



2 Anwendungen und Probleme

Probleme empirischer Schätzungen

- *Fehlende Variablen und Multikollinearität*
- *Endogenität / Simultanität*
- *Strukturbruch*
- *Heteroskedastie*
- *Autokorrelation*

Für die Interpretation der Schätzergebnisse werden
in der Ökonometrie eine Reihe von **Annahmen** getroffen:

- *Korrekte Spezifikation des theoretischen Modells*
- *Exogenität der erklärenden Variablen*
- *Strukturkonstanz*
- *unabhängige, identisch verteilte Residuen (i.i.d.)*

Wenn diese Annahmen nicht erfüllt sind, kann es zu verzerrten Schätzungen
der Koeffizienten und der Standardfehler (der Konfidenzintervalle) kommen.

Deshalb ist es wichtig, diese Annahmen immer wieder zu überprüfen.

Fehlende Variablen und Multikollinearität

Fehlende Variablen führen dann zu verzerrten Schätzergebnissen, wenn diese fehlenden Variablen mit den im Modell aufgenommenen Variablen korreliert sind.

Da nahezu alle ökonomischen Variablen miteinander korreliert sind, führen fehlende Variablen fast immer zu verzerrten Schätzergebnissen:

“Wahres” Modell: $y_t = \beta_0 + \beta_1 \cdot x_{1,t} + \beta_2 \cdot x_{2,t} + \varepsilon_t$

Geschätztes Modell: $y_t = \beta_{0,g} + \beta_{1,g} \cdot x_{1,t} + \varepsilon_{g,t}$

Der KQ-Schätzer “entfernt” alle Korrelation der Residuen mit den x -Variablen, daher führt eine Korrelation von x_1 und x_2 zu einer verzerrten Schätzung von β_1 .

Test

Überprüfe, welche Variablen relevant sind

- *ökonomische Plausibilität*
- *statistische Signifikanz*

Beispiel

- *Cobb/Douglas Produktionsfunktion*

Weitere Beispiele

- *Wirtschaftswachstum ← Zinsen, Weltkonjunktur*
- *Konsum ← Einkommen, Vermögen*
- *...*

Endogenität / Simultanität

Wenn die erklärenden Variablen in einer Schätzung endogen sind, d.h. durch die zu erklärende Variable beeinflusst werden, führt der OLS-Schätzer zu verzerrten Ergebnissen für die Koeffizienten.

Beispiel

Der Konsum ist abhängig vom Einkommen, und die Einkommensidentität gilt.

$$C = c' \cdot Y + \varepsilon$$

$$Y = C + I$$

Dann gilt:

$$\begin{aligned} C &= c' \cdot Y + (C - c' \cdot Y) \\ &= c' \cdot Y + (Y - I) - c' \cdot Y \\ &= c' \cdot Y + (1 - c') \cdot Y - I \end{aligned}$$

- D.h. die "wahren" Residuen sind korreliert mit Y
- Für den OLS-Schätzer gilt jedoch, dass die geschätzten Residuen unkorreliert mit den erklärenden Variablen sind
- Daraus folgt, dass der OLS-Schätzer verzerrt ist

Hinweis → Aufzeichnen!

Für das Vorgehen in diesem Fall bieten sich 3 Alternativen an:

1. Schätzung der reduzierten Form

Das Modell kann nach den endogenen Variablen aufgelöst werden.

$$\begin{aligned} C &= c' \cdot (C + I) + \varepsilon \\ (1 - c') \cdot C &= c' \cdot I + \varepsilon \\ C &= \frac{c'}{1 - c'} \cdot I + \varepsilon' \end{aligned}$$

Wenn I exogen ist, kann diese Gleichung konsistent geschätzt werden. Die marginale Konsumneigung c' kann aus dem geschätzten Koeffizienten berechnet werden.

2. Instrumentschätzer

Eine Alternative ist die Schätzung mit einem Instrumentvariablenschätzer (IV)

- *Dafür wird die endogene erklärende Variable (also Y) in einer ersten Stufe auf exogene, mit Y korrelierte Variablen geschätzt*
- *Die geschätzten Werte dieser Schätzung (der FIT) werden dann in der zweiten Stufe anstelle der Variablen für die Schätzung der Modellgleichung verwendet (TSLS – Two Stage Least Squares)*

3. Schätzung anhand verzögerter Werte der erklärenden Variablen

Zumindest bei Zeitreihendaten kann y_t keinen direkten Einfluss auf x_{t-1} ausüben (Ursache \rightarrow Wirkung)

Endogenität/Simultanität ist ein sehr weit verbreitetes Problem:

- *Es gibt kaum ökonomische Variablen, die exogen sind, die also nicht von anderen ökonomischen Variablen beeinflusst werden*
- *Man sollte also immer auf Simultanität testen (anhand der Methoden oben → liefern sie andere Ergebnisse?)*
- *Andererseits führen die Methoden oben meist zu einem Verlust an Effizienz der Schätzung:
Der OLS-Schätzer weist definitionsgemäß die geringste Varianz auf
→ alle anderen Schätzer sind weniger effizient
Man muss also den Verlust an Effizienz gegen den Simultanitäts-Fehler abwägen*
- *Manchmal ist der Simultanitätsfehler gering:*
 - *wenn die Varianz der Schätzgleichung gering ist*
 - *wenn die Varianz der erklärenden Variable groß ist*
 - *wenn die Endogenität nicht sehr ausgeprägt ist*
 - *Beispiel: Konsumfunktion*
- *Weitere Beispiele: Schätzung von Lohn-, Preis- und Produktivitätsgleichungen*

$$\Delta \ln p = \alpha_0 + \alpha_1 \cdot \Delta \ln w + \alpha_2 \cdot \Delta \ln \pi_l + \varepsilon_p$$

$$\Delta \ln w = \beta_0 + \beta_1 \cdot \Delta \ln p + \beta_2 \cdot \Delta \ln \pi_l + \beta_3 \cdot UR + \varepsilon_w$$

$$\Delta \ln \pi_l = \gamma_0 + \gamma_1 \cdot \Delta \ln w + \gamma_2 \cdot \Delta \ln p + \gamma_3 \cdot \Delta \ln Q + \varepsilon_\pi$$

Hier ist die Endogenität wichtig → Identifikationsproblem

Strukturbruch

Strukturbruch bedeutet, dass sich die Struktur (die Koeffizienten, die Standardabweichung) des Modells über die Zeit (oder im Querschnitt z.B. mit der Unternehmensgröße) ändert

- *Manchmal gibt es einen a priori bekannten Zeitpunkt, an dem sich das Modell möglicherweise geändert hat (deutsche Vereinigung, Ölpreisschocks, Euro).*
- *Manchmal ändern sich die Koeffizienten langsam über die Zeit (Änderung der Konsumgewohnheiten).*

Die Schätzung über einen Strukturbruch hinweg impliziert eine Fehlspezifikation des Modells

Test auf Strukturbruch

- *Chow-Test:*

Schätzung des Modells

- *für die Zeit vor dem Strukturbruch (Ostdeutschland)*
- *nach dem Strukturbruch (Westdeutschland)*
- *für die gesamte Zeitperiode (Gesamtdeutschland)*
- *F-Test für die Gleichheit der Koeffizienten*

- *Ein alternativer Test ist der Chow Vorhersage Test*

- *schätze das Modell für die Periode bis zum Strukturbruch*
- *berechne auf der Basis dieser Schätzgleichung eine Vorhersage für die Zeitperiode nach dem Strukturbruch*
- *untersuche, ob die Abweichungen signifikant größer sind (F- oder χ^2 -Test)*
Dieser Test liefert auch sinnvolle Ergebnisse, wenn der Strukturbruch noch nicht lange zurückliegt.

– *Alternativer Test auf Strukturbruch: Rekursive Schätzungen*

Rekursive Schätzungen sind eine hilfreiche Möglichkeit, strukturelle Veränderungen in der Modellgleichung zu untersuchen:

- *Das Vorgehen*
 - *schätze die Modellgleichung für ein kurzes Sample am Anfang des Beobachtungszeitraums (es geht grundsätzlich auch anders herum)*
 - *berechne aus der Gleichung eine Vorhersage für die nächste Periode*
 - *bestimme den Vorhersagefehler (das sogenannte rekursive Residuum)*
 - *verlängere das Sample um eine Periode . . . usw.*

- *Ein formaler Test auf Strukturbruch ist der CuSum-Test (Cumulated Sum of recursive Residuals)*
 - *Dafür wird die kumulierte Summe der rekursiven Residuen berechnet*
 - *Wenn beispielsweise die rekursiven Residuen immer (sehr häufig) positiv sind, ist das ein Zeichen für einen Strukturbruch in der Gleichung*
 - *Für die Summe der Residuen kann ein Konfidenzband berechnet werden*
 - *Über- oder unterschreitet die Summe der Residuen das Konfidenzband, liegt ein signifikanter Strukturbruch vor*

- *Ein Test auf eine bestimmte Art der Heteroskedastie ist der CuSum²-Test*
Hierbei werden die Quadrate der rekursiven Residuen aufsummiert
Über- oder unterschreitet diese Summe das Konfidenzband,
dann liegt eine signifikante Änderung der Varianz der Residuen vor

- *Man kann sich auch die Koeffizienten der rekursiven Schätzungen ansehen*
Aus der Entwicklung dieser Koeffizienten können mögliche Ursachen für einen Strukturbruch näher eingegrenzt werden (z.B. die Saisonfigur hat sich geändert)

Beispiel I Schätzung einer Einkommensfunktion

Beispiel II Schätzung einer Konsumfunktion

Heteroskedastie

Heteroskedastie bedeutet, dass die Varianz der Residuen in der Stichprobe nicht konstant ist.

Beispiele:

Bei Zeitreihendaten:

Ansteigen der Varianz über die Zeit,
Anstieg der Varianz mit dem Einkommen.

Bei Querschnittsdaten:

Höhere absolute Varianz bei großen Einheiten
(Große Unternehmen, hohes Einkommen),
geringere relative Varianz (Veränderungsraten)
bei großen Einheiten (Risikostreuung).

Heteroskedastie führt nicht zu verzerrter Schätzung der Koeffizienten,
aber zu verzerrter Schätzung der Standardfehler (der Konfidenzintervalle).
Außerdem reduziert sie die Effizienz der Schätzung.

Test auf Heteroskedastie

Weit verbreitet und in EViews integriert ist der White-Test.

Testidee ist der Versuch der Erklärung der Streubreite der Residuen durch andere Variablen, also z.B.:

Erkläre die Residuenquadrate ε_t^2 durch die Niveaus und die Quadrate der erklärenden Variablen.

Signifikanztest für Heteroskedastie:

F-Test (Test, ob R^2 signifikant größer als Null)

Teststatistik:

$$F_{k-1, T-k} = \frac{R^2}{1 - R^2} \cdot \frac{T - k}{k - 1}$$

$T - k$: Zahl der Freiheitsgrade

$k - 1$: Zahl der Restriktionen, Zahl der geschätzten Koeffizienten $- 1$ (Konstante)

Bereinigung um Heteroskedastie

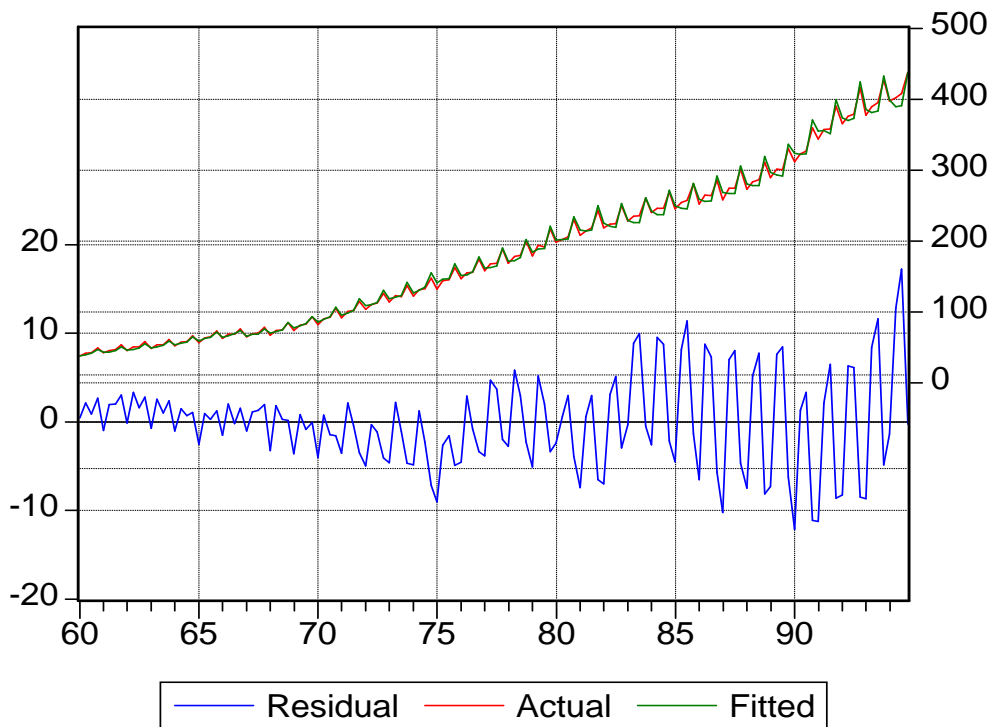
1. Teilen der gesamten Gleichung durch die Variable, die die Varianz beeinflusst (z.B. das Einkommen), Gewichtung
2. Bereinigung um Heteroskedastie mit Hilfe der geschätzten Werte der Test-Gleichung des White-Tests
3. Andere Spezifikation des theoretischen Modells, z.B. logarithmische Spezifikation

Beispiel

Schätzung einer Konsumfunktion in Abhängigkeit des verfügbaren Einkommens

```
=====
LS // Dependent Variable is KONSUMW
Sample: 1960:1 1994:4
Included observations: 140 after adjusting endpoints
=====
      Variable      Coefficient Std. Error T-Statistic Prob.
=====
           C          0.407818   0.866981    0.470389    0.6388
           YVW          0.872722   0.003517   248.1528    0.0000
=====
R-squared          0.997764      Mean dependent var 184.6276
Adjusted R-squared 0.997748      S.D. dependent var 111.6542
S.E. of regression 5.298797      Sum squared resid 3874.661
Log likelihood     -431.0914      F-statistic        61579.81
Durbin-Watson stat 1.654655      Prob(F-statistic) 0.000000
=====
```

Tatsächliche, geschätzte Werte, Residuen



```

=====
White Heteroskedasticity Test:
=====
F-statistic      36.51121    Probability    0.000000
Obs*R-squared   48.67642    Probability    0.000000
=====

```

```

Test Equation:
LS // Dependent Variable is RESID^2
Sample: 1960:1 1994:4
Included observations: 140

```

```

=====
Variable      Coefficient Std. Error T-Statistic Prob.
=====
C             -7.743614   9.459041  -0.818647  0.4144
YVW           0.131753   0.092857   1.418875  0.1582
YVW^2         0.000125   0.000187   0.669733  0.5042
=====
R-squared      0.347689    Mean dependent var 27.67615
Adjusted R-squared 0.338166    S.D. dependent var 41.75801
S.E. of regression 33.97147    Sum squared resid 158106.3
Log likelihood  -690.7080    F-statistic      36.51121
Durbin-Watson stat 2.064741    Prob(F-statistic) 0.000000
=====

```

Schätzung in Logarithmen

LS // Dependent Variable is LOG(KONSUMW)

Sample: 1960:1 1994:4

Included observations: 140 after adjusting endpoints

```
=====
Variable      Coefficient Std. Error T-Statistic Prob.
=====
C              -0.078648   0.017249  -4.559593   0.0000
LOG(YVW)       0.989736   0.003328  297.3719    0.0000
=====
R-squared      0.998442    Mean dependent var 5.004162
Adjusted R-squared 0.998431    S.D. dependent var 0.692430
S.E. of regression 0.027431    Sum squared resid 0.103841
Log likelihood  305.8063    F-statistic      88430.05
Durbin-Watson stat 1.689312    Prob(F-statistic) 0.000000
=====
```

Teilen durch das verfügbare Einkommen

LS // Dependent Variable is KONSUMW/YVW

Sample: 1960:1 1994:4

Included observations: 140 after adjusting endpoints

```
=====
Variable      Coefficient Std. Error T-Statistic Prob.
=====
C              0.865629   0.003364  257.3279    0.0000
1/YVW          1.542208   0.360907   4.273141    0.0000
=====
R-squared      0.116855    Mean dependent var 0.877250
Adjusted R-squared 0.110455    S.D. dependent var 0.024838
S.E. of regression 0.023426    Sum squared resid 0.075734
Log likelihood  327.9001    F-statistic      18.25973
Durbin-Watson stat 1.823971    Prob(F-statistic) 0.000036
=====
```

Autokorrelation

Autokorrelation bedeutet, dass die Residuen zum Zeitpunkt t korreliert sind mit den Residuen der Vorperioden.

Autokorrelation der Residuen führt nicht unbedingt (aber meistens) zu verzerrten Schätzungen der Koeffizienten, aber zu verzerrten Schätzungen der Standardabweichung.

Beispiele

– Positive Autokorrelation 1. Ordnung

- Persistenz der Residuen, ein großes positives Residuum heute führt zu einem großen positiven Residuum in der nächsten Periode
- Autokorrelation der Residuen ist ein Hinweis auf fehlende Variablen (fast alle ökonomischen Variablen weisen positive Autokorrelation 1. Ordnung auf)
- Fehlende Variablen implizieren eine Fehlspezifikation des Modells und führen fast immer zu verzerrten Schätzungen der Parameter (siehe oben)
- Irgendwo muss die Autokorrelation ja herkommen
→ Hinweis auf Fehlspezifikation

– Positive Autokorrelation 4. Ordnung

Fehler bei der Berücksichtigung der Saisonstruktur in der Gleichung

Exkurs: Saisonbereinigung

der Daten	konstante Saisonfaktoren, additiv oder multiplikativ, in Eviews: $seas \times xb$	gleitende Durchschnitte
in der Schätzung	Saisondummies, konstant, für sub-samples oder interagiert mit trend	AR(4) oder MA(4) Prozesse, die mitgeschätzt werden

Häufig stehen neben Originaldaten alternativ saisonbereinigte Daten zur Verfügung (X11, Berliner Verfahren oder ein ähnliches Saisonbereinigungsverfahren)

Zeitreihenstruktur vieler Quartalsdaten

$$x_t - \rho_4 \cdot x_{t-4} = \rho_1 \cdot (x_{t-1} - \rho_4 \cdot x_{t-5}) + \varepsilon_t$$

Vorsicht: Instabilität, wenn die autoregressive Wurzel ≥ 1 ,

→ Schätzung in Differenzen, aber das ist ein anderes Modell

Test auf Autokorrelation

1. Erkläre die Residuen anhand der Residuen der Vorperioden, t -Test für die Koeffizienten, F -Test der Gleichung

2. Breusch/Godfrey-Test

Erkläre die Residuen anhand der Residuen der Vorperioden

und der erklärenden Variablen des theoretischen Modells, F -Test der Gleichung

Bereinigung um Autokorrelation

a) – 1. Differenzen der Variablen

– 4. Differenzen der Variablen

– Quasi-Differenzen der Variablen: $x_t - \rho_i \cdot x_{t-i}$

→ in EViews Aufnahme von $AR(1)$, $AR(4)$ in die Schätzung

Achtung: Eine Schätzung in Differenzen ist ein anderes Modell

b) Schätzung eines erweiterten Modells:

– Aufnahme fehlender Variablen,

– Berücksichtigung einer verzögerten Anpassung im theoretischen Modell

Beispiel

Produktionsfunktion und technischer Fortschritt

(totale Faktorproduktivität oder Einsatz von Rohstoffen,
Energieeinsatz als stark autokorrelierte Variable)

```
=====
LS // Dependent Variable is LOG(YT)
Sample: 1960:1 1989:4
Included observations: 120
=====
```

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-4.289145	1.517816	-2.825866	0.0056
@SEAS(1)	-0.069824	0.005463	-12.78073	0.0000
@SEAS(2)	-0.034393	0.005543	-6.204790	0.0000
@SEAS(3)	0.019634	0.005877	3.340839	0.0011
T	0.012716	0.004160	3.056678	0.0028
T^2	-4.63E-05	1.31E-05	-3.540673	0.0006
LOG(K)	0.237311	0.227287	1.044103	0.2987
LOG(Q)	0.344839	0.055403	6.224201	0.0000
LOG(LT)	0.546957	0.163346	3.348464	0.0011
LOG(H)	0.691310	0.122896	5.625155	0.0000

```
=====
R-squared          0.995029      Mean dependent var 5.589478
Adjusted R-squared 0.994622      S.D. dependent var 0.259055
S.E. of regression 0.018997      Sum squared resid 0.039699
Log likelihood     310.5622      F-statistic        2446.441
Durbin-Watson stat 1.174896      Prob(F-statistic) 0.000000
=====
```

Zeitreihenanalyse der Residuen

```

=====
LS // Dependent Variable is RES
Sample: 1961:3 1989:4
Included observations: 114 after adjusting endpoints
=====

```

Variable	Coefficient	Std. Error	T-Statistic	Prob.
C	-0.000187	0.001086	-0.172092	0.8637
RES(-1)	0.343691	0.097442	3.527128	0.0006
RES(-2)	-0.052885	0.100400	-0.526742	0.5995
RES(-3)	0.062613	0.072477	0.863901	0.3896
RES(-4)	0.672380	0.069444	9.682353	0.0000
RES(-5)	-0.206988	0.095720	-2.162420	0.0328
RES(-6)	0.012003	0.093102	0.128925	0.8977

```

=====
R-squared                0.596084      Mean dependent var-0.000968
Adjusted R-squared      0.573435      S.D. dependent var 0.017661
S.E. of regression      0.011535      Sum squared resid 0.014237
Log likelihood          350.5652      F-statistic        26.31781
Durbin-Watson stat      1.973950      Prob(F-statistic) 0.000000
=====

```

```

=====
Breusch-Godfrey Serial Correlation LM Test:
=====
F-statistic          17.15475      Probability      0.000000
Obs*R-squared        68.83756      Probability      0.000000
=====

```

```

Test Equation:
LS // Dependent Variable is RESID

```

```

=====
Variable      Coefficient Std. Error T-Statistic Prob.
=====
C              -1.428387   1.249411  -1.143248   0.2556
@SEAS(1)      -0.002138   0.003776  -0.566282   0.5724
@SEAS(2)      -0.001382   0.004043  -0.341951   0.7331
@SEAS(3)      -0.000219   0.004469  -0.048988   0.9610
T              -0.004758   0.003074  -1.547661   0.1248
T^2            1.52E-05    9.68E-06   1.570078   0.1195
LOG(K)         0.256921   0.169386   1.516775   0.1324
LOG(Q)         0.084357   0.046904   1.798510   0.0751
LOG(LT)        -0.194611   0.126037  -1.544072   0.1257
LOG(H)         -0.015287   0.109821  -0.139198   0.8896
RESID(-1)      0.292982   0.105062   2.788646   0.0063
RESID(-2)     -0.025329   0.107444  -0.235738   0.8141
RESID(-3)     -0.025941   0.105906  -0.244939   0.8070
RESID(-4)      0.673722   0.104016   6.477121   0.0000
RESID(-5)     -0.183104   0.108944  -1.680707   0.0959
RESID(-6)      0.041569   0.112417   0.369776   0.7123
RESID(-7)      0.097294   0.108700   0.895063   0.3729
RESID(-8)      0.062488   0.105344   0.593181   0.5544
=====

```

```

=====
R-squared          0.573646      Mean dependent var 6.79E-16
Adjusted R-squared 0.502587      S.D. dependent var 0.018265
S.E. of regression 0.012882      Sum squared resid 0.016926
Log likelihood     361.7114      F-statistic         8.072824
Durbin-Watson stat 1.958904      Prob(F-statistic)  0.000000
=====

```