UNIVERSITÄT
BIELEFELD
Fakultät für
Wirtschaftswissenschaften

Einführung Mikroökonometrie

310208 (VO)

Dietmar Bauer

Wiederholung

- Gepoolte Querschnittsdaten liegen vor, wenn wir mehrere gleichartige Datensätze zu verschiedenen Zeiten (oder in verschiedenen Populationen) erheben.
- Sind die gleichen Variablen enthalten, dann können wir das gleiche Modell schätzen.
- Das können wir entweder getrennt für jeden Datensatz oder gemeinsam im vereinten (=gepoolten) Datensatz machen.
- Getrennt ist besser, wenn die Parameter nicht übereinstimmen.
- Gemeinsam ist besser, wenn die Parameter gleich sind: Das schreit nach einem statistischen Test!

Wiederholung (II)

- Letztes Mal haben wir gesehen, dass wir die beiden getrennten Schätzungen auch in ein gemeinsames Modell 'einbetten' können.
- Das ist zunächst nur ein algebraischer Trick.

$$\begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix} = \begin{pmatrix} X_{(1)} & 0 \\ 0 & X_{(2)} \end{pmatrix} \begin{pmatrix} \beta_{(1)} \\ \beta_{(2)} \end{pmatrix} + \begin{pmatrix} U_{(1)} \\ U_{(2)} \end{pmatrix}$$
$$Y = X\alpha + U$$

Andere Parameterisierung

Betrachten wir zunächst den einfachsten Fall: nur ein Regressor:

- Stichprobe 1: $\mathbf{y}_i = \beta_0 + \beta_1 \mathbf{x}_i + \mathbf{u}_i$
- Stichprobe 2: $\mathbf{y}_i = \beta_2 + \beta_3 \mathbf{x}_i + \mathbf{u}_i$

Wir verwenden die Dummy-Variable $\mathbf{z}_i$, die anzeigt, ob ein Datenpunkt zu Stichprobe 2 gehört.

Dann können wir schreiben:

$$\mathbf{y}_i = \beta_0 + \beta_1 \mathbf{x}_i + (\beta_2 - \beta_0)\mathbf{z}_i + (\beta_3 - \beta_1)\mathbf{z}_i \mathbf{x}_i + \mathbf{u}_i$$

$$\begin{array}{l|c|r} \text{Stichprobe 1} & \mathbf{z}_i = 0 & \mathbf{y}_i = \beta_0 + \beta_1 \mathbf{x}_i + \mathbf{u}_i \\ \text{Stichprobe 2} & \mathbf{z}_i = 1 & \mathbf{y}_i = \beta_0 + \beta_1 \mathbf{x}_i + (\beta_2 - \beta_0) + (\beta_3 - \beta_1)\mathbf{x}_i + \mathbf{u}_i \end{array}$$

Andere Parametrisierung:

Betrachte die folgende Gleichung: $\mathbf{z}_i^{(j)} = \delta(j = 2)$ (Dummy für Periode 2)

$$\mathbf{y}_i^{(j)} = (\beta_{(1)})'\mathbf{X}_i^{(j)} + \gamma'(\mathbf{z}_i^{(j)}\mathbf{X}_i^{(j)}) + \mathbf{u}_i^{(j)}, \qquad (*)$$

für $i = 1, ..., n(j), j = 1, 2$. Damit für $j = 1$:

$$\mathbf{y}_i^{(1)} = (\beta_{(1)})'\mathbf{X}_i^{(1)} + \mathbf{u}_i^{(1)}$$

Für $j = 2$:

$$\mathbf{y}_i^{(2)} = (\beta_{(1)})'\mathbf{X}_i^{(2)} + \gamma'\mathbf{X}_i^{(2)} + \mathbf{u}_i^{(2)} = (\beta_{(1)} + \gamma)'\mathbf{X}_i^{(2)} + \mathbf{u}_i^{(2)}$$

und damit

$$\beta_{(1)} + \gamma = \beta_{(2)}$$

Die Koeffizienten sind gleich, dann und nur dann, wenn $\gamma = 0$.

Stellt man die Beobachtungen des zweiten Datensatzes unter den ersten, kann man beide Gleichungen so gemeinsam schätzen.

In Matrixschreibweise:

$$
Y = \left(\begin{array}{c} \mathbf{y}_1^{(1)} \\ \mathbf{y}_2^{(1)} \\ \vdots \\ \mathbf{y}_{n(1)}^{(1)} \\ \hline \mathbf{y}_1^{(2)} \\ \vdots \\ \mathbf{y}_{n(2)}^{(2)} \end{array} \right), \quad
X = \left(\begin{array}{cccc|cccc}
1 & \mathbf{x}_{1,1}^{(1)} & \dots & \mathbf{x}_{1,K}^{(1)} & 0 & 0 & \dots & 0 \\
1 & \mathbf{x}_{2,1}^{(1)} & \dots & \mathbf{x}_{2,K}^{(1)} & 0 & 0 & \dots & 0 \\
\vdots & \vdots & & \vdots & \vdots & & & \vdots \\
1 & \mathbf{x}_{n(1),1}^{(1)} & \dots & \mathbf{x}_{n(1),K}^{(1)} & 0 & 0 & \dots & 0 \\
\hline
1 & \mathbf{x}_{1,1}^{(2)} & \dots & \mathbf{x}_{1,K}^{(2)} & 1 & \mathbf{x}_{1,1}^{(2)} & \dots & \mathbf{x}_{1,K}^{(2)} \\
\vdots & & & \vdots & \vdots & \vdots & & \vdots \\
1 & \mathbf{x}_{n(2),1}^{(2)} & \dots & \mathbf{x}_{n(2),K}^{(2)} & 1 & \mathbf{x}_{n(2),1}^{(2)} & \dots & \mathbf{x}_{n(2),K}^{(2)}
\end{array} \right)
$$

$$
Y = \left(\begin{array}{c} Y_{(1)} \\ Y_{(2)} \end{array} \right), \qquad X = \left(\begin{array}{c|c} X_{(1)} & 0 \\ \hline X_{(2)} & X_{(2)} \end{array} \right), \qquad \alpha = \left(\begin{array}{c} \beta^{(1)} \\ \hline \gamma \end{array} \right)
$$

Das ergibt den KQ-Schätzer:

$$
\begin{aligned}
\hat{\alpha} &= (X'X)^{-1}X'Y \\
&= \begin{pmatrix} X'_{(1)}X_{(1)} + X'_{(2)}X_{(2)} & X'_{(2)}X_{(2)} \\ X'_{(2)}X_{(2)} & X'_{(2)}X_{(2)} \end{pmatrix}^{-1} \begin{pmatrix} X'_{(1)}Y_{(1)} + X'_{(2)}Y_{(2)} \\ X'_{(2)}Y_{(2)} \end{pmatrix} \\
&= \left(\begin{bmatrix} 0 & 0 \\ 0 & (X'_{(2)}X_{(2)})^{-1} \end{bmatrix} + \begin{bmatrix} I_{K+1} \\ -I_{K+1} \end{bmatrix} (X'_{(1)}X_{(1)})^{-1} \begin{bmatrix} I_{K+1} & -I_{K+1} \end{bmatrix} \right) \begin{bmatrix} X'_{(1)}Y_{(1)} + X'_{(2)}Y_{(2)} \\ X'_{(2)}Y_{(2)} \end{bmatrix} \\
&= \begin{pmatrix} 0 \\ \hat{\beta}_{(2)} \end{pmatrix} + \begin{bmatrix} I_{K+1} \\ -I_{K+1} \end{bmatrix} \hat{\beta}_{(1)} = \begin{pmatrix} \hat{\beta}_{(1)} \\ \hat{\beta}_{(2)} - \hat{\beta}_{(1)} \end{pmatrix}
\end{aligned}
$$

Daher ist der KQ-Schätzer für γ gleich der Differenz der einzeln geschätzten Gleichungen.

Nachdem wir beide Datensätze in einer Gleichung haben, können wir die KQ-Theorie verwenden, um folgendes Theorem zu zeigen:

Theorem (Chow-Test)

Wenn das gemeinsame Modell $(*)$ die Annahmen KLR erfüllt und somit $\beta^{(1)} = \beta^{(2)}$ gilt, dann ist die Test-Statistik

$$F = (\hat{\boldsymbol{\beta}}_{(2)} - \hat{\boldsymbol{\beta}}_{(1)})'(M_1^{-1} + M_2^{-1})^{-1}(\hat{\boldsymbol{\beta}}_{(2)} - \hat{\boldsymbol{\beta}}_{(1)})/(K\hat{\sigma}^2)$$

mit $M_j = X'_{(j)}X_{(j)}$ verteilt nach einer F-Verteilung mit K und $n(1) + n(2) - 2K$ Freiheitsgraden.

Da die beiden Datensätze unabhängig voneinander sind, sind auch die beiden Schätzer $\hat{\boldsymbol{\beta}}_{(1)}$ und $\hat{\boldsymbol{\beta}}_{(2)}$ voneinander unabhängig. Daher gilt:

$$\mathrm{Var}(\hat{\boldsymbol{\beta}}_{(2)} - \hat{\boldsymbol{\beta}}_{(1)}) = \mathrm{Var}(\hat{\boldsymbol{\beta}}_{(1)}) + \mathrm{Var}(\hat{\boldsymbol{\beta}}_{(2)}) = \sigma^2 M_1^{-1} + \sigma^2 M_2^{-1} = \sigma^2(M_1^{-1} + M_2^{-1})$$

Chow-Test:

- Nullhypothese: $H_0 : \gamma = 0$.
- Gegenhypothese: $H_1 : \gamma \neq 0$.
- Test-Statistik: F verteilt nach F mit K und $n(1) + n(2) - 2K$ Freiheitsgraden.
- Test mit Konfidenz $1 - \alpha$ verwirft, wenn p-Wert kleiner als α ist.

3. Gepoolte Querschnittsdaten

Man kann zeigen, dass

$$F = \frac{(SSR_p - SSR_1 - SSR_2)}{SSR_1 + SSR_2} \frac{n(1) + n(2) - 2K}{K}$$

wobei SSR_1 und SSR_2 die Residuenquadratsumme der gemeinsamen Schätzung sind und SSR_p jene, die unter der Nullhypothese, also in einem Modell mit $\gamma = 0$ erhalten würde.

Interpretation:

- Darstellung von F mit Differenz der Schätzer zeigt, dass der Test verwirft, wenn die geschätzten Koeffizienten zu unterschiedlich sind.
- Darstellung mit Residuenquadratsumme zeigt, dass der Test dann verwirft, wenn die Residuenquadratsumme stark verringert wird, wenn man in den beiden Datensätzen unterschiedliche Koeffizienten zulässt.

Beweis der Darstellung: Der Beweis benutzt die Blockmatrizeninversion:

$$(X'X)^{-1} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} (\underbrace{D - CA^{-1}B}_{D_\perp})^{-1} \begin{bmatrix} -CA^{-1} & I \end{bmatrix}$$

wobei $A = M_1 + M_2$ und $B = C = D = M_2$. Dann gilt $\hat{\gamma} = [0, I](X'X)^{-1}X'Y$ und somit

$$\begin{aligned}
\hat{\gamma}'(M_1^{-1} + M_2^{-1})^{-1}\hat{\gamma} &= Y'X(X'X)^{-1} \begin{bmatrix} 0 \\ I \end{bmatrix} (D_\perp^{-1})^{-1}[0, I](X'X)^{-1}X'Y \\
&= Y'X \left(\begin{bmatrix} -A^{-1}B \\ I \end{bmatrix} D_\perp^{-1} D_\perp D_\perp^{-1} \begin{bmatrix} -CA^{-1} & I \end{bmatrix} \right) X'Y \\
&= Y'X \left((X'X)^{-1} - \begin{bmatrix} I \\ 0 \end{bmatrix} (M_1 + M_2)^{-1}[I, 0] \right) X'Y \\
&= Y' \left(X(X'X)^{-1}X' - I + I - X_1(M_1 + M_2)^{-1}X_1' \right) Y \\
&= Y'(I - X_1(X_1'X_1)^{-1}X_1)Y - Y'(I - X(X'X)^{-1}X')Y
\end{aligned}$$

Diese Ausdrücke kennen wir als Residuenquadratsumme der Regression von Y auf X_1, die ersten Spalten der Datenmatrix X (das ergibt SSR_p) beziehungsweise auf X (das ist $SSR_1 + SSR_2$).

3. Gepoolte Querschnittsdaten

Datensatz *CPS*78_85.*RAW* in Wooldridge: Durchschnittseinkommen als Funktion der Ausbildung.

- Wir haben schon gesehen, dass die Ausbildung und die Erfahrung und einige andere Variable einen Einfluss haben.
- Wir haben auch gesehen, dass Logarithmen zu Homoskedastizität führen.
- Jetzt werden zwei Wellen der Erhebung in 78 und 85 gemeinsam verwendet.
- Die Inflation ist nicht berücksichtigt, sodass schon allein deswegen damit gerechnet werden muss, dass die beiden Datensätze nicht mit dem gleichen Modell beschrieben werden können: Die Inflation kann als multiplikativer Faktor gesehen werden.

3. Gepoolte Querschnittsdaten

```
Call:
lm(formula = lwage ~ educ + south + female + exper + expersq +
union)

Residuals:
Min        1Q   Median        3Q       Max
-2.69617 -0.27930 -0.01868  0.29430  2.31078

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.48834432  0.08518813   5.733 0.000  ***
educ         0.08692861  0.00560766  15.502 < 0.00 ***
south       -0.06856410  0.03089558  -2.219 0.0267 *
female      -0.25227665  0.02837287  -8.891 < 0.00 ***
exper        0.03210662  0.00392788   8.174 0.000  ***
expersq     -0.00045035  0.00008542  -5.272 0.000  ***
union        0.13216848  0.03341396   3.955 0.000  ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4551 on 1077 degrees of freedom
Multiple R-squared: 0.3009,Adjusted R-squared: 0.297
F-statistic: 77.24 on 6 and 1077 DF,  p-value: < 2.2e-16
```

3. Gepoolte Querschnittsdaten

```
Call:
lm(formula = lwage ~ educ + south + female + exper + expersq +
union, subset = (y85 == 0))

Residuals:
Min       1Q   Median       3Q      Max
-2.55578 -0.23808  0.01535  0.21889  1.31010

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.4677126  0.0983131   4.757 2.52e-06 ***
educ         0.0765125  0.0065206  11.734  < 2e-16 ***
south       -0.0157631  0.0370184  -0.426   0.6704
female      -0.3156381  0.0344931  -9.151  < 2e-16 ***
exper        0.0249556  0.0047169   5.291 1.77e-07 ***
expersq     -0.0002857  0.0001016  -2.812   0.0051 **
union        0.2008916  0.0375991   5.343 1.35e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.387 on 543 degrees of freedom
Multiple R-squared: 0.3835,Adjusted R-squared: 0.3766
F-statistic: 56.29 on 6 and 543 DF,  p-value: < 2.2e-16
```

3. Gepoolte Querschnittsdaten

```
Call:
lm(formula = lwage ~ educ + south + female + exper + expersq +
union, subset = (y85 == 1))

Residuals:
Min       1Q   Median       3Q      Max
-2.22617 -0.27678 -0.00632  0.28205  2.10718

Coefficients:
Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.6560242  0.1214769   5.400 1.01e-07 ***
educ         0.0880661  0.0079352  11.098  < 2e-16 ***
south       -0.1024121  0.0422090  -2.426 0.015588 *
female      -0.2337426  0.0385604  -6.062 2.57e-09 ***
exper        0.0336233  0.0053604   6.273 7.41e-10 ***
expersq     -0.0005097  0.0001177  -4.329 1.79e-05 ***
union        0.1916060  0.0504258   3.800 0.000162 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.436 on 527 degrees of freedom
Multiple R-squared: 0.3251,Adjusted R-squared: 0.3174
F-statistic:  42.3 on 6 and 527 DF,  p-value: < 2.2e-16
```

3. Gepoolte Querschnittsdaten

Die Zahlen ändern sich, aber ist das genug, um auf andere Koeffizienten schließen zu können?

Chow-Test:

SSR_p	SSR_1	SSR_2	n	K	F	F_c	p
223.09	81.32	100.19	1084	6	35.02	2.02	0

p-Wert kleiner als 0.05 $\Rightarrow$ Test verwirft Gleichheit der Koeffizienten ganz klar.

Das sagt uns aber noch nicht, warum der Test verwirft.

3. Gepoolte Querschnittsdaten

```
Call: lm(formula = lwage ~ y85 + educ + I(south * y85) + I(y85 * educ) +
exper + expersq + union + female + I(y85 * female))

Residuals:
Min        1Q   Median       3Q      Max
-2.55656 -0.25804  0.00597  0.26069  2.08614

Coefficients:
Estimate  Std. Error t value             Pr(>|t|)
(Intercept)      0.46595135  0.09323821   4.997 0.00000677924112228 ***
y85              0.18037940  0.12577523   1.434             0.15182
educ             0.07458564  0.00665882  11.201 < 0.0000000000000002 ***
I(south * y85)  -0.10318974  0.03968689  -2.600             0.00945 **
I(y85 * educ)    0.01599902  0.00937712   1.706             0.08826 .
exper            0.02917736  0.00356123   8.193 0.00000000000000719 ***
expersq         -0.00039218  0.00007738  -5.068 0.00000047307554696 ***
union            0.19796150  0.03025619   6.543 0.00000000009324648 ***
female          -0.31740547  0.03652471  -8.690 < 0.0000000000000002 ***
I(y85 * female)  0.08345961  0.05117571   1.631             0.10322
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4116 on 1074 degrees of freedom
Multiple R-squared: 0.4298,Adjusted R-squared: 0.425
F-statistic: 89.94 on 9 and 1074 DF,  p-value: < 0.0000000000000022
```

3. 1. Der FD-Schätzer

Gepoolte Querschnittsdaten tauchen oft in der Politikanalyse auf: erklärt anhand der Verbrechensrate.

- Angenommen Sie erkennen ein Problem mit Kriminalität in einer Stadt in der Form einer hohen Kriminalitätsrate.
- Sie setzen eine Maßnahme zur Behebung des Problems.
- Nach einer gewissen Zeit monitoren Sie die Kriminalitätsrate.
- Können Sie anhand der Entwicklung der Kriminalitätsrate erkennen, ob Sie das Problem behoben haben?
- Nein, da sich auch andere Dinge parallel verändert haben könnten. Zum Beispiel die Arbeitslosenquote. Oder aber es könnte ein genereller bundesweiter Trend sein. Oder ...

Wie können Sie untersuchen, ob wirklich Ihre Maßnahme den Ausschlag gegeben hat?

3. 1. Der FD-Schätzer

Zunächst brauchen Sie ein Modell, dass die wesentlichen Einflussfaktoren beinhaltet:

Daten von 46 US Städten in 2 Jahren (=92 Beobachtungen insgesamt; 1982 und 1987)

- `crmrte`: jährliche Zahl von 'indizierten' Verbrechen pro 1000 Einwohner
- `unem:` Arbeitslosenquote: in Prozent (Mittelwert etwa 8).
- `lpolpc:` log policemen per 1000 cap.
- `lpop:` log population.
- `llawexpc:` log law enforcement expenditure per capita

3. 1. Der FD-Schätzer

```
Call:
lm(formula = crmrte ~ unem + lpolpc + lpop + llawexpc)

Residuals:
   Min     1Q Median     3Q    Max
-42.44 -21.28  -4.55  12.12  80.83

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  77.8434    82.7924   0.940  0.34971
unem         -0.7574     0.9304  -0.814  0.41784
lpolpc       41.4583    12.2506   3.384  0.00107 **
lpop         -4.8734     4.3008  -1.133  0.26027
llawexpc      8.6031    10.9378   0.787  0.43368
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 27.7 on 87 degrees of freedom
Multiple R-squared: 0.1762,Adjusted R-squared: 0.1383
F-statistic: 4.653 on 4 and 87 DF,  p-value: 0.001886
```

Oder doch nicht? Es fehlen sicher noch viele Einflussfaktoren.

3. 1. Der FD-Schätzer

Je nach Stadt gibt es viele Einflussfaktoren, nicht alle sind wahrscheinlich gut zu messen. Auf jeden Fall wäre das ein riesiges Modell.

Alternative:

$$\mathbf{y}_{i,t} = \beta_0 + \beta_1 \delta_{t=87} + \beta' \mathbf{x}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}$$

Hier verwenden wir zwei Subskripts, um explizit zu machen, dass wir Beobachtungen von den gleichen Individuen zu verschiedenen Zeitpunkten haben.

- Konstante plus eine potentiell andere Konstante im zweiten Jahr 1987 (durch δ_{87})
- Regressoren, die in jedem Jahr gleich wirken: $\mathbf{x}_{i,t}$.
- individuenspezifische Konstanten $\mathbf{a}_i$, die sich über die Zeit hinweg nicht ändern, aber von Stadt zu Stadt andere Werte haben (können)
- Fehlerterme $\mathbf{u}_{i,t}$

Anmerkung: Die Individuen sind im konkreten Fall die Städte, wir sprechen abwechselnd von Individuen oder von Städten.

3. 1. Der FD-Schätzer

Welchen Charakter haben die $\mathbf{a}_i$'s?

- fettgedruckt, also Zufallsvariablen
- wenn wir nur Individuum *i* betrachten: selbe Eigenschaften wie die Koeffizienten β: individuenspezifische Konstante.
- Wir brauchen wiederholte Beobachtungen $T \geq 2$, sonst haben wir zu viele Koeffizienten, um sie schätzen zu können
- Durch diese Sichtweise ergeben sich sogenannte *fixe Effekte*. Das heißt aber nicht, dass wir sie beliebig setzen können oder dass die nicht auch eine Verteilung haben können.

3. 1. Der FD-Schätzer

Namen für $\mathbf{a}_i$:

- individuenspezifische Effekte
- unbeobachtete Effekte
- fixe Effekte
- unbeobachtete Heterogenität

Für jedes Individuum haben wir so einen Term, den wir nicht gut schätzen können aus nur 2 Beobachtungen pro Individuum.

3. 1. Der FD-Schätzer

Aber:

$$
\begin{aligned}
\mathbf{y}_{i,2} &= \beta_0 + \beta_1 + \beta' \mathbf{x}_{i,2} + \mathbf{a}_i + \mathbf{u}_{i,2} \\
\mathbf{y}_{i,1} &= \beta_0 + \beta' \mathbf{x}_{i,1} + \mathbf{a}_i + \mathbf{u}_{i,1} \\
\mathbf{y}_{i,2} - \mathbf{y}_{i,1} &= \beta_1 + \beta' (\mathbf{x}_{i,2} - \mathbf{x}_{i,1}) + \mathbf{u}_{i,2} - \mathbf{u}_{i,1} \\
\Delta \mathbf{y}_i &= \beta_1 + \beta' \Delta \mathbf{x}_i + \Delta \mathbf{u}_i
\end{aligned}
$$

Die ersten Differenzen eliminieren die individuenspezifischen Effekte, so dass sie nicht modelliert werden müssen!

3. 1. Der FD-Schätzer

```
Call:
lm(formula = ccrmrte ~ cunem + clpolpc + clpop + cllawexp)

Residuals:
    Min      1Q  Median      3Q     Max
-32.910 -13.238  -4.345  13.917  43.747

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  17.6792     7.9176   2.233   0.0311 *
cunem         2.3870     0.8892   2.684   0.0104 *
clpolpc      82.5031    32.9810   2.502   0.0165 *
clpop       -15.7915    36.9820  -0.427   0.6716
cllawexp    -10.9950    16.8392  -0.653   0.5174
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 19.15 on 41 degrees of freedom
  (46 observations deleted due to missingness)
Multiple R-squared: 0.2575,Adjusted R-squared: 0.1851
F-statistic: 3.556 on 4 and 41 DF,  p-value: 0.01402
```

Interpretation des Modells:

- 1987 war die Kriminalitätsrate generell um 17.7 Vergehen pro Jahr pro 1000 Einwohner höher als 1982.
- Jedes Prozent an zusätzlicher Arbeitslosigkeit erhöht die Kriminalitätsrate um ca. 2.4 Vergehen.
- Auch die Zahl der Polizisten (`clpolpc:` change in log policemen per capita) zeigt eine positive Korrelation: je mehr Polizisten, umso mehr Vergehen. Keine Kausalität, sondern nur eine Gleichzeitigkeit!
- Bevölkerungszuwachs verringert die Kriminalitätsrate: das ist kontraintuitiv, der Koeffizient ist nicht statistisch signifikant.
- Ausgaben zur Erhöhung der Sicherheit haben den erwarteten negativen Koeffizienten sind aber auch nicht statistisch signifikant.

3. 1. Der FD-Schätzer

Allgemeiner:

- Wenn wir mehrere Jahre hinweg beobachten, können wir jeweils zwischen zwei Jahren Differenzen ziehen.
- Dies führt dann zu Gleichungen, in denen wir mittels KQ-Schätzern die Koeffizienten schätzen können.
- Diese Schätzer werden FD-Schätzer (First Difference) genannt.
- Das sind die einfachsten Paneldatenschätzer.

3. 1. Der FD-Schätzer

Wie lauten die Annahmen, unter denen die FD-Schätzer 'gute' Eigenschaften haben?

Annahme (FD: first differences)

FD.1: Der datengenerierende Prozess erfüllt die Gleichung

$$\mathbf{y}_{i,t} = \beta' \mathbf{X}_{i,t} + \mathbf{a}_i + \mathbf{u}_{i,t}, \quad , t = 1, ..., T, i = 1, ..., I$$

wobei $T \geq 2$. **FD.2:** Wir beobachten eine Zufallsstichprobe im Querschnitt (die Daten sind unabhängig über die Individuen hinweg).

FD.3: Kein Regressor ist konstant über die Zeit und über Individuen hinweg und es existieren keine perfekten linearen Beziehungen zwischen den Regressoren.

FD.4 (strikte Exogenität): Wenn $\mathbf{X}_i \in \mathbb{R}^{T \times K}$ die Regressormatrix für Individuum i bezeichnet dann gelte:

$$\mathbb{E}(\mathbf{u}_{i,t} | \mathbf{X}_i, \mathbf{a}_i) = 0$$

Unter diesen Annahmen ist der FD-Schätzer unverzerrt.

3. 1. Der FD-Schätzer

Annahme (FD: first differences (cont'ued))

FD.5: $\mathrm{Var}(\Delta \mathbf{u}_{i,t}|\mathbf{X}_i) = \sigma^2, t = 2, ..., T.$

FD.6: Für alle $t \neq s$ sind die differenzierten Fehler unkorreliert in der Zeit:

$$\mathrm{Cov}(\Delta \mathbf{u}_{i,t}, \Delta \mathbf{u}_{i,s}|\mathbf{X}_i) = 0$$

Unter diesen Annahmen ist der FD-Schätzer der BLUS. Besonders FD.6 ist sehr speziell.

Können wir andere Transformationen finden, die die Daten besser nutzen?