
Kapitel 9

Analyse der Modellstruktur

Dummies
Strukturbrüche

Analyse der Modellstruktur

- Datenfenster: increasing und moving windows
- Dummy-Variable
- Strukturbruch,
- Tests auf Strukturbrüche: Chow-Test, Chow Forecast Test, CUSUM-Test
- Rekursive Residuen

Increasing window oder Rekursive OLS-Schätzung

Das Modell lautet in Matrixform: $y = X\beta + u$
 y, u : $(n \times 1)$ -Spaltenvektoren; X : Dimension $n \times k$,
 β : $(k \times 1)$ -Vektor

Wir wählen ein Datenfenster, das stets mit der ersten
Beobachtung beginnt, dessen Ende (in t) mit t wächst.
 b_t : OLS-Schätzer für β aus Beobachtungen (y_i, x_i) , $i=1, \dots, t$
$$b_t = (X_t' X_t)^{-1} X_t' y_t, \quad t=k+1, \dots, n \quad \text{und}$$
$$\text{Var}\{b_t\} = \sigma^2 (X_t' X_t)^{-1}$$

X_t : Dimension $(t \times k)$, y_t : $(t \times 1)$ -Vektor (vgl. Ö1 K3/F29,36)

Konsumfunktion

OLS-Anpassung an österreichische Jahres-Daten

1954 bis 1999:

$$\hat{C} = 1.948 + 0.897 Y$$

rekursiv

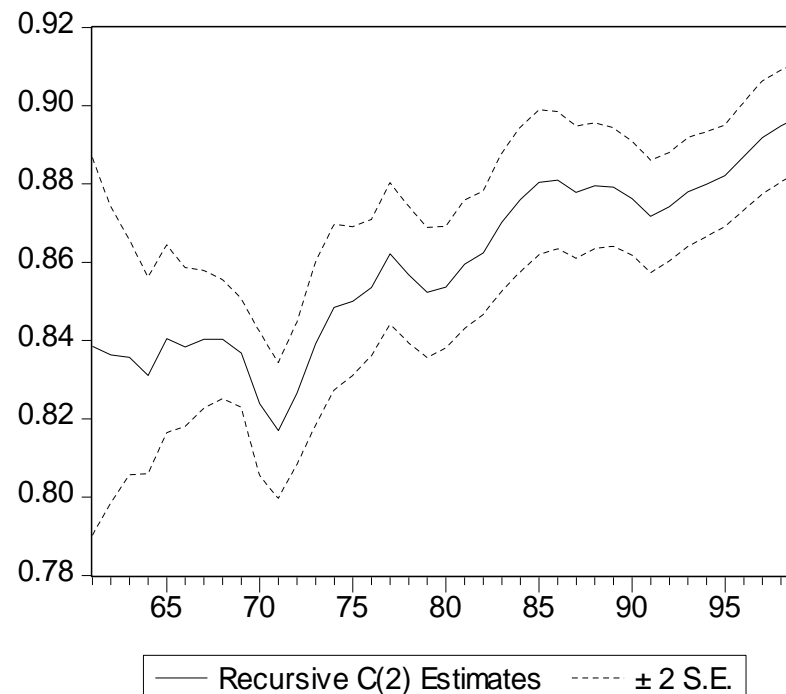
(increasing window)

geschätzte marginale
Konsumneigung und
zugehöriges

95%-Konfidenzband.

1971: Einbruch

1972-99: Anstieg



Dummy-Variable

Dummy-Variable:

Regressor, der ein Ereignis oder einen Umstand anzeigt;
er hat den Wert 1 in Perioden, in denen der Umstand
zutrifft, sonst den Wert 0

Beispiele:

	t	D08	Nach08
■ Periode einer Hochkonjunktur/Stagnation	-----		
■ Zusammenbruch d Interbankenmarktes	.	0	0
■ Region (Stadt/Land)	07	0	0
■ Saison eines Jahres	08	1	1
■ Markieren von Ausreißer	09	0	1
■ Zeit vor/nach Ölpreis-Schock	.	0	1

Dummy-Variable für Saison

Saison-Dummies für Quartalsdaten sind definiert als:

$$Q_{it} = 1, \quad i\text{-tes Quartal}$$

$$Q_{it} = 0, \quad \text{sonst}$$

Frühlings-Dummy Q_{1t} hat den Wert 1 in jedem ersten Quartal und sonst 0; analog das Sommer-Dummy ($i = 2$), etc.

Beachte: Für jede Periode ($t = 1, \dots, n$) gilt

$$Q_{1t} + Q_{2t} + Q_{3t} + Q_{4t} = 1$$

Man verwendet in der Regressionsgleichung entweder 4 Dummies und keinen Interzept, oder 3 Dummies und den Interzept.

Dummy-Variable für Saison

Jahr	Q	D1	D2	D3	D4		C
2010	Q1	1	0	0	0		1
2010	Q2	0	1	0	0		1
2010	Q3	0	0	1	0		1
2010	Q4	0	0	0	1		1
2011	Q1	1	0	0	0		1
2011	Q2	0	1	0	0		1
2011	Q3	0	0	1	0		1
2011	Q4	0	0	0	1		1

$$D1 + D2 + D3 + D4 = 1$$

EViews: $D2 = @seas(2)$

Modelle für Quartalsdaten

Das Modell $Y = \alpha + \beta X + u$ berücksichtigt keine Saison

Modell mit seasonspezifischem Interzept und Anstieg, wobei nur für die 1-ten Quartale, 2-ten Quartale, ...

$$Y_t = \alpha_1 + \beta_1 X_t + u_t \quad \text{nur 1-te Quartale}$$

$$Y_t = \alpha_2 + \beta_2 X_t + u_t \quad \text{nur 2-te Quartale}$$

$$Y_t = \alpha_3 + \beta_3 X_t + u_t \quad \text{nur 3-te Quartale}$$

$$Y_t = \alpha_4 + \beta_4 X_t + u_t \quad \text{nur 4-te Quartale}$$

geschätzt wird.

Schreibweise mit Saison-Dummyvariablen Q_{it} :

$$Y_t = \sum_i \alpha_i Q_{it} + \sum_i \beta_i Q_{it} X_t + u_t \quad \text{für alle Datenpunkte}$$

Modelle für Quartalsdaten

Modell mit seasonspezifischem Interzept:

$$Y_t = \alpha_1 Q_{1t} + \alpha_2 Q_{2t} + \alpha_3 Q_{3t} + \alpha_4 Q_{4t} + \beta X_t + u_t$$

Die α_i messen die Effekte der einzelnen Quartale.

Saisonspezifischer Interzept mit Konstanter:

$$Y_t = \alpha_1 + \delta_2 Q_{2t} + \delta_3 Q_{3t} + \delta_4 Q_{4t} + \beta X_t + u_t$$

mit $\delta_i = \alpha_i - \alpha_1, \quad i = 2, 3, 4.$

- α_1 misst den Effekt des 1-ten Quartals,
- die δ_i messen die Unterschiede der einzelnen Quartale zum 1-ten.
- Im 1.Quartal steht nur beim Interzept eine 1. (Datenmatrix)
- Die anderen Q-Effekte werden zw Interzept und Q geteilt.
Z.B.: $\alpha_2 = \delta_2 + \alpha_1$

Modelle für Quartalsdaten, Forts.

Modell mit gemeinsamem Interzept, aber
saisonspezifischem Anstieg

$$Y_t = \alpha + \sum_i \beta_i Q_{it} X_t + u_t \quad \text{bzw.}$$

$$Y_t = \alpha + \beta_1 Q_{1t} X_t + \beta_2 Q_{2t} X_t + \beta_3 Q_{3t} X_t + \beta_4 Q_{4t} X_t + u_t$$

Strukturbruch

Der datengenerierende Prozess kann in Teilbereichen des Beobachtungszeitraums durch das gleiche Modell beschrieben werden;

In den Teilbereichen (Regimen) liegen aber unterschiedliche Parameterwerte vor.

Strukturbruch-Analyse:

- ❑ Gibt es Teilbereiche mit unterschiedlichen Strukturen?
- ❑ Wann hat der Strukturbruch stattgefunden? Schätzung des Zeitpunktes des Strukturbruchs (*change point*)
- ❑ Change point detection Tests dienen zum Auffinden von change points. (Es werden in der Regel alle möglichen Zeitpunkte durchprobiert, und der mit dem max Ausschlag getestet.)

Chow-Test

Chow-Test: Zum Entscheiden, ob unterschiedliche Strukturen vorliegen, oder nicht

Voraussetzungen:

1. Teilbereiche mit konstanter Struktur können identifiziert werden
2. Der Zeitpunkt, zu dem der Übergang von einem Regime zum anderen stattgefunden hat, ist bekannt
3. Die Anzahl der Beobachtungen in jedem Regime, so dass das Modell an die Daten jedes einzelnen Regimes angepasst und die Residuen bestimmt werden können, ist ausreichend

Oft erlauben Dummies das Modellieren von Regimen

Chow-Test, Forts.

Nullhypothese:

die Regressionsparameter sind in allen Teilbereichen des Beobachtungszeitraums gleich

Alternative:

zu bestimmten Zeitpunkten ändert das Interzept oder/und einige oder alle anderen Regressionsparameter ihre Werte

Chow-Test, Forts.

Modell mit zwei Regimen:
$$\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} X_1 & 0 \\ 0 & X_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$$

die partitionierten Größen y , X , β und u entsprechen den Größen vor und nach dem Strukturbruch

Nullhypothese (kein Strukturbruch) wird mittels F -Test überprüft: $H_0: \beta_1 = \beta_2$ (restringiertes Modell)

$$F = \frac{S_R - S}{S} \frac{n - 2k}{k}$$

S : Summe der Fehlerquadrate aus den separat geschätzten Modellen aus beiden Regimen, $S = S_1 + S_2$

S_R : Summe der Fehlerquadrate im Modell unter H_0

Chow-Test, Forts.

Die F -Statistik folgt bei Zutreffen von H_0

- der F -Verteilung $F(k, n-2k)$ bei normalverteilten Störgrößen
- näherungsweise der Chi-Quadrat-Verteilung $\chi^2(k)$ bei großem n

$S = S_1 + S_2$:

S_1 stammt aus der Schätzung für Periode 1 der Länge n_1 ,

S_2 stammt aus der Schätzung für Periode 2 der Länge n_2 .

Die zugehörigen Freiheitsgrade sind $(n_1 - k)$ und $(n_2 - k)$.

Für S $(n-2k) = (n_1 - k) + (n_2 - k)$.

Die Freiheitsgrade von S_R sind: $n - k$

Daher: Freiheitsgrade für den Zähler $(S_R - S)$ gleich k .

Konsumfunktion, Forts.

OLS-Anpassung der Konsumgleichung $C_t = \alpha + \beta Y_t + u_t$
an 2 Regime

1. 1954 bis 1971: $b = 0.817$, $S_1 = 200.68$
2. 1972 bis 1999: $b = 0.824$, $S_2 = 5107.17$

F-Statistik:

$$F = \frac{6899.7 - (200.68 + 5107.17)}{200.68 + 5107.17} \frac{46 - 4}{2} = 6.30$$

p -Wert: 0.004

Eviews: $p = 1 - @cfdist(x, df1, df2)$ bzw.

$p = 1 - @cchisq(x, df)$

Chow-Test für m Regime

Verallgemeinerung auf m Regime

$$H_0: \beta_1 = \dots = \beta_m$$

F -Statistik

$$F = \frac{S_R - \sum_i S_i}{\sum_i S_i} \frac{n - mk}{(m-1)k}$$

S_i : Summe der Fehlerquadrate im Modell des i -ten Regimes

$(i = 1, \dots, m)$

Verteilung von F : $F([m-1]k, n-mk)$ oder $\chi^2([m-1]k)$

Chow's Prognosetest

Ändert sich die Struktur am Ende des Beobachtungszeitraums in einem Bereich kürzer als k , $p < k$, ist der Chow-Test nicht anwendbar.

Anpassen des Modells $y = X\beta + u$ an Beobachtungen

$t = 1, \dots, n-p$ gibt OLS-Schätzer b

Prognose $\hat{y}_f = X_f b$ für die Zeitpunkte $t = n-p+1, \dots, n$

Der Prognosetest prüft die Nullhypothese, dass das Modell auch im Prognosebereich gültig ist:

$H_0: y_f = X_f \beta + u$ *F-Statistik in Prognosefehlern e_f*

$$F = \frac{1}{e'e} e_f' \left[I_p + X_f (X'X)^{-1} X_f' \right] e_f \frac{n-p-k}{p}$$

Prognosetest: Berechnung von F

alternativ rechnet man einfacher

1. Anpassen des Modells an die $n-p$ Beobachtungen;
Summe der Fehlerquadrate S_D für „Datenbereich“ D
2. Anpassen des Modells an alle n Beobachtungen;
Summe der Fehlerquadrate S_{D+F} für „Daten- und Prognosebereich“ $D+F$
3. Einsetzen in F -Statistik

$$F = \frac{S_{D+F} - S_D}{S_D} \frac{n - p - k}{p}$$

F ist nach $F(p, n-p-k)$ verteilt.

Konsumfunktion, Forts.

Wir betrachten den gesamten Bereich D+F 1954-1999, unterteilen in den „Datenbereich“ D 1954-1991, und in den „Prognosebereich“ F 1992-1999.

OLS-Anpassung der Konsumfunktion für beide Perioden

1. 1954 bis 1991, D: $S_D = 3400.90$
2. 1954 bis 1999, D+F: $S_{D+F} = 6899.69$

Der p -Wert für die F-Statistik des Prognosetests ist 0.0006.

$$F = \frac{6899.69 - 3400.90}{3400.90} \frac{46 - 8 - 2}{8} = 4.63$$

Quandt-Andrews-Test

Dieser Test ermöglicht die Bestimmung des Zeitpunkts an dem der Strukturbruch stattfindet.

Nullhypothese:

Es liegt kein Strukturbruch in $t=1, \dots, T$ vor.

Alternative:

Es gibt einen Strukturbruch bei t^* (unbekannt).

Teststatistik ist das Maximum aller Chow-Tests für *allen* möglichen Zeitpunkten in $t=1 < t_1, \dots, t_2 < T$:

$$\max_t F_t$$

Die Verteilung der Teststatistik ist nicht-standard / tabelliert.
(Im Anfangs- u Endbereich kann nicht auf Strukturbruch getestet werden.)

Rekursive Residuen

Das wahre Modell sei $y = Xb + u$.

Rekursive Residuen sind transformierte 1-Schritt Prognosefehler:

$$w_t = \frac{Y_t - x'_t b_{t-1}}{\sqrt{1 + x'_t (X'_t X_t)^{-1} x_t}}, \quad t = k+1, \dots, n$$

b_{t-1} ist der OLS-Schätzer von b für die Periode $t-1$.

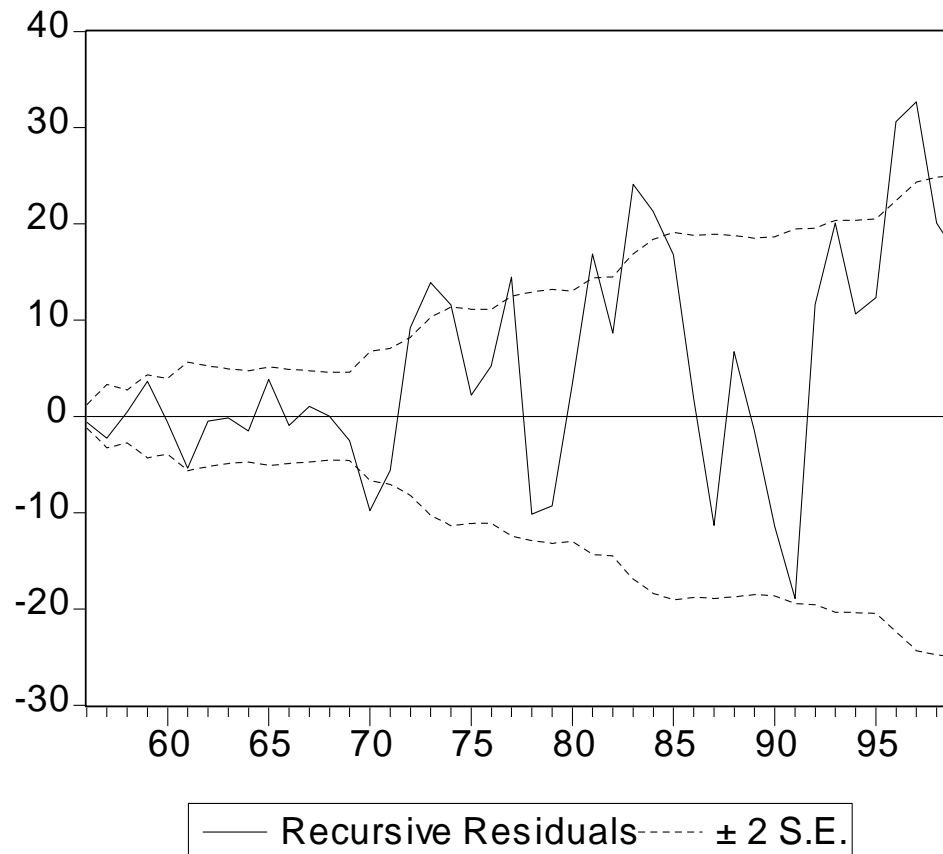
Der $(n-k)$ -Vektor w folgt (bei normalverteilten Störgrößen)

$$w \sim N(0, \sigma^2 I)$$

Die Elemente von w sind unabhängig. Daher eignen sie sich gut für die Konstruktion von Tests auf Strukturstabilität.

Konsumfunktion, Forts.

Plot der
rekursiven
Residuen
1956-99



Weitere Tests zur Strukturstabilität

Tests auf Basis der rekursiven Residuen sind:

- ❑ CUSUM Test (cumulative sum)
- ❑ MOSUM Test (moving sum)
- ❑ CUSUM-SQ Test (cumulative sum of squares)

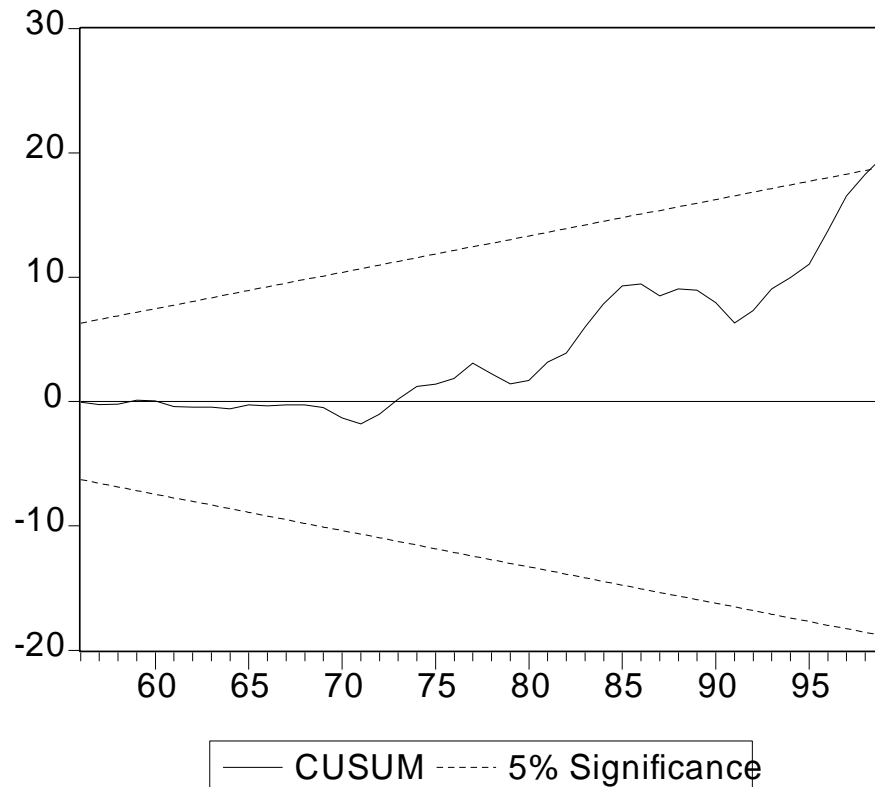
CUSUM Test:

$$W_t = \frac{1}{s} \sum_{r=k+1}^t w_r$$

s^2 ist die Residuenvarianz aus der Schätzung des Modells über die gesamte Periode.

Kritische Schranken gibt es von Brown(1975)

Konsumfunktion, Forts.



CUSUM Test:
W-Werte mit den
zugehörigen
kritischen Schranken