

Empirische Wirtschaftsforschung

Panelschätzungen

Ulrich Fritsche

Universität Hamburg

Email: `ulrich.fritsche@wiso.uni-hamburg.de`

Inhalt

Einführung

Pooled OLS

SUR

Fixed Effects

Random Effects

Vergleich Fixed und Random Effects

Zusammenfassung

(1) Einführung

- ▶ Panelschätzungen als Kombination von Querschnitts- und Zeitreihendaten.
- ▶ Anwendung: Betrachtung und Analyse derselben Gruppe von Individuen, Firmen, Ländern,... im Zeitverlauf.

Beispiel: Grunfelds' Investitionsdaten¹

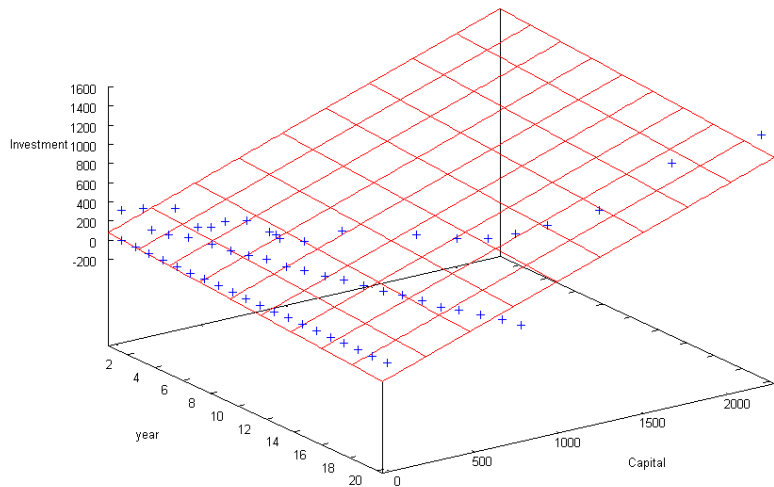
Investitionsgüternachfrage INV_{it} von 10 Firmen über 20 Jahre (1935-1954), in Abhängigkeit vom erwarteten Gewinn V_{it} und dem optimalen Kapitalstock K_{it} .²

$$INV_{it} = f(V_{it}, K_{it}) \quad (1)$$

¹ Grunfeld/Griliches (1960): "Is Aggregation Necessarily Bad", *Review of Economics and Statistics*

² in Millionen US-Dollar

Beispiel: Graphische Darstellung



firm i	year t	INV	V	K
1	1	317.6	3078.5	2.8
1	2	391.8	4661.7	52.6
1	3	410.6	5387.1	156.9
1	4	257.7	2792.2	209.2
2	1	209.9	1362.4	53.8
2	2	355.3	1807.1	50.5
2	3	469.9	2676.3	118.1
2	4	262.3	1801.9	260.2
3	1	33.1	1170.6	97.8
3	2	45	2015.8	104.4
3	3	77.2	2803.3	118
3	4	44.6	2039.7	156.2

Formale Darstellung: Wdh. OLS

$$y_t = \beta_1 + \beta_2 x_{t2} + \dots + \beta_m x_{tm} + u_t; t = 1, 2, \dots, T \quad (2)$$

$$\mathbf{y}_{T \times 1} = \mathbf{X}_{T \times m} \boldsymbol{\beta}_{m \times 1} + \mathbf{u}_{T \times 1} \quad (3)$$

$$\mathbf{y}_{T \times 1} = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{pmatrix} \quad \mathbf{X}_{T \times m} = \begin{pmatrix} 1 & X_{12} & \cdots & X_{1m} \\ 1 & X_{22} & \cdots & X_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{T2} & \cdots & X_{Tm} \end{pmatrix} \quad \boldsymbol{\beta}_{m \times 1} =$$
$$\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_m \end{pmatrix} \quad \mathbf{u}_{T \times 1} = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{pmatrix}$$

Panelmodell

$$y_{it} = \beta_{1it} + \beta_{2it}x_{2it} + \dots + \beta_{mit}x_{mit} + u_{it};$$
$$t = 1, 2, \dots, T; i = 1, 2, \dots, N \quad (4)$$

$$\mathbf{y}_{NT \times 1} = \mathbf{X}_{NT \times m} \boldsymbol{\beta}_{NT \times m} + \mathbf{u}_{NT \times 1} \quad (5)$$

$$\mathbf{y}_{NT \times 1} = \begin{pmatrix} y_{11} \\ y_{22} \\ \vdots \\ y_{NT} \end{pmatrix} \quad \mathbf{X}_{NT \times m} = \begin{pmatrix} 1 & X_{1,12} & \cdots & X_{1,1m} \\ 1 & X_{2,22} & \cdots & X_{2,2m} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{N,T2} & \cdots & X_{N,Tm} \end{pmatrix}$$

$$\boldsymbol{\beta}_{N, Tm \times 1} = \begin{pmatrix} \beta_{1,11} \\ \beta_{2,22} \\ \vdots \\ \beta_{N,Tm} \end{pmatrix} \quad \mathbf{u}_{NT \times 1} = \begin{pmatrix} u_{11} \\ u_{22} \\ \vdots \\ u_{NT} \end{pmatrix}$$

Vorteile von Panelschätzungen

Inhaltliche Vorteile

- ▶ Detailliertere Analyse
Beispiel: Konsumfunktion über 100 Individuen:
Einkommenselastizität von 2% im Durchschnitt. Aber: Alle Konsumenten? Nur die Hälfte?
- ▶ Identifikation der Vergangenheit einzelner Individuen, Länder,
...

⇒ Bessere Politikempfehlungen

Statistische Vorteile

- ▶ Höhere Effizienz der Schätzung durch große Anzahl von Beobachtungen
- ▶ Einbeziehung unbeobachtbarer Parameter

Charakteristika von Panelmodellen

- ▶ "balanced" und "unbalanced" Panel
- ▶ Verhältnis von Zeitdimension T und Querschnittsdimension N
 - ▶ $T > N$: Investitionsbeispiel, Makrodaten
 - ▶ $T < N$: Mikrostudien, z.B. das Sozioökonomische Panel des DIW
- ▶ statische und dynamische Panel

Schätzung von Panelmodellen

Problem: Die Gleichung

$$y_{it} = \beta_{1it} + \beta_{2it}x_{2it} + \beta_{3it}x_{3it} + u_{it}; \quad (6)$$
$$t = 1, 2, \dots, T; i = 1, 2, \dots, N$$

ist nicht schätzbar, da wir mehr Parameter ($3 \times N \times T$) als Beobachtungen ($N \times T$) haben.

- ▶ Pooled OLS
- ▶ Seemingly Unrelated Regression (SUR)
- ▶ Fixed Effects
- ▶ Random Effects
- ▶ Arellano-Bond-Schätzer
- ▶ und viele mehr: Pooled Mean Group Estimator, ...

(2) Pooled OLS

Annahmen:

- ▶ Koeffizienten verändern sich nicht über die Zeit: $\beta_{mit} = \beta_{mi}$
- ▶ Koeffizienten sind gleich für alle Individuen: $\beta_{mi} = \beta_m$
- ▶ Die Varianz ist dieselbe für alle Individuen: $Var[u_{it}] = Var[u_t]$
- ▶ Es besteht keine Abhängigkeit zwischen den Individuen:
 $Cov[u_{it}, u_{jt}] = 0$
- ▶ Es besteht keine Autokorrelation: $Cov[u_{it}, u_{is}] = 0$

Damit wird Gleichung (6) zu

$$y_{it} = \beta_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (7)$$

Pooled OLS (cont'd)

Schätzung von Gleichung (7) mit OLS in Gretl ergibt:

Modell 4: Gepoolte KQ, mit 200 Beobachtungen

Mit 10 Querschnittseinheiten

Zeitreihenlänge = 20

Abhängige Variable: Inv

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	-43,0245	9,49790	-4,5299	0,0000
V	0,115374	0,00583018	19,7892	0,0000
K	0,231931	0,0254649	9,1079	0,0000
Mittel d. abh. Var.	145.9068	Stdabw. d. abh. Var.	216.8855	
Summe d. quad. Res.	1749128	Stdfehler d. Regress.	94.22749	
R^2	0.813144	Korrigiertes R^2	0.811247	
$F(2, 197)$	428.6435	P-Wert(F)	1.75e-72	

Pooled OLS (cont'd)

Test auf gleiche Konstanten für alle Individuen:

H_0 : Konstante ist gleich für alle Firmen

Test-Statistik: $F(9, 188) = 48.9915$

mit p-Wert $P(F(9,188) > 48.9915) = 0.000000$

Das heißt, die Verwendung des Pooled OLS-Schätzers ist zu restriktiv.

(3) Seemingly Unrelated Regression (SUR)

Aufgabe der Annahmen

- ▶ $\beta_{mi} = \beta_m$, d.h., die Koeffizienten können für jedes Individuum unterschiedlich sein.
- ▶ $Cov[u_{it}, u_{jt}] = 0$, d.h., die Störterme können untereinander korreliert sein.

Beispiel für zwei Firmen:

- ▶ General Electric ($i = GE$) und Westinghouse ($j = WE$)
- ▶ Die Störterme $u_{GE,t}$ und $u_{WE,t}$ fangen den Einfluss von Variablen auf, die auf beide Investitionsnachfragen gleich wirken.
 - ▶ Zinssätze, gesamtwirtschaftliche Nachfrage,...

SUR (cont'd): Schätzung

Schätze

$$\begin{aligned} INV_{GE,t} &= \beta_{1GE} + \beta_{2GE} V_{GE,t} + \beta_{3GE} K_{GE,t} + u_{GE,t} \\ INV_{WE,t} &= \beta_{1WE} + \beta_{2WE} V_{WE,t} + \beta_{3WE} K_{WE,t} + u_{WE,t} \quad (8) \end{aligned}$$

mit

$$\text{Cov}[u_{GE,t}, u_{WE,t}] = \sigma_{GE,WE} \neq 0$$

Damit wird die Varianz-Kovarianz-Matrix $\mathbf{V}(\mathbf{u})$ aus dem OLS-Modell

$$\mathbf{V}_{OLS}(\mathbf{u}) = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \sigma^2 \mathbf{I}_T$$

zu

$$\mathbf{V}_{SUR}(\mathbf{u}) = \begin{bmatrix} \sigma_{GE}^2 \mathbf{I} & \sigma_{GE,WE} \mathbf{I} \\ \sigma_{WE,GE} \mathbf{I} & \sigma_{WE}^2 \mathbf{I} \end{bmatrix} = \begin{bmatrix} \sigma_{GE}^2 & \sigma_{GE,WE} \\ \sigma_{WE,GE} & \sigma_{WE}^2 \end{bmatrix} \otimes \mathbf{I}_T$$

SUR (cont'd): Schätzung

Dieses Modell kann dann mit dem SUR-Schätzer als **Generalized Least Squares (GLS)** bzw. **Weighted Least Squares** in zwei Schritten geschätzt werden:

- ▶ Schätze die Varianz-Kovarianz-Matrix mit OLS:³

$$\hat{\sigma}_{ij}^2 = \frac{(\mathbf{y}_{GE} - \mathbf{X}_{GE}\hat{\beta}_{GE})'(\mathbf{y}_{WE} - \mathbf{X}_j\hat{\beta}_j)}{T} = \frac{\hat{\mathbf{u}}_i'\hat{\mathbf{u}}_j}{T} \quad (9)$$

- ▶ Nutze die geschätzte Varianz-Kovarianz-Matrix zur Berechnung des SUR-Schätzers:

$$\hat{\beta}_{SUR} = (\mathbf{X}'\hat{\mathbf{V}}_{SUR}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}_{SUR}^{-1}\mathbf{y} \quad (10)$$

³Wdh.: Im OLS-Fall hatten wir $\hat{\sigma}^2 = \frac{1}{T-m}\hat{\mathbf{u}}'\hat{\mathbf{u}}$.

SUR (cont'd): Vergleich mit Einzelschätzung

Gleichungssystem, SUR
Schätzer: Scheinbar unverbundene Regressionen

Gleichung 1: SUR, benutze die Beobachtungen 1935–1954 ($T = 20$)
Abhängige Variable: inv_ge

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	-27,7193	27,0328	-1,0254	0,3174
v_ge	0,0383102	0,0132901	2,8826	0,0092
k_ge	0,139036	0,0230356	6,0357	0,0000
R-Quadrat	0,692557	Korrigiertes R-Quadrat	0,656388	

Gleichung 1: KQ, benutze die Beobachtungen 1935–1954 ($T = 20$)
Abhängige Variable: inv_ge

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	-9,95631	31,3742	-0,3173	0,7548
v_ge	0,0265512	0,0155661	1,7057	0,1063
k_ge	0,151694	0,0257041	5,9015	0,0000
R^2	0,705307	Korrigiertes R^2	0,670637	

SUR (cont'd): Vergleich mit Einzelschätzung

Gleichung 2: SUR, benutze die Beobachtungen 1935–1954 ($T = 20$)
Abhängige Variable: inv_we

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	-1,25199	6,95635	-0,1800	0,8590
v_we	0,0576298	0,0134110	4,2972	0,0004
k_we	0,0639781	0,0489010	1,3083	0,2056
R-Quadrat	0,740401	Korrigiertes R-Quadrat	0,709860	

Gleichung 2: KQ, benutze die Beobachtungen 1935–1954 ($T = 20$)
Abhängige Variable: inv_we

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	-0,509390	8,01529	-0,0636	0,9501
v_we	0,0528941	0,0157065	3,3677	0,0037
k_we	0,0924065	0,0560990	1,6472	0,1179
R^2	0,744446	Korrigiertes R^2	0,714381	

SUR (cont'd): Vor- und Nachteile

Vorteil:

Kleinere Standardfehler im Vergleich zu Einzelschätzungen

Nachteil:

Ist die Querschnittsdimension N "groß", müssen viele Gleichungen geschätzt werden

Lösung: Fixed Effects

(4) Fixed Effects

Unterschiede in den Querschnittsdaten werden über die Konstante eingefangen, die Effekte der übrigen Regressoren sind über alle Individuen gleich.

$$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (11)$$

mit $u_{it} \sim N(0, \sigma^2)$. Keine Abhängigkeit der Residuen von der Zeit und den Individuen.

Für die Schätzung gibt es zwei Möglichkeiten:

1. Least Squares Dummy Variable (LSDV) Schätzer
2. Within - Groups Schätzer

Fixed Effects (cont'd): LSDV Schätzer

Idee: Verwende eine Niveaudummyvariable für jedes Individuum:

$$D_{1i} = \begin{cases} 1 : i = 1 \\ 0 : \text{andere} \end{cases} \quad D_{2i} = \begin{cases} 2 : i = 1 \\ 0 : \text{andere} \end{cases} \quad \dots$$

Damit erhalten wir für das Beispiel:⁴

$$INV_{it} = \beta_{11}D_{1i} + \beta_{12}D_{2i} + \dots + \beta_{1,10}D_{10i} + \beta_2V_{2it} + \beta_3K_{3it} + u_{it} \quad (12)$$

⁴Keine weitere Konstante, da sonst perfekte Multikollinearität!

LSDV Beispiel Schätzung

Modell 6: Gepoolte KQ, mit 200 Beobachtungen

Mit 10 Querschnittseinheiten

Zeitreihenlänge = 20

Abhängige Variable: Inv

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
D_1	-69,1435	49,6855	-1,3916	0,1657
D_2	100,862	24,9137	4,0485	0,0001
D_3	-235,119	24,4183	-9,6288	0,0000
D_4	-27,6350	14,0698	-1,9641	0,0510
D_5	-115,317	14,1620	-8,1427	0,0000
D_6	-23,0736	12,6612	-1,8224	0,0700
D_7	-66,6829	12,8376	-5,1943	0,0000
D_8	-57,3586	13,9856	-4,1013	0,0001
D_9	-87,2770	12,8851	-6,7735	0,0000
D_{10}	-6,54627	11,8199	-0,5538	0,5803
V	0,109771	0,0118549	9,2596	0,0000
K	0,310644	0,0173704	17,8835	0,0000
Mittel d. abh. Var.	145.9068	Stdabw. d. abh. Var.	216.8855	
Summe d. quad. Res.	522855.2	Stdfehler d. Regress.	52.73656	
R^2	0.944144	Korrigiertes R^2	0.940876	
$F(11, 188)$	288.8925	P-Wert(F)	2.1e-111	

Fixed Effects: Within - Groups - Schätzer

Problem des LSDV-Schätzers: Hohe Anzahl von zu schätzenden Parametern

⇒ Großer Rechenaufwand (Beispiel: 1000 Individuen in einem repräsentativen Sample über 20 Jahre)

Lösung: Within-Groups-Schätzer

Umformung von Gleichung (11):

$$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (11)$$

Berechne den Durchschnitt für jedes Individuum i über die Zeit T :

$$\begin{aligned} \frac{1}{T} \sum_{t=1}^T y_{it} &= \beta_{1i} + \beta_2 \frac{1}{T} \sum_{t=1}^T x_{2it} + \beta_3 \frac{1}{T} \sum_{t=1}^T x_{3it} + \frac{1}{T} \sum_{t=1}^T u_{it} \\ \bar{y}_i &= \beta_{1i} + \beta_2 \bar{x}_{2i} + \beta_3 \bar{x}_{3i} + \bar{u}_i \end{aligned} \quad (13)$$

Fixed Effects: Within - Groups - Schätzer

Subtrahiere Gleichung (13) von (11):

$$\begin{aligned}y_{it} - \bar{y}_i &= \beta_2(x_{2it} - \bar{x}_{2i}) + \beta_3(x_{3it} - \bar{x}_{3i}) + (u_{it} - \bar{u}_i) \\ \tilde{y}_{it} &= \beta_2 \tilde{x}_{2it} + \beta_3 \tilde{x}_{3it} + \tilde{u}_{it}\end{aligned}\quad (14)$$

Schätzung dieser Gleichung mit OLS ergibt dieselben Schätzer für β_2 und β_3 sowie die Residuen u_{it} wie in der LSDV-Schätzung.

within group: KQ, benutze die Beobachtungen 1–200
Abhängige Variable: invd

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
vd	0,109771	0,0115517	9,5026	0,0000
kd	0,310644	0,0169261	18,3530	0,0000
Mittel d. abh. Var.	-7,50e-07	Stdabw. d. abh. Var.	106,1833	
Summe d. quad. Res.	522855,1	Stdfehler d. Regress.	51,38757	
R^2	0,766968	Korrigiertes R^2	0,765791	
$F(2, 198)$	325,8342	P-Wert(F)	2,37e-63	

Within-Groups-Schätzer: Offene Fragen: 1

Schätzer für die Konstanten?

Diese können berechnet werden über

$$\beta_{1i} = \bar{y}_i - \hat{\beta}_2 \bar{x}_{2i} - \hat{\beta}_3 \bar{x}_{3i} \quad (15)$$

mit $\hat{\beta}_{2,3}$ als geschätzte Koeffizienten aus Gleichung (14). Dies ergibt dieselben Werte wie in der Beispielschätzung des LSDV-Schätzers.

Aus dem Beispiel erhalten wir für die erste Firma im Datensatz:

$$\begin{aligned} \hat{\beta}_1 &= I\bar{N}V_1 - \hat{\beta}_2 \bar{V}_1 - \hat{\beta}_3 \bar{K}_1 \\ &= 608.02 - 0.109771 \times 4333.845 - 0.310644 \times 648.4350 \\ &= -69.1435 \\ &= D_1 \text{ aus der LSDV-Schätzung} \end{aligned} \quad (16)$$

Within-Groups-Schätzer: Offene Fragen: 2

Die Standardfehler der Within-Groups- und der LSDV-Schätzung sind verschieden

	LSDV	Within Groups
s.e. β_2	0.0118549	0.0115517
s.e. β_3	0.0173704	0.0169261

Dies liegt an der unterschiedlichen Zahl der Freiheitsgrade bei der Schätzung der Varianz.

Within-Groups-Schätzer: Offene Fragen: 2 (cont'd)

Einzelgleichungsfall:

$$\hat{\sigma}^2 = \frac{1}{T - m} \times RSS \quad (17)$$

In der LSDV-Schätzung:

$$\hat{\sigma}_{LSDV}^2 = \frac{1}{NT - N - m} \times RSS = \frac{1}{200 - 10 - 2} \times RSS = \frac{1}{188} \times RSS \quad (18)$$

im Vergleich zur Within-Groups-Schätzung:

$$\hat{\sigma}_{WG}^2 = \frac{1}{NT - m} \times RSS = \frac{1}{200 - 2} \times RSS = \frac{1}{198} \times RSS \quad (19)$$

Das heißt, im 2.Fall werden die s.e. zu niedrig und damit die t-Werte zu hoch geschätzt.

Within-Groups-Schätzer: Offene Fragen: 2 (cont'd)

Dieses Problem lässt sich jedoch durch einen Korrekturfaktor beheben. Gretl verwendet diesen automatisch, d.h., eine Schätzung unter Verwendung des Within-Groups-Schätzers liefert folgendes Ergebnis:

Fixed Effects: Fixed-effects, mit 200 Beobachtungen
Mit 10 Querschnittseinheiten
Zeitreihenlänge = 20
Abhängige Variable: Inv

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	-58,7290	12,4463	-4,7186	0,0000
V	0,109771	0,0118549	9,2596	0,0000
K	0,310644	0,0173704	17,8835	0,0000
Mittel d. abh. Var.	145.9068	Stdabw. d. abh. Var.	216.8855	
Summe d. quad. Res.	522855.2	Stdfehler d. Regress.	52.73656	
R^2	0.944144	Korrigiertes R^2	0.940876	
$F(11, 188)$	288.8925	P-Wert(F)	2.1e-111	

Der Schätzer für die Konstante ist der Durchschnitt über die 10 Schätzer für jede einzelne Firma.

(5) Random Effects

- ▶ Nachteil des Fixed Effects - Schätzers:
Charakteristika, die für jedes Individuum über die Zeit gleich bleiben, können nicht geschätzt werden.
- ▶ Beispiel: Geschlecht: Dummy = 1 für "Frau", dann ist
$$\tilde{D}_{it} = D_{it} - \bar{D}_i = 0$$

⇒ Lösung: Random Effects

- ▶ Idee: Beobachtungen sind eine zufällige Stichprobe aus einem größeren Sample
⇒ Besonders plausibel bei Umfrage- und Mikrodaten.

Random Effects: Formale Darstellung

In Gleichung (11) des Fixed Effects - Modells,

$$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it}, \quad (11)$$

ersetze die Konstante β_{1i} durch

$$\beta_{1i} = \bar{\beta}_1 + e_i \quad (20)$$

mit $E[e_i] = 0$, $Var[e_i] = \sigma_e^2$, $Cov[e_i, e_j] = 0$
und $\bar{\beta}_1$: Durchschnitt aller Individuen

Das heißt, die individuellen Effekte wirken über den Störterm e_i und nicht mehr über die Konstante. e_i ändert sich je nachdem welches Individuum **zufällig** gezogen wird.

Random Effects: Formale Darstellung

Verwende Gleichung (20) in Gleichung (11):

$$\begin{aligned}y_{it} &= \bar{\beta}_1 + e_i + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \\ &= \bar{\beta}_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + (e_i + u_{it}) \\ &= \bar{\beta}_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + \nu_{it}\end{aligned}\tag{21}$$

Annahmen:

- ▶ Individuelle Effekte sind über die Zeit unkorreliert:
 $Cov(e_{it}, e_{is}) = 0$
- ▶ Keine Korrelation zwischen individuellen Effekten und dem Störterm der gesamten Schätzung: $Cov(e_i, u_{it}) = 0$

Schätzung von Random Effects

Problem: Autokorrelation in den Residuen ν_{it}

$$\begin{aligned} \text{Cov}(\nu_{it}, \nu_{is}) &= E(\nu_{it}\nu_{is}) \\ &= E[(e_i + u_{it})(e_i + u_{is})] \\ &= E(e_i^2) + E(e_i u_{is}) + E(u_{it} e_i) + E(u_{it} u_{is}) \\ &= \sigma_e^2 + 0 + 0 + 0 \\ &= \sigma_e^2 \end{aligned} \tag{22}$$

Illustration:

Individuum	Zeitpunkt	ν_{it}
1	1	$e_1 + u_{11}$
1	2	$e_1 + u_{12}$
2	1	$e_2 + u_{21}$
2	2	$e_2 + u_{22}$

Schätzung von Random Effects (cont't)

Unter Verwendung von Gleichung (22) erhält man den Autokorrelationskoeffizient ρ der Residuen ν_{it} als:

$$\rho = \text{corr}(\nu_{it}, \nu_{is}) = \frac{\text{Cov}(\nu_{it}, \nu_{is})}{\sqrt{\text{Var}(\nu_{it})\text{Var}(\nu_{is})}} = \frac{\sigma_e^2}{\sigma_e^2 + \sigma_u^2} \quad (23)$$

Das heißt, ρ geht **nicht** mit der Zeit zurück.

Erinnerung

Überhöhte geschätzte Standardfehler bei Vorliegen von Autokorrelation, OLS ist nicht mehr der **beste** Schätzer (BLUE)

⇒ Schätzung mit **Generalized Least Squares (GLS)**

Schätzung von Random Effects: GLS

Transformiere die Schätzgleichung des Random Effects Modells aus Gleichung (21):

$$y_{it} = \bar{\beta}_1 + \beta_2 x_{2it} + \beta_3 x_{3it} + \nu_{it} \quad (21)$$

Berechne die Abweichung von den Durchschnitten der Beobachtungen für i :

$$\begin{aligned} y_{it}^* &= \bar{\beta}_1 x_{1it}^* + \beta_2 x_{2it}^* + \beta_3 x_{3it}^* + \nu_{it}^* & (24) \\ (y_{it} - \theta \bar{y}_i) &= \bar{\beta}_1 (1 - \theta) + \beta_2 (x_{2it} - \theta \bar{x}_{2i}) \\ &+ \beta_3 (x_{3it} - \theta \bar{x}_{3i}) + (\nu_{it} - \theta \bar{\nu}_i) \end{aligned}$$

Der Transformationsparameter θ korrigiert für die Autokorrelation und ist definiert als:

$$\theta = 1 - \frac{\sigma_u}{\sqrt{T\sigma_e^2 + \sigma_u^2}} \quad (25)$$

R^2 wird von Gretl nicht berechnet.
(Siehe Verbeek (2007), S.369 zu R^2 in Panelmodellen.)

(6) Vergleich Fixed und Random Effects

Individuelle Heterogenität:

- ▶ **Fixed Effects:** Konstante, gleichbleibende Charakteristika nicht schätzbar
- ▶ **Random Effects:** Störterm e_i

Beispielschätzung: "National Longitudinal Survey (NLS)"

- ▶ **Daten**
 - ▶ 716 Individuen, 5 Jahre
- ▶ **Variablen**
 - ▶ abhängige Variable: $\ln(Wage)$
 - ▶ unabhängige Variablen: *EXPER*, Arbeitserfahrung; *TENURE*, Beschäftigungsdauer; *EDUC*, Ausbildungszeit vor Jobantritt in Jahren; *BLACK*, 1 wenn Schwarz; *SOUTH*, 1 wenn Wohnort im Süden; *UNION*, 1 wenn Gewerkschaftsmitglied

Vergleich Fixed und Random Effects (cont'd)

Fixed-effects, mit 3580 Beobachtungen

Mit 716 Querschnittseinheiten

Zeitreihenlänge = 5

Abhängige Variable: l wage

Weggelassen wg. exakter Kollinearität: BLACK EDUC

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	1,45003	0,0401400	36,1244	0,0000
exper	0,0410832	0,00662001	6,2059	0,0000
exper2	-0,000409052	0,000273333	-1,4965	0,1346
tenure	0,0139089	0,00327784	4,2433	0,0000
tenure2	-0,000896227	0,000205860	-4,3536	0,0000
south	-0,0163224	0,0361490	-0,4515	0,6516
union	0,0636972	0,0142538	4,4688	0,0000
Mittel d. abh. Var.	1,918238	Stdabw. d. abh. Var.	0,464607	
Summe d. quad. Res.	108,7985	Stdfehler d. Regress.	0,195110	
R^2	0,859172	Korrigiertes R^2	0,823644	

Test auf unterschiedliche Konstanten in Gruppen –

Nullhypothese: Die Gruppen haben gleiche Konstanten

Teststatistik: $F(715, 2858) = 15,145$

mit p-Wert = $P(F(715, 2858) > 15,145) = 0,0000$

Vergleich Fixed und Random Effects (cont'd)

Exakte Multikollinearität:

Aus Gleichung (11) der Schätzung des Fixed Effects Modells für Individuum 1:

$$y_{it} = \beta_{1i} + \beta_2 x_{2it} + \beta_3 x_{3it} + u_{it} \quad (26)$$

$$\ln(WAGE)_{1t} = \beta_{11} + \beta_2 BLACK + \beta_3 EDUC + \dots$$

$$\ln(WAGE)_{1t} = \beta_{11} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_3 \begin{pmatrix} 12 \\ 12 \\ \vdots \\ 12 \end{pmatrix} + \dots$$

Vergleich Fixed und Random Effects (cont'd)

Random-effects (GLS), mit 3580 Beobachtungen

Mit 716 Querschnittseinheiten

Zeitreihenlänge = 5

Abhängige Variable: l wage

	Koeffizient	Std. Fehler	t-Quotient	P-Wert
const	0,533929	0,0798828	6,6839	0,0000
exper	0,0436170	0,00635758	6,8606	0,0000
exper2	-0,000560959	0,000262608	-2,1361	0,0327
tenure	0,0141541	0,00316656	4,4699	0,0000
tenure2	-0,000755342	0,000194726	-3,8790	0,0001
south	-0,0818117	0,0224109	-3,6505	0,0003
black	-0,116737	0,0302087	-3,8643	0,0001
union	0,0802353	0,0132132	6,0724	0,0000
educ	0,0732536	0,00533076	13,7417	0,0000
Mittel d. abh. Var.	1,918238	Stdabw. d. abh. Var.	0,464607	
Summe d. quad. Res.	526,5595	Stdfehler d. Regress.	0,383944	
Log-Likelihood	-1648,811	Akaike-Kriterium	3315,621	
Schwarz-Kriterium	3371,269	Hannan-Quinn	3335,459	

$$\hat{\sigma}_\varepsilon^2 = 0,0380681$$

$$\hat{\sigma}_u^2 = 0,115887$$

$$\theta = 0,743683$$

Tests im Random Effects Modell: Breusch-Pagan

Breusch-Pagan-Test auf Vorliegen von Random Effects

- ▶ $H_0 : \sigma_e^2 = 0 \Rightarrow \nu_{it} = u_{it}$
- ▶ Teststatistik

$$\xi_{LM} = \frac{NT}{2(T-1)} \left\{ \frac{\sum_{i=1}^N \left(\sum_{t=1}^T \hat{u}_{it} \right)^2}{\sum_{i=1}^N \sum_{t=1}^T \hat{u}_{it}^2} - 1 \right\} \quad (27)$$

Beispiel

Breusch-Pagan –

Nullhypothese: Varianz des idiosynkratischen Fehlers = 0

Asymptotische Teststatistik: $\chi^2(1) = 3859,28$

mit p-Wert = 0,00000

Tests im Random Effects Modell: Hausmann

Hausmann-Test auf Exogenität der individuellen Effekte

- ▶ $H_0 : Cov(e_i, x_{it}) = 0$
- ▶ Fixed Effects Schätzer ist konsistent⁵ wenn H_0 abgelehnt wird, nicht jedoch der Random Effects Schätzer
- ▶ Teststatistik:

$$\xi_H = (\hat{\beta}_{FE} - \hat{\beta}_{RE})' [\hat{V}(\hat{\beta}_{FE}) - \hat{V}(\hat{\beta}_{RE})]^{-1} (\hat{\beta}_{FE} - \hat{\beta}_{RE}) \quad (28)$$

Beispiel

Hausman-Test –

Nullhypothese: GLS-Schätzungen sind konsistent

Asymptotische Teststatistik: $\chi^2(6) = 20,5231$

mit p-Wert = 0,00223382

⇒ Fixed Effects ist zu bevorzugen. Mehr dazu in der nächsten Vorlesung.

(7) Zusammenfassung

- ▶ Pooled OLS
- ▶ Seemingly Unrelated Regression
- ▶ Fixed und Random Effects
- ▶ Tests und Vergleich von Fixed und Random Effects
- ▶ **Literatur:**
 - Dougherty (2007), Kapitel 14 (Stine)
 - Lucke (2006), SUR (Stine)