Berner Fachhochschule
Haute école spécialisée bernoise
Bern University of Applied Sciences

CAS Datenanalyse

Kapitel 5: Dummy-Variablen

Prof. Dr. Raúl Gimeno
FRM,CAIA, PRM

Dummy-Variable (binäre Variable)

Dummy-Variable = binäre Variable mit den Ausprägungen 1 und 0 → zeigt das Zutreffen eines bestimmen Umstandes

Wert 1 → Umstand trifft zu, sonst Wert 0.

Zweck: Quantifizierung von qualitativen Variablen → Auswirkungen qualitativer Unterschiede untersuchen

Beispiele:

- Zuordnung einer Person zu einer Gruppe
- Konjunktur/ Stagnation
- Zeit vor/nach Ölpreis-Schock
- Regionen (Stadt/Land)
- Saisonen des Jahres

Fragen:

Verdienen Männer im Erwartungswert signifikant mehr als Frauen?

Wie gross ist der Einkommensunterschied im Erwartungswert?

Dummy-Variablen: Beispiel

Lohnsatz von 12 Personen

D_w = 1 wenn weiblich und sonst 0

D_m = 1 wenn männlich und sonst 0

Regression: $\hat{y} = 18.21 - 1.96D_w$

Bedingter Erwartungswert für den Lohnsatz von Frauen (D_w =1):

$E(y|D_w = 1) = 18.21 - 1.96*1 = 16.25$

Bedingter Erwartungswert für den Lohnsatz von Männern (D_w =0):

$E(y|D_w = 0) = 18.21 - 1.96*0 = 18.21$

Interzept = Mittelwert!

	Y Lohnsatz	D_w F = 1	D_m M = 1	Σ	Frauen	Männer
	15.02	1	0	1	15.02	
	18.33	0	1	1		18.33
	18.81	0	1	1		18.81
	15.88	1	0	1	15.88	
	18.58	0	1	1		18.58
	17.04	1	0	1	17.04	
	17.27	0	1	1		17.27
	16.94	1	0	1	16.94	
	17.71	0	1	1		17.71
	16.36	1	0	1	16.36	
	18.57	0	1	1		18.57
	16.26	1	0	1	16.26	
Summe		6	6	12		
Mittelwert	17.23				16.25	18.21

Hinweis: 2 Spalten für D_w und D_m nur aus didaktischen Gründen dargestellt. Sonst nur 1 Dummy-Variable D_w (Referenzgruppe = Männer)

Dummy-Variablen: Interpretation

Regressionschätzung: $\hat{y} = 18.21 - 1.96D_w$

Lohnsatzdifferenz zwischen Frauen und Männern

$E(y|D_w = 1) - E(y|D_w = 1) = 16.25 - 18.21 = -1.96$

Interpretation: Frauen verdienen im Erwartungswert (Durchschnitt) um 1.96 Geldeinheiten weniger als Männer → Koeffizient der Dummy-Variable misst den Unterschied zur Referenzkategorie (Männer → $D_w = 0$)

Der bedingte Erwartungswert von y der Referenzkategorie (Männer) wird durch das Interzept gemessen.

Beide Dummy-Variablen dürfen nicht als Regressoren verwendet werden → sonst Multikollinearität → Dummy-Variable für die Referenzgruppe nicht benutzen!

$\hat{y} = b_1 + b_2D_w + \cancel{b_3D_m}$

Dummy-Variablen Falle

Auswirkung wenn beide Dummy-Variablen als Regressoren verwendet werden → fehlerhaftes Beispiel:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_{12} \end{pmatrix} = \beta_1 \begin{pmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} + \beta_2 \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} + \beta_3 \begin{pmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_{12} \end{pmatrix} \qquad \mathbf{X} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \\ \vdots & \vdots & \vdots \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$$

Konsequenz: **1** = $D_w + D_m$

Spalte 1 = Spalte 2 + Spalte 3

→Spalten der Matrix **X** sind linear abhängig: $D_w + D_m = 1$

→ $\mathbf{X'X}$ singulär →OLS-Schätzer $\mathbf{b} = (\mathbf{X'X})^{-1}\mathbf{X'y}$ ist nicht definiert!

→ Dummy Variablen Falle!

Marginaler Effekt

Marginaler Effekt für Dummy-Variablen =
Differenz der bedingten Erwartungswerte

Regressionsfunktion: $\mathbf{y} = \beta_1 + \beta_2 D + \mathbf{u}$

$E(y \mid D = 1) = \beta_1 + \beta_2$ $\qquad$ $E(y \mid D = 0) = \beta_1$

Marginaler Effekt: $\quad E(y \mid D = 1) - E(y \mid D = 0) = \beta_2$

Interzept = b_1: Schätzer für den bedingten Mittelwert der Kategorie $D = 0$

$b_1 + b_2$: Schätzer für den bedingten Mittelwert der Kategorie $D = 1$

Unterschiede im Interzept

Regression mit einer Dummy Variable und einer erklärenden Variable:

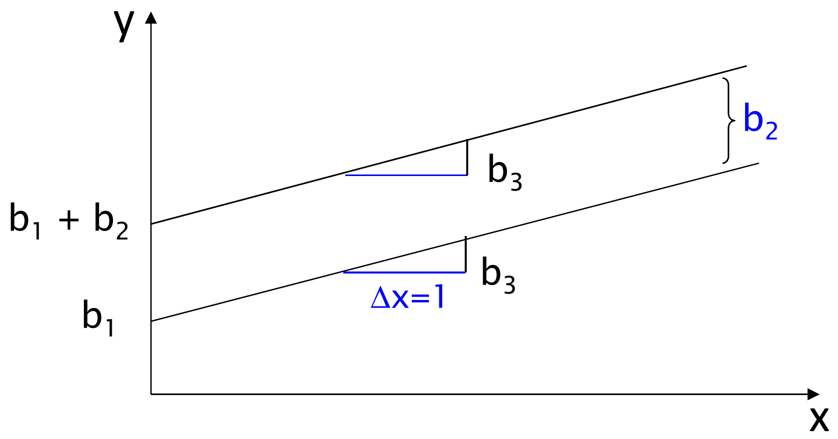
$$y_t = \beta_1 + \beta_2 D_t + \beta_3 x_t + u_t$$

$D = 0 \rightarrow\ y_t = \beta_1 + \beta_3 x_t + u_t$ $\qquad E(y|D = 0) = \beta_1 + \beta_3 x_t$

$D = 1 \rightarrow\ y_t = \underbrace{(\beta_1 + \beta_2)}_{\text{Interzept}} + \beta_3 x_t + u_t$ $\qquad E(y|D = 1) = \beta_1 + \beta_2 + \beta_3 x_t$

$\hat{y}_t = b_1 + b_2 D_t + b_3 x_t$

Marginaler Effekt: $E(y|D = 1) - E(y|D = 0) = \beta_2 \rightarrow$ Unterschied im Interzept



Wenn
$\beta_1 > 0,\ \beta_2 > 0,\ \beta_3 > 0$
Dummy führt zu einer Parallelverschiebung der Regressionsgeraden um den Betrag b_2.
Steigung b_3 bleicht gleich

Unterschiede im Interzept und Steigung

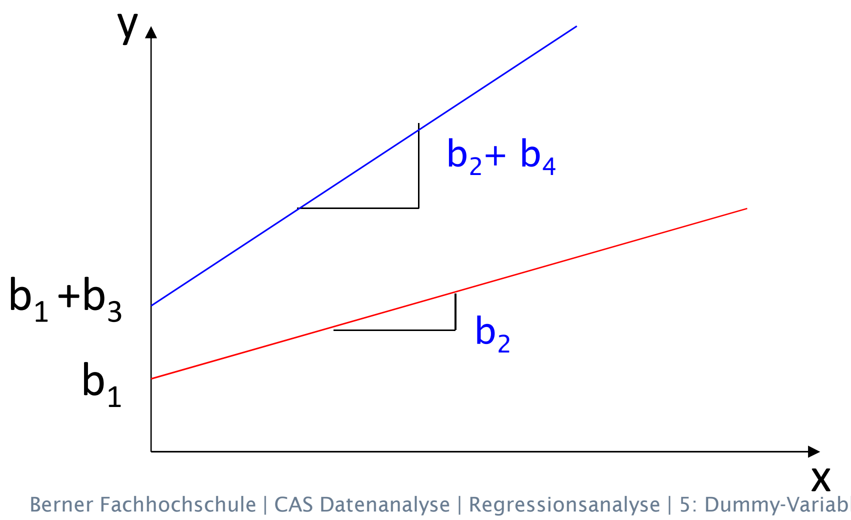
$$y_t = \beta_1 + \beta_2 x_t + \beta_3 D_t + \beta_4 \underbrace{D_t x_t}_{} + u_t$$

Interaktion zwischen Dummy und Regressor (Interaktionsterm)

$D = 0 \rightarrow\ y_t = \beta_1 + \beta_2 x_t + u_t$ $\qquad E(y|D=0) = \beta_1 + \beta_2 x_t$

$D = 1 \rightarrow\ y_t = (\beta_1+\beta_3) + (\beta_2+\beta_4)x_t + u_t$ $\qquad E(y|D=1) = (\beta_1+\beta_3) + (\beta_2+\beta_4)x_t$

Marginaler Effekt: $E(y|D=1) - E(y|D=0) = \beta_3 + \beta_4 x_t$



Wenn
$\beta_1 >0,\ \beta_2 >0,\ \beta_3 >0,\ \beta_4 >0$
→ Zwei Regressionsgeraden mit unterschiedlichen Koeffizienten und Interzepten.

Interaktionseffekte

Frage: Wirkt sich der Familienstand (Dummy: verheiratet/ledig) für Männer und Frauen unterschiedlich auf y (Stundenlohn) aus?

$D_V = 1$ für verheiratete und sonst 0

$D_W = 1$ für eine Frau und sonst 0

Modell: $y_t = \beta_1 + \beta_2 D_w + \beta_3 D_V + \beta_4 D_w D_V + \beta_5 x_t + u_t$

(1) Unverheirateter Mann: $E(y | D_W=0, D_V=0) = \beta_1 + \beta_5 x$

(2) Unverheiratete Frau: $E(y | D_W=1, D_V=0) = \beta_1 + \beta_2 + \beta_5 x$

(3) Verheirateter Mann: $E(y | D_W=0, D_V=1) = \beta_1 + \beta_3 + \beta_5 x$

(4) Verheiratete Frau: $E(y | D_W=1, D_V=1) = \beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 x$

Erwarteter Lohnunterschied zwischen verheirateten und unverh. Frauen

(4)–(2): $E(y | D_W=1, D_V=1) - E(y | D_W=1, D_V=0) = \beta_3 + \beta_4$

Erwarteter Lohnunterschied zwischen verheirateten Frauen und Männern:

(4)–(3): $E(y | D_W=1, D_V=1) - E(y | D_W=0, D_V=1) = \beta_2 + \beta_4$

2 Dummy Variablen → 4 Fälle

Interaktionseffekte

Alternative Möglichkeit:

$D_{MV} = 1$ für verheiratete Männer und sonst 0

$D_{WV} = 1$ für verheiratete Frauen und sonst 0

$D_{WL} = 1$ für ledige Frauen und sonst 0

Referenzkategorie: ledige Männer

Modell: $y_t = \beta_1 + \beta_2 D_{MV} + \beta_3 D_{WV} + \beta_4 D_{WL} + \beta_5 x_t + u_t$

Lohngleichung:

Modell 1: $y_t = b_1 + b_2 D_w + b_3 D_v + e_t$

Modell 2: $y_t = b_1 + b_2 D_w + b_3 D_v + b_4 D_w D_v + e_t$

b_1 gibt den Durchschnittslohn der Kategorie «unverheiratete Männer» in Modell 2, aber nicht im Modell 1!

Interpretation von Dummies in log-lin Gleichungen

Modell: $\ln(y_t) = b_1 + b_2x_t + b_3D_w + e_t$

Um wieviel Prozent unterscheidet sich $\hat{y}$ für D = 1 von der Kategorie mit D = 0, wenn x_t konstant gehalten wird (ceteris paribus)?

D = 1: $E(\ln y | D = 1) = b_1 + b_2x_t + b_3$

D = 0: $E(\ln y | D = 0) = b_1 + b_2x_t$

$E(\ln y | D = 1) - E(\ln y | D = 0) = b_3$

$$E[\ln(y|D=1) - \ln(y|D=0)] = E\left[\ln\left(\frac{(y|D=1)}{(y|D=0)}\right)\right] = b_3$$

$$E\left[\frac{(y|D=1)-(y|D=0)}{(y|D=0)}\right] \times 100 = [\exp(\beta_3) - 1]100$$

Berücksichtigung der systematischen Verzerrung wegen $E[\exp(e)] = 0.5\sigma^2$

Genauerer Schätzwert: $(\exp[b_3 - 0.5\widehat{\mathrm{var}}(b_3)] - 1) \times 100$

Difference-in-Difference

Situation: In einer Stadt wurde eine neue Umfahrungsstrasse gebaut.

Frage: Welche Auswirkungen hat dies auf die Immobilienpreise gehabt?

«Was-wäre-wenn» Frage: Wenn die Strasse gebaut wurde, fehlt das Kontrafaktum (wie wären die Preise, wenn die Strasse nicht gebaut worden wäre).

Mögliche Lösung: Sie vergleichen einfach den Mittelwert der Grundstückpreise vor dem Bau mit den Preisen nach dem Bau der Umfahrungsstrasse.

Problem: Während des Baus sind die Immobilienpreise im Allgemeinen gestiegen. Man würde diesen Preisanstieg fälschlicherweise der Umfahrungsstrasse zuschreiben.

Lösung: Preise vor und nach dem Bau mit den Grundstückpreisen einer nicht betroffenen Region der Stadt vergleichen.

Treatment und Kontrollgruppe

Treatment Gruppe: Gruppe, die von einer Veränderung betroffen wurde (bzw. der einer Behandlung zuteil wurde)

Kontrollgruppe: Wurde keiner Behandlung zugeteilt.

Before: Periode vor der Veränderung

After: Periode nach der Veränderung

	Treatment Gruppe die Betroffenen	Kontrollgruppe die Nicht-Betroffenen
Before (vor)	T_B	C_B
After (nach)	T_A	C_A

Schätzung der verursachten Preisänderung → Differenz der Differenz der Mittelwerte

Difference-in-Difference: DiD = $(T_A - T_B) - (C_A - C_B)$

Problem: Kaum genügende vergleichbare Immobilienpreise in den Gruppen → Immobilien unterscheiden sich in Bezug auf Grösse, Lage, Ausstattung…

Difference-in-Difference: Regression

Lösung mittels Regressionsgleichung:

$y_i = \beta_1 + \beta_2 treat + \beta_3 after + \beta_4 treat{\times}after + \beta_5 x_i + u_i$

Dummy Variablen

$treat \begin{cases} = 1 \text{ wenn in Treatment Gruppe} \\ = 0 \text{ wenn in Kontrollgruppe} \end{cases}$ $after \begin{cases} = 0 \text{ vor Treatment} \\ = 1 \text{ nach Treatment} \end{cases}$

	Treatment-Gruppe	Kontrollgruppe	Differenz
Before	$\beta_1 + \beta_2 + \beta_5 x_i$	$\beta_1 + \beta_5 x_i$	β_2
After	$\beta_1 + \beta_2 + \beta_3 + \beta_4 + \beta_5 x_i$	$\beta_1 + \beta_3 + \beta_5 x_i$	$\beta_2 + \beta_4$
Differenz	$\beta_3 + \beta_4$	β_3	β_4

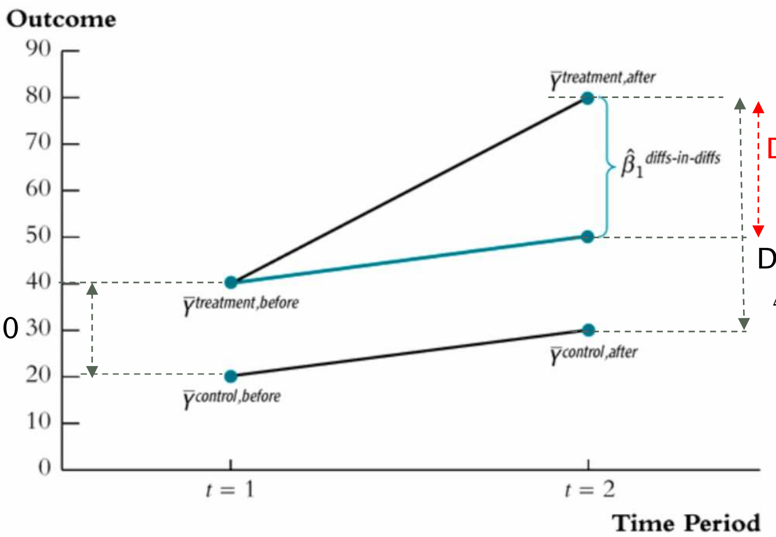
β_4= Koeffizient des Interaktionsterms → Difference-in-Difference Schätzer

Problem: Der DiD-Schätzer ist nur bei einer tatsächlichen Zufallsauswahl der Treatment Gruppe anwendbar.

Sozialwissenschaft → echte Zufallsauswahl nur selten möglich!

Difference-in-Difference

Das Vorgehen lässt sich an der folgenden Graphik veranschaulichen.



DiD = (80-30) –(40-20) = 30

Differenz Betroffenen-Nicht-Betrof.
Δ = 80-30 =50

Δ = 40-20 = 20

Beispiele: Auswirkung von Mindestlöhnen auf die Beschäftigung
Auswirkung von Einwanderung auf die Arbeitslosigkeit der Einheimischen

Strukturbruchmodell

- Untersuchung: Was ist der Einfluss des Wirtschaftswachstums auf die Veränderung der Erwerbslosenquote in Deutschland?
- Jahre 2003-2005 Reform der Sozialsysteme (Hartz IV) unter Kanzler Gerhard Schröder.
- Zeitraum: 1992-2014
- Daten: reale BIP-Wachstumsrate und Veränderung der Erwerbslosenquote
- Zeitraum vor Umsetzung der Reform: 1992-2004
- Zeitraum nach Umsetzung der Reform: 2005-2014
- Strukturbruch: Veränderung in der parametrischen Struktur des wahren Zusammenhangs → abrupte Veränderung eines oder mehrerer Parameterwerte.

Strukturbruchmodell

Getrennte Behandlung → zwei Regressionsmodelle:

Phase I: $y_t = \alpha_I + \beta_I x_t + u_t$ Phase II: $y_t = \alpha_{II} + \beta_{II} x_t + u_t$

α_I , β_I: wahre Parameter der Phase I

α_{II} , β_{II}: wahre Parameter der Phase II

Frage: Gibt es einen Strukturbruch?

→ Haben sich beide Parameterwerte verändert?

→ Hat sich nur ein Parameterwert verändert?

Vier Möglichkeiten:

Fall 1: $\alpha_I = \alpha_{II}$, $\beta_I = \beta_{II}$ → kein Strukturbruch

Fall 2: $\alpha_I \neq \alpha_{II}$, $\beta_I = \beta_{II}$ → Strukturbruch im Niveauparameter

Fall 3: $\alpha_I = \alpha_{II}$, $\beta_I \neq \beta_{II}$ → Strukturbruch im Steigungsparameter

Fall 4: $\alpha_I \neq \alpha_{II}$, $\beta_I \neq \beta_{II}$ → Strukturbruch im Niveau- & Steigungsparameter

Frage: Wie kann ein Regressionsmodell spezifiziert werden, das die Phasen I und II in einer einzigen Gleichung zusammenfasst und dabei die 4 Fälle als spezielle Variante enthält?

Strukturbruchmodell

Parameter der Phase II

$\alpha_{II} = \alpha_I + \gamma$ → Differenz zwischen den Niveauparametern

$\beta_{II} = \beta_I + \delta$ → Differenz zwischen den Steigungsparametern

Phase II Modell: $y_t = \alpha_{II} + \beta_{II} x_t + u_t = \alpha_I + \gamma + (\beta_I + \delta) x_t + u$

$$y_t = \alpha_I + \gamma + \beta_I x_t + \delta x_t + u$$

Phase I: $y_t = \alpha_I + 0 \times \gamma + \beta_I x_t + 0 \times \delta x_t + u$ $\quad t = 1, 2, \ldots. T_I$

Phase II: $y_t = \alpha_I + 1 \times \gamma + \beta_I x_t + 1 \times \delta x_t + u$ $\quad t = T_I+1, T_I+2, \ldots. T$

Strukturbruchmodell

Mithilfe einer Dummy-Variable können zwei Gleichungen in einer zusammengefasst.

$$D_t = \begin{cases} 0 \text{ wenn } t = 1, 2 \ldots T_I \\ 1 \text{ wenn } t = T_I+1, T_I+2 \ldots . T \end{cases}$$

Strukturbruchmodell: $y_t = \alpha_I + \gamma D_t + \beta_I x_t + \delta D_t x_t + u_t$

Fall 1: $\gamma = 0$ und $\delta = 0 \rightarrow y_t = \alpha_I + \beta_I x_t + u_t \rightarrow$ kein Strukturbruch

Fall 2: $\gamma \neq 0$ und $\delta = 0 \rightarrow y_t = \alpha_I + \gamma D_t + \beta_I x_t + u_t$
$\rightarrow$ Strukturbruch im Niveauparameter

Fall 3: $\gamma = 0$ und $\delta \neq 0 \rightarrow y_t = \alpha_I + \beta_I x_t + \delta D_t x_t + u_t$
$\rightarrow$ Strukturbruch im Steigungsparameter

Fall 4: $\gamma \neq 0$ und $\delta \neq 0 \rightarrow y_t = \alpha_I + \gamma D_t + \beta_I x_t + \delta D_t x_t + u_t$
$\rightarrow$ Strukturbruch im Niveau- und Steigungsparameter

Schätzung und Interpretation des Strukturbruchmodells

K = 4 (Interzept + 3 Steigungsparameter)
N = 23 (Zeitperiode 1992 - 2014) df = N-K = 23-4 = 19

	Koeff.	se(.)	t-Wert	p-Wert
Konstante	0.809	0.219	3.695	0.002
Wachstum	-0.374	0.126	-2.975	0.008
Dummy	-1.106	0.286	-3.862	0.001
Inter-Dummy	0.233	0.140	1.664	0.112

$a_1 = 0.809$: Niveauparameter der Phase I $\rightarrow$Nullwachstum würde die Erwerbslosenquote in Phase I um 0.809 Prozentpunkte erhöhen.

$b_I = -0.374$: Ein zusätzlicher Prozentpunkt reales Wirtschafts-wachstum senkt die Erwerbslosenquote um 0.374 Prozentpunkte.

$\hat{\gamma} = -1.106$: $\alpha_{II} = \alpha_I + \gamma \rightarrow$ Der Niveauparameter der Phase II ist um 1.106 kleiner als derjenige der Phase I

$\hat{\delta} = 0.233$: $\beta_{II} = \beta_I + \delta \rightarrow$ Der Steigungsparameter der Phase II ist um 0.233 über demjenigen der Phase I

Konklusion

$b_{II} = b_I + \hat{\delta} = -0.374 + 0.233 = -0.141$

$a_{II} = a_I + \hat{\gamma} = 0.809 - 1.106 = -0.296$

In der Phase II ergibt sich bei Nullwachstum ein leichter Rückgang der Erwerbslosenquote (a_{II} = -0.297) während sich in Phase I ein deutlicher Anstieg der Erwerbslosenquote ergab (a_I = 0.809).

Die Empfindlichkeit der Erwerbslosenquote ist in Bezug auf das Wirtschaftswachstum in Phase II (b_{II} = -0.141) geringer als in Phase I (b_{II} = -0.374)

Getrennte Schätzungen

	Koeff.	se(.)	t-Wert	p-Wert
Konstante	0.809	0.206	3.939	0.002
Wachstum	-0.374	0.118	-3.171	0.009
Konstante	-0.296	0.199	-1.489	0.175
Wachstum	0.141	0.066	-2.129	0.066

Vernachlässigung des Strukturbruchs

Welche Schätzprobleme ergeben sich, wenn man fälschlicherweise unterstellt, es gäbe keinen Strukturbruch?

Modell: $y_t = \alpha + \beta x_t + u_t$

	Koeff.	se(.)	t-Wert	p-Wert
Konstante	0.216	0.169	1.277	0.216
Wachstum	-0.186	0.071	-2.613	0.016

Falls der wahre Wirkungszusammenhang durch einen Strukturbruch gekennzeichnet ist, sind die Schätzer a und b des obigen Modells vollkommen wertlos, denn die Parameter α und β existieren nicht.

→ Schätzung eines Scheinzusammenhangs!

Diagnose mittels F-Test

Nullhypothese H_0: kein Strukturbruch $\leftrightarrow \gamma = \delta = 0$

Unrestringiertes Modell: $y_t = \alpha_I + \gamma D_t + \beta_I x_t + \delta D_t x_t + u_t$

Restringiertes Modell: $y_t = \alpha + \beta x_t + u_t$

Schätzung: $y_t = 0.216 - 0.186 x_t \rightarrow RSS_r = 9.664$

F-Statistik: $F = \frac{(RSS_r - RSS)/L}{RSS/(N-K)} \approx F_{(K-1,N-K)}$

K = 4 (Interzept + 3 Steigungsparameter)

Anzahl Restriktionen: L = 2

Freiheitsgrade: N-K = 23-4 = 19

$$F_e = \frac{(9.664-5.252)/2}{5.252/19} = 7.978$$

Signifikanzniveau: $\alpha = 5\%$

Kritischer Wert: $F_c(0.095,2,19) = 3.52$

$F_e > F_c \rightarrow H_0$ kann verworfen werden $\rightarrow$ Strukturbruch hat stattgefunden

Diagnose mittels t-Test

Nullhypothese H_0: kein Strukturbruch $\leftrightarrow \gamma = \delta = 0$

Frage: Liegt Fall 1 vor oder nicht?

Es wurde nicht zwischen den Fällen 2,3 und 4 differenziert.

Fall 2: $\gamma \neq 0$ und $\delta = 0 \rightarrow y_t = \alpha_I + \gamma D_t + \beta_I x_t + u_t$

Fall 3: $\gamma = 0$ und $\delta \neq 0 \rightarrow y_t = \alpha_I + \beta_I x_t + \delta D_t x_t + u_t$

Fall 4: $\gamma \neq 0$ und $\delta \neq 0 \rightarrow y_t = \alpha_I + \gamma D_t + \beta_I x_t + \delta D_t x_t + u_t$

Unterscheidung Fall 3-4: Nullhypothese H_0: $\gamma = 0$

$t_e = -3.862 < t_c(0.95,19) = 2.09 \rightarrow H_0$ verwerfen

Unterscheidung Fall 2-4: Nullhypothese H_0 : $\delta = 0$

$t_e = 1.664 < t_c(0.95,19) = 2.09 \rightarrow H_0$ nicht verwerfen

Chow-Test

Unrestringiertes Modell: $y_t = \alpha_I + \gamma D_t + \beta_I x_t + \delta D_t x_t + u_t$

Restringiertes Modell: $y_t = \alpha + \beta x_t + u_t$

K = 1 # exogene Variablen

RSS_r: Summe der Residuenquadrate des restringierten Modells über alle Perioden

RSS_I: Summe der Residuenquadrate in der Phase I

T_{II}: Anzahl Beobachtungen in der Phase II

$$F = \frac{(RSS_r - RSS_I)/L}{RSS_I/(N-K)} \approx F_{(T_{II}, T_I - K)}$$

Oft möchte man unmittelbar nach einer umgesetzten wirtschaftspolitischen Massnahmen überprüfen, ob sich eine strukturelle Veränderung im Wirkungszusammenhang zwischen endogener und exogenen Variablen ereignet hat.

Problem: Nur eine Beobachtung für Phase II → kein F-Test!

Chow-Test: Beispiel

Wir sind im Jahre 2005. Hat die zu Beginn des Jahres 2005 eingeführten Hartz IV Gesetze noch im gleichen Jahr zu einem Strukturbruch im Wirkungszusammenhang zwischen dem Wirtschaftswachstum und der Veränderung der Erwerbslosenquote geführt?

Nur die Jahresdaten 1992-2005 liegen vor.

Beobachtungsumfang: T = 14, T_I = 13 und T_{II} = 1

T_{II} = 1 →der einfache F-Test kann nicht angewendet werden → Chow-Test

Beobachtungen 1992-2005 (T = 14)

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	0,837418	0,191022	4,384	0,0009	***
gdp	−0,382037	0,113095	−3,378	0,0055	***

Mittel d. abh. Var.	0,357143	Stdabw. d. abh. Var.	0,640570
Summe d. quad. Res.	2,734252	Stdfehler d. Regress.	0,477341

RSS_r

Beobachtungen 1992-2004 (T = 13)

	Koeffizient	Std.-fehler	t-Quotient	p-Wert	
const	0,809461	0,205515	3,939	0,0023	***
gdp	−0,374142	0,117975	−3,171	0,0089	***

Mittel d. abh. Var.	0,323077	Stdabw. d. abh. Var.	0,653394
Summe d. quad. Res.	2,676189	Stdfehler d. Regress.	0,493244

RSS_I

T_I-K = 13 - 2 = 11

$$F = \frac{(RSS_r - RSS_I)/L}{RSS_I/(N-K)}$$

$$F = \frac{(2.734 - 2.676)/1}{2.676/11} = 0.239$$

F(0.095,1,11) = 4.84

H_0: kein Strukturbruch

$F_e < F_c \rightarrow H_0$ kann nicht verworfen werden

Chow-Test: Test auf Strukturbrüche

Regression (1): Stichprobe von Männern

Regression (2): Stichprobe von Frauen

Frage: Unterscheiden sich die Koeffizienten beider Gruppen?

Sie vermuten, dass in der Grundgesamtheit zwischen zwei Gruppen oder Perioden Unterschiede bestehen:

- Gruppe 1/ Periode 1:

  Regression 1: $y_t = \beta_1 + \beta_2 x_{2t} + \beta_3 x_{3t} + \ldots + \beta_k x_{kt} + u_t$ $\qquad$ $t = 1,2,..,n_1$

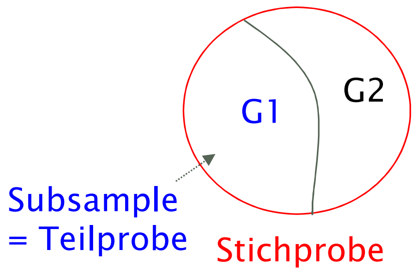
- Gruppe 2/ Periode 2:

  Regression 2: $y_t = \beta_1' + \beta_2' x_{2t} + \beta_3' x_{3t} + \ldots + \beta_k' x_{kt} + u_t'$ $\qquad$ $t = n_1+1,\ldots,N$

- Nullhypothese H_0: $\beta_1 = \beta_1', \quad \beta_2 = \beta_2', \ \ldots \ \beta_k = \beta_k'$

Unrestringiertes Modell: zwei unabhängige geschätzten Gleichungen für beide Gruppen (Subsamples = Teilproben)

- Gruppe 1: $y_t = b_1 + b_2 x_{2t} + b_3 x_{3t} + \ldots + b_k x_{kt} + e_t \quad \longrightarrow RSS^1$
- Gruppe 2: $y_t = b_1' + b_2' x_{2t} + b_3' x_{3t} + \ldots + b_k' x_{kt} + e_t' \quad \longrightarrow RSS^2$



Chow-Test: F-Statistik

Man kann zeigen, dass die nicht-restringierte Quadratsumme der Residuen die Summe der Quadratsummen der Residuen der beiden Gleichungen (1) und (2) ist: $RSS = RSS^1 + RSS^2$

- Freiheitsgraden: $(n_1 - K) + (N - n_1 - K) = N - 2K$
- H_0 wahr $\Leftrightarrow$ alle Koeffizienten der beiden Modelle übereinstimmen
- Das restringierte Modell wird über alle N Beobachtungen (beide Gruppen oder Perioden) geschätzt.

  $y_t = b_1^* + b_2^* x_{2t} + b_3^* x_{3t} + \ldots + b_k^* x_{kt} + e_t^* \quad \longrightarrow SSR_r$

  N Beob. gesamte Stichprobe

Man kann zeigen, dass folgende Teststatistik für die Nullhypothese F-verteilt ist:

$$F = \frac{(RSS_r - RSS)/K}{RSS/(N-2K)} \sim F_{K,N-2K}$$

N: Anzahl Beobachtungen

K: Anzahl der geschätzten Koeffizienten

- Gültigkeit: Nur wenn Varianz der Störterme σ^2 in beiden Gruppen gleich gross ist!