Beispiel zu 1): kausale Effekte

Forschungsfrage: Sind exportierende Betriebe produktiver als nichtexportierende?

$$\ln ap_{it} = \beta_0 + \beta_1 exp_{it} + \mathbf{w}_{it} \boldsymbol{\gamma} + v_{it}$$

Linearer Ansatz mit zeitkonstanten Koeffizienten.
Fokus auf β_1 .

Das ist noch kein Modell. Wichtig: Annahmen über v_{it} .



Beispiel und Panel-Modell Notation (1)

$$\ln ap_{it} = \underbrace{\beta_0 + \beta_1 \exp_{it} + \mathbf{w}_{it} \boldsymbol{\gamma}} + v_{it}$$

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + v_{it}$$

$$= \mathbf{x}_{it} \boldsymbol{\beta} + c_i + u_{it}$$

Konstante(n),
Treatment-Variable (\exp_{it}),
Kontrollvariablen (\mathbf{w}_{it})
alle in \mathbf{x}_{it} subsumiert

idiosynchratische
Fehlerkomponente

zeitkonstante, individuenspezifische
Komponente des Fehlerterm,
„unbeobachtete Heterogenität“



Beispiel und Panel-Modell Notation (2)

$$\ln ap_{it} = \underbrace{\beta_0 + \beta_1 exp_{it} + \mathbf{w}_{it} \boldsymbol{\gamma}} + v_{it}$$

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + v_{it}$$

$$= \mathbf{x}_{it} \boldsymbol{\beta} + c_i + u_{it}$$

Notation von Josef Brüderl

$$y_{it} = \beta_1 x_{it} + u_{it}$$

$$u_{it} = v_i + \varepsilon_{it}$$

Unsere Notation folgt weitestgehend

Jeffrey Wooldridge, *Econometric Analysis of Cross Section and Panel Data*, MIT Press 2002.

Beispiel und Panel-Modell

$$\begin{aligned} \ln ap_{it} &= \underbrace{\beta_0 + \beta_1 exp_{it} + \mathbf{w}_{it} \boldsymbol{\gamma}} + v_{it} \\ y_{it} &= \mathbf{x}_{it} \boldsymbol{\beta} + c_i + u_{it} \end{aligned}$$

Forschungsziel: Partieller Effekt von *exp* auf Mittelwert von *lnap*
gegeben Kontrollvariablen **und** unbeobachtete Heterogenität

Im Allgemeinen $\frac{\partial}{\partial x_j} E(y_{it} \mid \mathbf{x}_{it}, c_i)$

Im linearen Panel Modell mit idiosynchratischen u_{it} :

$$E(u_{it} \mid \mathbf{x}_{it}, c_i) = 0 \Rightarrow E(y_{it} \mid \mathbf{x}_{it}, c_i) = \mathbf{x}_{it} \boldsymbol{\beta} + c_i$$

$$\Rightarrow \frac{\partial}{\partial x_j} E(y_{it} \mid \mathbf{x}_{it}, c_i) = \beta_j$$

Lineares Panel-Modell

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + v_{it}, \quad v_{it} = c_i + u_{it}$$

$$E(y_{it} \mid \mathbf{x}_{it}, c_i) = \mathbf{x}_{it} \boldsymbol{\beta} + c_i \quad \Rightarrow \quad \frac{\partial}{\partial x_j} E(y_{it} \mid \mathbf{x}_{it}, c_i) = \beta_j$$

$$E(y_{it} \mid \mathbf{x}_{it}) = E_{c_i \mid \mathbf{x}_{it}} [E(y_{it} \mid \mathbf{x}_{it}, c_i)] = \mathbf{x}_{it} \boldsymbol{\beta} + E(c_i \mid \mathbf{x}_{it})$$

$$\text{Im Allgemeinen } \frac{\partial}{\partial x_j} E(y_{it} \mid \mathbf{x}_{it}) = \beta_j + \frac{\partial}{\partial x_j} E(c_i \mid \mathbf{x}_{it}) \neq \beta_j$$

➔ Ignorieren wir c_i , dann ist der partielle Effekt nicht der gewünschte (kausale, originäre, ceteris paribus) Effekt von x_j , geg. beobachtete Kontrollvariablen und unbeobachtete Heterogenität

Lineares Panel-Modell

$$y_{it} = \mathbf{x}_{it} \boldsymbol{\beta} + v_{it}, \quad v_{it} = c_i + u_{it} \quad E(y_{it} | \mathbf{x}_{it}, c_i) = \mathbf{x}_{it} \boldsymbol{\beta} + c_i$$

$$\frac{\partial}{\partial x_j} E(y_{it} | \mathbf{x}_{it}, c_i) = \beta_j \quad \frac{\partial}{\partial x_j} E(y_{it} | \mathbf{x}_{it}) = \beta_j + \frac{\partial}{\partial x_j} E(c_i | \mathbf{x}_{it})$$

Es sei $E(c_i | \mathbf{x}_{it}) = \delta_0 + \delta_1 x_{1,it} + \dots + \delta_j x_{j,it} + \dots + \delta_K x_{K,it}$

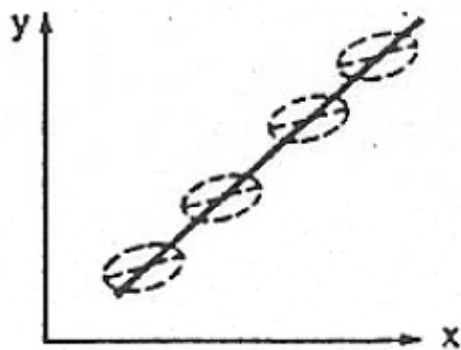
$$\Rightarrow \frac{\partial}{\partial x_j} E(y_{it} | \mathbf{x}_{it}) = \beta_j + \delta_j$$

Im Beispiel: Exportieren tendenziell die „starken“ Betriebe ($\delta_j > 0$)
dann überschätzt die Regression ohne Berücksichtigung von
 c_i im Mittel den „wahren“ Exporteffekt (β_j)

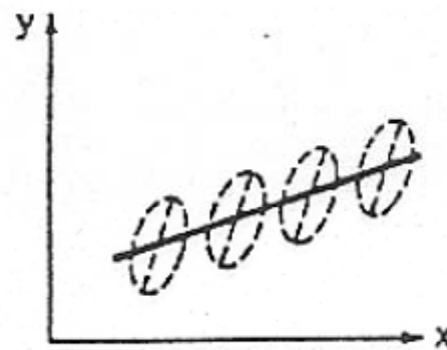
Aber: wenn x_j und c unabhängig (unkorriert) sind, dann ist $\delta_j = 0$
und $\frac{\partial}{\partial x_j} E(y_{it} | \mathbf{x}_{it}) = \beta_j$

Warum c_i wichtig ist..

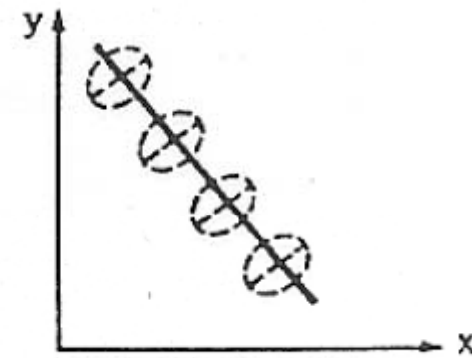
Korrelation zwischen c_i und u_{it} und seine Konsequenzen



1a)



1b)



1c)

Quelle: Hsiao (1986), S.7

Referenzergebnis: Pooled OLS

- OLS Regression von y_{it} auf \mathbf{x}_{it} (d.h. exp_{it} und \mathbf{w}_{it})
- .. als ob $N \cdot T$ unabhängige Querschnittsbeobachtungen zur Verfügung wären.
- Panelstruktur wird ignoriert, insb. $v_{it} = c_i + u_{it}$
- OLS-Koeffizientenschätzer von exp_{it} wahrscheinlich zu hoch
- OLS-Standardfehler sind zu klein, da T -Beobachtungen zum selben i eben keine unabhängige, „frische“ Information sind

Variablen im Exportbeispiel

Abhängige Variable

$$\ln ap_{it} = \ln \left(\frac{\text{Umsatz}_{it}}{\text{tätige Personen}_{it}} \right)$$

Erklärende Variablen

Export-Dummy

$$\exp_{it} = 1 \quad \text{wenn } \text{Auslandsumsatz}_{it} > 0$$

Betriebsgröße:

$$\ln tP_{it} = \ln(\text{tätige Personen}_{it})$$

Qualität der Belegschaft:

$$\text{arbeiterant}_{it} = \frac{\text{Arbeiter}_{it}}{\text{tätige Personen}_{it}}$$

$$\text{lohn}_{it} = \frac{\text{Lohnsumme}_{it}}{\text{tätige Personen}_{it}} - \text{Durchschnittslohn}(\text{HG}, \text{Region})_t$$

Dummies für

- Osten
- Hauptgruppen (IG, GG, VG, Basis VL)
- für hochwertige und Spitzentechnologie
- für Mehrbetriebsunternehmen
- für die Jahre 1996 bis 2006.



Pooled-OLS Schätzung (1)

```
reg lnape exp lntP arbeiterant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu jahr1996- jahr2005
```

Source	SS	df	MS	Number of obs = 490350		
				F(22,490327) = 8418.99		
Model	76271.5849	22	3466.89022	Prob > F = 0.0000		
Residual	201913.723490327	.411794013		R-squared = 0.2742		
				Adj R-squared = 0.2741		
Total	278185.308490349	.567321047		Root MSE = .64171		
lnape	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
exp	.2261	.00213	106.15	0.000	.2219	.2302
lntP	.0106877	.0009545	11.20	0.000	.0088169	.0125586
arbeiterant	-.2059455	.0053656	-38.38	0.000	-.2164619	-.1954291
lohn	.0002943	1.01e-06	291.99	0.000	.0002923	.0002963
hgd2	-.1745256	.0024932	-70.00	0.000	-.1794121	-.1696391
hgd3	-.2265739	.0046738	-48.48	0.000	-.2357344	-.2174135
hgd4	-.2309772	.0025499	-90.58	0.000	-.235975	-.2259794
ost	-.2075498	.0024272	-85.51	0.000	-.212307	-.2027926
techd1	-.1237306	.0049172	-25.16	0.000	-.1333681	-.114093
techd2	-.0394741	.0028505	-13.85	0.000	-.045061	-.0338872
:						

Pooled-OLS Schätzung (2)

```
reg lnap exp lntP arbeitertant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu jahr1996-  
jahr2005, cluster(bnr)
```

Linear regression

```
Number of obs = 490350
```

$$F(22, 68891) = 1037.78$$

Prob > F = 0.0000

(Std. Err. adjusted for 68892 clusters in bnr)

	lnap	Coef.	Robust Std. Err.	t	P> t	[95% Conf. Interval]
	exp	.2261	.005198	43.50	0.000	.2159 .2363
	lntP	.0106877	.0031677	3.37	0.001	.004479 .0168965
	arbeiterant	-.2059455	.0175763	-11.72	0.000	-.2403951 -.1714959
	lohn	.0002943	8.81e-06	33.42	0.000	.000277 .0003116
	hgd2	-.1745256	.0051605	-33.82	0.000	-.1846402 -.164411
	hgd3	-.2265739	.0092187	-24.58	0.000	-.2446425 -.2085053
	hgd4	-.2309772	.0075909	-30.43	0.000	-.2458553 -.2160991
	ost	-.2075498	.0066992	-30.98	0.000	-.2206803 -.1944193
	techd1	-.1237306	.0099162	-12.48	0.000	-.1431663 -.1042948
	techd2	-.0394741	.0059907	-6.59	0.000	-.0512158 -.0277324
	⋮					



Pooled-OLS Schätzung (2)

```
reg lnap exp lntP arbeitant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu jahr1996-  
jahr2005 cluster(bnr)
```

Linear regression

Number of obs = 490350

F(22, 68891) = 1037.78

Prob > F = 0.0000

(Std. Err. adjusted for 68892 clusters in bnr)

		Robust				
lnap		Coef.	Std. Err.	t	P> t	[95% Conf. Interval]

exp		.2261	.005198	43.50	0.000	.2159 .2363

arb Zum Vergleich

exp		.2261	.00213	106.15	0.000	.2219	.2302
hgd3		-.2265739	.0092187	-24.58	0.000	-.2446425	-.2085053
hgd4		-.2309772	.0075909	-30.43	0.000	-.2458553	-.2160991
ost		-.2075498	.0066992	-30.98	0.000	-.2206803	-.1944193
techd1		-.1237306	.0099162	-12.48	0.000	-.1431663	-.1042948
techd2		-.0394741	.0059907	-6.59	0.000	-.0512158	-.0277324

⋮

Pooled OLS

Unter welchen Bedingungen ist Pooled OLS konsistent?

Annahme POLS.1:

$$E(\mathbf{x}'_{it} v_{it}) = \mathbf{0} \quad t = 1, 2, \dots, T$$

- Das beinhaltet:

$$E(\mathbf{x}'_{it} u_{it}) = \mathbf{0} \quad \text{und} \quad E(\mathbf{x}'_{it} c_i) = \mathbf{0} \quad \forall t$$

D.h., erklärende Variablen und c_i sind unkorreliert

Panel-Regression

Modelle/Schätzer,
die Panelstruktur ($v_{it} = c_i + u_{it}$)
berücksichtigen

Fixed Effects Modell

$E(c_i | \mathbf{x}_{it})$ beliebig bzw.

$Cov(c_i, x_{j,it})$ beliebig

- Within-Schätzer
- First-Difference Schätzer

Random Effects Modell

$E(c_i | \mathbf{x}_{it}) = 0$

$Cov(c_i, x_{j,it}) = 0$

- pooled OLS (consistent)
- pooled GLS (efficient)

Random oder Fixed Effects?

- Traditionell wird c_i bezeichnet als
 - *Random Effect*, wenn es wie eine Zufallsvariable behandelt wird.
 - *Fixed Effect*, wenn es wie ein Parameter behandelt wird, der für jedes Individuum i geschätzt werden kann.
- In mikroökonometrischen Panels mit einer großen Zahl von Zufallsziehungen aus der Grundgesamtheit macht es fast immer Sinn, die unbeobachteten Effekte als Zufallsvariablen zu behandeln.

Random oder Fixed Effects?

- In der modernen Ökonometrie ist die Schlüsselfrage, ob c_i korreliert ist mit den beobachteten erklärenden Variablen oder nicht:
 - *Random Effect* wenn keine Korrelation vorliegt:
$$\text{Cov}(\mathbf{x}_{it}, c_i) = \mathbf{0}, \quad t = 1, 2, \dots, T$$
 - *Fixed Effect* bedeutet, dass man Korrelationen zwischen c_i und \mathbf{x}_{it} erlaubt



FE vs RE im Exportbeispiel

Was wäre in unserem Beispiel plausibel?

- Unbeobachtete Effekte, die die Produktivität beeinflussen, könnten z.B. Managementqualität, spezielles Know-How oder gute Kapitalausstattung sein.
- Diese Merkmale könnten durchaus mit dem Exportverhalten korreliert sein. Ein gutes Produkt lässt sich auch im Ausland besser verkaufen.
 - **Fixed Effects Modell** plausibler

Fixed Effects Modell

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

Annahme FE.1: strikte Exogenität

$$E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \quad t = 1, \dots, T$$

$$\text{mit } \mathbf{x}_i = (\mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT})$$

D.h., beliebige Beziehung zwischen \mathbf{x}_{it} und c_i
aber gegeben c_i gibt es **keine** Beziehung
zwischen u_{it} und den \mathbf{x}_{it} **aller** Perioden.

Strikte Exogenität (1)

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

$$E(u_{it} \mid \mathbf{x}_i, c_i) = 0$$

implizieren

$$E(y_{it} \mid \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots, \mathbf{x}_{iT}, c_i) = E(y_{it} \mid \mathbf{x}_{it}, c_i) = \mathbf{x}_{it}\boldsymbol{\beta} + c_i$$

- Die erste Gleichung bedeutet, dass nach Kontrolle von \mathbf{x}_{it} und c_i alle \mathbf{x}_{is} keinen Einfluss mehr auf y_{it} haben für $s \neq t$.
- Die zweite Gleichung beschreibt die funktionale Form von $E(y_{it} \mid \mathbf{x}_{it}, c_i)$.

Strikte Exogenität (2)

$$E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \quad t = 1, \dots, T$$

schließt aus

– verzögerte abhängige Variablen in \mathbf{x}_{it} :

$$y_{it} = \beta_1 y_{it-1} + c_i + u_{it} \quad ?$$

$$E(u_{it} \mid \mathbf{x}_i, c_i) = E(u_{it} \mid y_{iT-1}, \dots, y_{it}, \dots, y_{i1}, c_i) = 0$$

$$E(u_{it} \mid y_{it}, c_i) = E(u_{it} \mid \beta_1 y_{it-1} + c_i + u_{it}, c_i) \neq 0$$

Strikte Exogenität (2)

$$E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \quad t = 1, \dots, T$$

schließt aus

- verzögerte abhängige Variablen in \mathbf{x}_{it} :
- Feedbackeffekte von verzögerten Schocks ($u_{it-1}, u_{it-2}, \dots$) auf Elemente von \mathbf{x}_{it} (so dass $E(u_{it-1} \mid \mathbf{x}_{it}) \neq 0$)

Beispiel: $\ln ap_{it} = \beta_0 + \beta_1 exp_{it} + \mathbf{w}_{it} \boldsymbol{\gamma} + c_i + u_{it}$

- Ist exp_{it} strikt exogen?
- Exporttätigkeit heute könnte von vergangenen Produktivitätsschocks beeinflusst sein.
- Ashenfelter's Dip

Within Transformation

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

gilt auch im Durchschnitt für jedes i :

$$\bar{y}_i = \bar{\mathbf{x}}_i\boldsymbol{\beta} + c_i + \bar{u}_i \quad \text{mit} \quad \bar{y}_i = \frac{1}{T} \sum_{t=1}^T y_{it}$$

Abziehen der Gleichung voneinander ergibt

$$\underbrace{y_{it} - \bar{y}_i}_{\ddot{y}_{it}} = \underbrace{(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)}_{\ddot{\mathbf{x}}_{it}} \boldsymbol{\beta} + \underbrace{u_{it} - \bar{u}_i}_{\ddot{u}_{it}}$$

was die c_i eliminiert, aber auch alle zeitkonstanten erklärenden Variablen.

Within Schätzer („Fixed Effects Schätzer“):

Pooled OLS Schätzer von \ddot{y}_{it} auf $\ddot{\mathbf{x}}_{it}$:

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{FE} &= \left(\sum_{i=1}^N \ddot{\mathbf{x}}_i' \ddot{\mathbf{x}}_i \right)^{-1} \left(\sum_{i=1}^N \ddot{\mathbf{x}}_i' \ddot{y}_i \right) \\ &= \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it} \right)^{-1} \left(\sum_{i=1}^N \sum_{t=1}^T \ddot{\mathbf{x}}_{it}' \ddot{y}_{it} \right)\end{aligned}$$

$\hat{\boldsymbol{\beta}}_{FE}$ ist konsistent unter FE.1 (strikte Exogenität) und einer Rangbedingung .

First Difference Transformation

Bei zwei Perioden können die Gleichungen von Periode 1 und Periode 2 abgezogen werden:

$$\underbrace{y_{it} - y_{it-1}}_{\Delta y_{it}} = \underbrace{(\mathbf{x}_{it} - \mathbf{x}_{it-1})}_{\Delta \mathbf{x}_{it}} \boldsymbol{\beta} + \underbrace{u_{it} - u_{it-1}}_{\Delta u_{it}}$$

Lineares Modell in Differenzen ohne Konstante und ohne c_i aber auch ohne zeitkonstante erklärende Variablen

First Difference Schätzer ist der
Pooled OLS Schätzer der Regression

$$\Delta y_{it} = \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta u_{it}$$

$\hat{\boldsymbol{\beta}}_{FD}$ ist konsistent unter FE.1 (strikte Exogenität)
und einer Rangbedingung .

Es läßt sich zeigen, dass $E(u_{it} \mid \mathbf{x}_i, c_i) = 0, \Rightarrow E(\Delta \mathbf{x}_{it} \Delta u_{it}) = \mathbf{0}$



Within/FE-Schätzer

```
xtreg lnexp lntP arbeiterant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu jahr1996-  
jahr2005, fe
```

Fixed-effects (within) regression

Group variable: bnr

R-sq: within = 0.1085

between = 0.1169

overall = 0.1062

Number of obs = 490350

Number of groups = 68892

Obs per group: min = 1

avg = 7.1

max = 11

F(22,421436) = 2332.55

Prob > F = 0.0000

corr(u_i, Xb) = 0.1393

lnexp	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lnexp	.094	.002232	42.13	0.000	.0897	.0984
lntP	-.0849324	.0018994	-44.72	0.000	-.0886551	-.0812097
arbeiterant	.2467592	.0070208	35.15	0.000	.2329986	.2605198
lohn	.000111	6.12e-07	181.42	0.000	.0001098	.0001122
hgd2	.0440935	.0040725	10.83	0.000	.0361116	.0520754
hgd3	-.04193	.0084138	-4.98	0.000	-.0584208	-.0254392
hgd4	-.0738374	.00393	-18.79	0.000	-.0815401	-.0661348
ost	-.1189403	.086207	-1.38	0.168	-.2879033	.0500227
techd1	-.0001622	.0093895	-0.02	0.986	-.0185654	.018241
techd2	-.0083742	.0043336	-1.93	0.053	-.0168678	.0001194
mbu	.0477343	.0033557	14.22	0.000	.0411572	.0543114
mlu	.0160272	.0029988	5.34	0.000	.0101498	.0219047

⋮



FD-Schätzung

```
reg D.lnap D.(exp lntP arbeiterant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu) jahr1996-  
jahr2005, nocons
```

Source	SS	df	MS
Model	2708.8714	22	123.130518
Residual	30275.9277418861	.072281563	
Total	32984.7991418883	.078744659	

Number of obs = 418883
F(22,418861) = 1703.48
Prob > F = 0.0000
R-squared = 0.0821
Adj R-squared = 0.0821
Root MSE = .26885

D.lnap	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
D1. exp	.0658	.002038	32.28	0.000	.0618	.0698
D1. lntP	-.1131673	.0024425	-46.33	0.000	-.1179546	-.10838
D1.arbeiter	.2615805	.008184	31.96	0.000	.2455402	.2776209
D1. lohn	.000098	5.63e-07	174.18	0.000	.0000969	.0000991
D1. hgd2	.0171386	.0048712	3.52	0.000	.0075913	.026686
D1.hgd3	-.052439	.0097992	-5.35	0.000	-.0716451	-.033233
D1.hgd4	-.0560545	.0050806	-11.03	0.000	-.0660122	-.0460967
D1.ost	-.1005532	.109766	-0.92	0.360	-.3156913	.1145849
D1.techd1	-.0201129	.0102773	-1.96	0.050	-.0402561	.0000303
:						

FE(Within) und zeitkonstante X

$$y_{it} - \bar{y}_i = (\mathbf{x}_{it} - \bar{\mathbf{x}}_i) \boldsymbol{\beta} + u_{it} - \bar{u}_i$$

Within-Transformation eliminiert alle zeitkonstanten erklärenden Variablen

$$y_{it} = \theta_1 + \theta_2 d2_t + \dots + \theta_T d2_T + z_i \gamma_1 + d2_t z_i \gamma_2 + \dots + dT_t z_i \gamma_T + \mathbf{w}_{it} \boldsymbol{\delta} + c_i + u_{it}$$

$\theta_1 + \mathbf{z}_i \gamma_1$ kann nicht von c_i getrennt werden aber:
 Zeitperioden-effekte ($\theta_2, \dots, \theta_T$) und Differenzen der partielle Effekte zeitkonstanter Variablen ($\gamma_2, \dots, \gamma_T$) können (relativ zur Basisperiode) geschätzt werden
 (Bsp: Veränderung des gender wage gap) (Wooldridge 2002, S. 267)

Random Effects Modell

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

Annahme RE.1:

- (a) $E(u_{it} | \mathbf{x}_i, c_i) = 0, \quad t = 1, \dots, T$
- (b) $E(c_i | \mathbf{x}_i) = E(c_i) = 0$

Zusätzlich zur strikten Exogenität dürfen erklärende Variablen nicht mit c_i korreliert sein.



Schätzung im Random Effects Modell

Unter Annahme RE.1 gilt

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \quad E(v_{it} | \mathbf{x}_i) = 0, \quad t = 1, \dots, T$$

→ Pooled OLS ist konsistent.

Aber: selbst wenn

$$E(u_{it}u_{its}) = 0 \text{ für alle } t \neq s \text{ und } E(u_{it}^2 | \mathbf{x}_i) = E(u_{it}^2) = \sigma_u^2,$$

(d.h., u_{it} seriell unkorreliert und homoskedastisch)

$$E(v_{it}^2) = \sigma_c^2 + \sigma_u^2$$

$$E(v_{it}v_{is}) = E[(c_i + u_{it})(c_i + u_{is})] = E(c_i^2) = \sigma_c^2$$

→ GLS ist effizient



Schätzung im Random Effects Modell

Also

$$\Omega = E(\mathbf{v}_i \mathbf{v}_i') = \begin{pmatrix} \sigma_c^2 + \sigma_u^2 & \sigma_c^2 & \dots & \sigma_c^2 \\ \sigma_c^2 & \sigma_c^2 + \sigma_u^2 & \dots & \vdots \\ \vdots & \vdots & \ddots & \sigma_c^2 \\ \sigma_c^2 & \vdots & \vdots & \sigma_c^2 + \sigma_u^2 \end{pmatrix}$$

und

$$\hat{\boldsymbol{\beta}}_{FE} = \left(\sum_{i=1}^N \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^N \mathbf{X}_i' \hat{\boldsymbol{\Omega}}^{-1} \mathbf{y}_i \right)$$



RE-Schätzung

```
xtreg lnap exp lntP arbeiterant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu jahr1996-  
jahr2005, re
```

Random-effects GLS regression
Group variable: bnr

Number of obs = 490350
Number of groups = 68892

R-sq: within = 0.1036
between = 0.2307
overall = 0.2109

Obs per group: min = 1
avg = 7.1
max = 11

Random effects u_i ~ Gaussian
corr(u_i, X) = 0 (assumed)

Wald chi2(22) = 62877.10
Prob > chi2 = 0.0000

	lnap	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
	exp	.129	.00212	60.89	0.000	.125	.133
	lntP	-.0231146	.0015381	-15.03	0.000	-.0261292	-.0201
	arbeiterant	.0859501	.0063531	13.53	0.000	.0734982	.098402
	lohn	.000124	6.09e-07	203.63	0.000	.0001228	.0001252
	hgd2	-.0160926	.003498	-4.60	0.000	-.0229486	-.0092366
	hgd3	-.1061965	.0071419	-14.87	0.000	-.1201944	-.0921985
	hgd4	-.11685	.0034461	-33.91	0.000	-.1236042	-.1100958
	ost	-.2652447	.00635	-41.77	0.000	-.2776904	-.252799
	techd1	.0028429	.0078006	0.36	0.716	-.012446	.0181318
	techd2	-.0068413	.0038463	-1.78	0.075	-.0143799	.0006973
	mbu	.0857305	.003222	26.61	0.000	.0794154	.0920455
	mlu	.0688241	.0028348	24.28	0.000	.0632681	.0743802
	:						

- FE/FD sind konsistent unter RE-Annahmen
- RE und Pooled OLS sind inkonsistent unter FE-Annahme, falls c_i und \mathbf{x}_{it} korreliert sind.
- Vorbehalt: Linearität und Additivität
- Hausman-Test (vergleicht RE und FE Koeffizienten der zeitvariierenden Regressoren)

$$\left(\hat{\boldsymbol{\beta}}_{1,RE} - \hat{\boldsymbol{\beta}}_{1,FE} \right)^T \left[\hat{V}(\hat{\boldsymbol{\beta}}_{1,RE}) - \hat{V}(\hat{\boldsymbol{\beta}}_{1,FE}) \right]^{-1} \left(\hat{\boldsymbol{\beta}}_{1,RE} - \hat{\boldsymbol{\beta}}_{1,FE} \right)^{H_0} \sim \chi_M$$

Test: Ho: difference in coefficients not systematic

```
chi2(21) = (b-B)'[(V_b-V_B)^(-1)](b-B)
          = 10424.40
Prob>chi2 = 0.0000
```

Vergleich der Schätzungen

	Coef.	S.E.	z	P> z	[95% Conf. Intv]	
OLS	.226	.00520	43.50	0.000	.216	.236
RE	.129	.00212	60.89	0.000	.125	.133
FE	.094	.00223	42.13	0.000	.090	.098
FD	.066	.00204	32.28	0.000	.062	.070

Geschätzter Exporteffekt schwankt zwischen

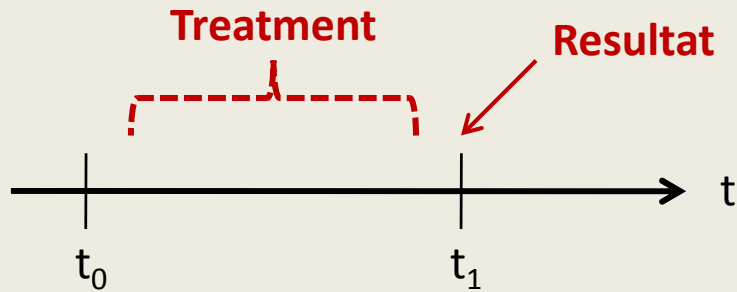
$$e^{\hat{\beta}_{1,FD}} - 1 = e^{0.66} - 1 = 0.068 \quad \text{und} \quad e^{\hat{\beta}_{1,OLS}} - 1 = e^{0.226} - 1 = 0.254$$

OLS und RE ignorieren c_i und überschätzen Exporteffekt.

FE bzw. FD Schätzungen sind überzeugender.

Aber schätzen sie 'kausalen Effekt' (im Sinne von Rubin/Brüderl) ?

Kausaler Effekt im 2-Perioden Sonderfall



id	t	Y	D
i	t_0	Y_{0it_0}	0
i	t_1	Y_{1it_1}	1
\vdots	\vdots	\vdots	\vdots
j	t_0	Y_{0jt_0}	0
j	t_1	Y_{0jt_1}	0
\vdots	\vdots	\vdots	\vdots

$$ATOT = E[Y_{1it_1} - Y_{0it_1} | D_{it_1} = 1]$$

$$= E[Y_{1it_1} | D_{it_1} = 1] - E[Y_{0it_1} | D_{it_1} = 1]$$

$$\hat{ATOT}_{DID} = \frac{1}{N_1} \sum_{i=1}^{N_1} (Y_{1it_1} - Y_{0it_0}) - \frac{1}{N_0} \sum_{j=1}^{N_0} (Y_{0jt_1} - Y_{0jt_0})$$

$$= [(\bar{Y}_{1t_1})_1 - (\bar{Y}_{0t_0})_1] - [(\bar{Y}_{0t_1})_0 - (\bar{Y}_{0t_0})_0]$$



Kausaler Effekt im 2-Perioden Sonderfall

$$\begin{aligned} \hat{ATOT}_{DID} &= \frac{1}{N_1} \sum_{i=1}^{N_1} (Y_{1it_1} - Y_{0it_0}) - \frac{1}{N_0} \sum_{j=1}^{N_0} (Y_{0jt_1} - Y_{0jt_0}) \\ &= \left[(\bar{Y}_{1t_1})_1 - (\bar{Y}_{0t_0})_1 \right] - \left[(\bar{Y}_{0t_1})_0 - (\bar{Y}_{0t_0})_0 \right] \end{aligned}$$

Im 2-Perioden Fall gilt:

DID-Schätzer = FD-Schätzer von β_1 in

$$\Delta y_{it} = \beta_0 + \beta_1 \Delta D_{it} + \Delta u_{it}$$

Heckman et. al. (1997): FE ist der „close cousin“ von DID

DID, FD, FE schätzen kausalen Effekt!

<i>id</i>	<i>t</i>	<i>Y</i>	<i>D</i>
<i>i</i>	<i>t</i> ₀	<i>Y</i> _{0<i>i</i><i>t</i>₀}	0
<i>i</i>	<i>t</i> ₁	<i>Y</i> _{1<i>i</i><i>t</i>₁}	1
⋮	⋮	⋮	⋮
<i>j</i>	<i>t</i> ₀	<i>Y</i> _{0<i>j</i><i>t</i>₀}	0
<i>j</i>	<i>t</i> ₁	<i>Y</i> _{0<i>j</i><i>t</i>₁}	0
⋮	⋮	⋮	⋮

Kausaler Effekt im 2-Perioden Sonderfall

$$\begin{aligned}\hat{ATOT}_{DID} &= \frac{1}{N_1} \sum_{i=1}^{N_1} (Y_{1it_1} - Y_{0it_0}) - \frac{1}{N_0} \sum_{j=1}^{N_0} (Y_{0jt_1} - Y_{0jt_0}) \\ &= \left[\left(\bar{Y}_{1t_1} \right)_1 - \left(\bar{Y}_{0t_0} \right)_1 \right] - \left[\left(\bar{Y}_{0t_1} \right)_0 - \left(\bar{Y}_{0t_0} \right)_0 \right]\end{aligned}$$

Identifizierungsbedingung:

$$E(Y_{0t_1} - Y_{0t_0} | D = 1) = E(Y_{0t_1} - Y_{0t_0} | D = 0)$$

ist mit (Variante von) „selection on the unobservables“ vereinbar

$$E(c_i | D = 1) \neq E(c_i | D = 0)$$

'Panelnormalfall'

id	t	Y	D	ΔD
11	t_0	Y_{0t_0}	0	
11	t_1	Y_{1t_1}	1	1
12	t_0	Y_{1t_0}	1	
12	t_1	Y_{1t_1}	1	0
13	t_0	Y_{0t_0}	0	
13	t_1	Y_{0t_1}	0	0
14	t_1	Y_{1t_0}	1	
14	t_1	Y_{0t_1}	0	-1

Falls $T > 2$:

“multiplicity of contrasts
that are sometimes
available” H,L&S

Sequential Treatments



Rubin-Modell und FE Panel-Modell

$$Y_{1it_1} = \mu_{1t_1}(X) + U_{1it_1}$$

$$Y_{0it_1} = \mu_{0t_1}(X) + U_{0it_1}$$

$$Y_{0it_1} = \beta_0^0 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

$$Y_{1it_1} = \beta_0^1 + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

The treatment dummy can be systematically related to the persistent component of the error term. This makes FE particularly suitable for applications where participation in a program is determined by preprogram attributes that also affect Y_{it} :

Wooldridge (2002, S.278)

Wozu Paneldaten?

Josef Brüderl:

1. zur Schätzung kausaler Effekte trotz unbeobachteter Heterogenität (Marriage-Premium for Men?)
2. zur Analyse individueller Dynamik
3. zur präziseren Schätzung (mehr Beobachtungen)



Beispiel zu 2): Persistenz im Export

Problem: Wird in jeder Periode neu über die Höhe der Produktivität entschieden?

Durch spezielles Know-How (z.B. durch Patente geschützt) könnte es zu Persistenzen in der Produktivität kommen, d.h. die Produktivität in einer Periode ist auch abhängig vom Wert in der Vorperiode.

$$y_{it} = \alpha y_{i,t-1} + c_i + u_{it}$$

$$\begin{aligned} \text{Cor}(y_{it}, y_{i,t-1}) &= \text{Cor}(\alpha y_{i,t-1} + c_i + u_{it}, y_{i,t-1}) \\ &= \alpha + \text{Cor}(c_i, y_{i,t-1}) \\ &= \alpha + \frac{(1-\alpha)}{[1 + (1-\alpha)\sigma_u^2] / [(1+\alpha)\sigma_c^2]} \end{aligned}$$

Bsp: y_{it} ist Studienfleiß. Lohnt es sich, Schüler zu bekehren (weil es positive State Dependence gibt ($\alpha > 0$)) oder ist Fleiß „angeboren“ (hohes c_i)

Dynamisches Modell:

$$y_{it} = \alpha y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it}$$

Durch die verzögerte abhängige Variable können die Standardmodelle nicht angewendet werden.

Bsp. FE:

$$\ddot{y}_{i,t-1} = \boxed{y_{i,t-1}} - \frac{1}{T-1} (y_{i1} + \dots + \boxed{y_{it}} + \dots + y_{i,T-1})$$

$$\ddot{u}_{it} = \boxed{u_{it}} - \frac{1}{T-1} (u_{i2} + \dots + \boxed{u_{i,t-1}} + \dots + u_{iT})$$

Lösung: FD-Transformation, IV-Ansatz

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta u_{it} \quad \text{für } t = 3, \dots, T$$

Mögliche Instrumente für $\Delta y_{i,t-1}$:

- Um eine weitere Periode verzögerte erste Differenz $\Delta y_{i,t-2}$, „verbraucht“ jedoch weitere Periode
- Niveau aus $t-2$: $y_{i,t-2}$, ist mit $\Delta y_{i,t-1}$ korreliert, jedoch nicht mit Δu_{it} (wenn keine Autokorrelation vorliegt)

Dynamisches Panelmodell

Lösung: FD-Transformation, IV-Ansatz

$$\Delta y_{it} = \alpha \Delta y_{i,t-1} + \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta u_{it} \quad \text{für } t = 3, \dots, T$$

Mögliche Instrumente für :

- Niveau aus $t-2$: y_{it-2} , ist mit Δy_{it-1} korreliert, jedoch nicht mit Δu_{it} (wenn keine Autokorrelation vorliegt).

- Wenn $T > 2$ gibt es viele solcher Instrumentvariablen

In Periode 3: y_{i1} ist ein Instrumente für Δy_{i3}

In Periode 4: y_{i1}, y_{i2} sind Instrumente für Δy_{i3}

In Periode 5: y_{i1}, y_{i2}, y_{i3} sind Instrumente für Δy_{i3}

In Periode 6: $y_{i1}, y_{i2}, y_{i3}, y_{i4}$ sind Instrumente für Δy_{i3}

Verbesserung: Verwendung aller möglicher Lags

$$\begin{aligned}
 Z_i &= \begin{bmatrix} y_{i1} & 0 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} \end{bmatrix} & \begin{matrix} t \\ 3 \\ 4 \\ \vdots \\ T \end{matrix} \\
 &= \begin{bmatrix} y_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & [y_{i1}, y_{i2}] & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & [y_{i1}, \dots, y_{i,T-2}] \end{bmatrix}
 \end{aligned}$$

Momente aus den Instrumenten (GMM):

$$E(\mathbf{Z}_i' \Delta \mathbf{u}_i) = \mathbf{0}$$

- Mehrgleichungssystem, i.A. keine eindeutige Lösung

Zu minimierende Kriteriumsfunktion:

$$J_N = \left(\frac{1}{N} \sum_{i=1}^N \Delta \mathbf{u}_i' \mathbf{Z}_i \right) \mathbf{W}_N \left(\frac{1}{N} \sum_{i=1}^N \mathbf{Z}_i' \Delta \mathbf{u}_i \right)$$

Two-Step Gewichtungsmatrix

$$\mathbf{W}_N = \left[\frac{1}{N} \sum_{i=1}^N \left(\mathbf{Z}_i' \hat{\Delta \mathbf{u}}_i \hat{\Delta \mathbf{u}}_i' \mathbf{Z}_i \right) \right]^{-1}$$

mit konsistenten Schätzern $\hat{\Delta \mathbf{u}}_i$ aus einer Schätzung im ersten Schritt.



2-stufige GMM Schätzung

```
xtabond2 lnep L.lnap exp lntP arbeiterant lohn hgd2 hgd3 hgd4 ost techd1 techd2 mbu mlu  
jahr1997-jahr2005, gmm(L.lnap) iv(exp lntP arbeiterant lohn hgd2 hgd3 hgd4 ost techd1  
techd2 mbu mlu jahr1997-jahr2005)nolevel nocons nomata twostep
```

Dynamic panel-data estimation, two-step difference GMM

```
-----  
Group variable: bnr                      Number of obs      =    356008  
Time variable : jahr                    Number of groups   =    55415  
Number of instruments = 66              Obs per group: min =         1  
Wald chi2(21) =    3611.55                avg =         6.42  
Prob > chi2    =         0.000              max =         9  
-----
```

lnep	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
lnep					
L1.	.3252009	.0143708	22.63	0.000	.2970346 .3533672
exp	.0591962	.003998	14.81	0.000	.0513603 .0670321
lntP	-.1757293	.0141978	-12.38	0.000	-.2035564 -.1479022
arbeiterant	.3159173	.0283449	11.15	0.000	.2603624 .3714722
lohn	.0001044	.0000166	6.28	0.000	.0000718 .000137
hgd2	.0217152	.0068951	3.15	0.002	.008201 .0352294
hgd3	-.0468468	.0166921	-2.81	0.005	-.0795628 -.0141308
hgd4	-.0631432	.0092034	-6.86	0.000	-.0811815 -.045105
ost	-.0576256	.1522563	-0.38	0.705	-.3560423 .2407912
techd1	-.0204412	.0161387	-1.27	0.205	-.0520726 .0111901
techd2	-.0027803	.0072628	-0.38	0.702	-.0170151 .0114546
:					

```
-----  
Arellano-Bond test for AR(1) in first differences: z = -27.63 Pr > z = 0.000  
Arellano-Bond test for AR(2) in first differences: z = 0.57 Pr > z = 0.567  
-----
```

- Fixed-Effects Logit

with
$$P(y_{i1} = 1 | c_i, \beta) = \frac{\exp(c_i + \mathbf{x}'_{i2}\beta)}{1 + \exp(c_i + \mathbf{x}'_{i2}\beta)},$$

it follows that the conditional probability is given by

$$P((0, 1) | t_i = 1, c_i, \beta) = \frac{\exp((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\beta)}{1 + \exp((\mathbf{x}_{i2} - \mathbf{x}_{i1})'\beta)},$$

- Random Effects Probit

$$P(\text{Mieter}_{h,t}) = \Phi(\beta_0 + \beta_1 \rho_{p,r} + \beta_2 T_{h,t} + \beta_3 \text{Mieter}_{h,t-1} + \mathbf{x}_{h,t}^\top \boldsymbol{\delta} + \mathbf{z}_{h,t}^\top \boldsymbol{\theta} + c_h)$$

► c_h : unbeobachtete Heterogenität

► $c_h | (\text{Mieter}_{h0}, \boldsymbol{\rho}_h, \mathbf{T}_h, \mathbf{x}_h) \sim N(\alpha_0 + \alpha_1 \text{Mieter}_{h0} + \boldsymbol{\alpha}_2^\top \boldsymbol{\rho}_h + \boldsymbol{\alpha}_3^\top \mathbf{T}_h + \boldsymbol{\alpha}_4^\top \mathbf{x}_h, \sigma_a^2)$