Übung: Autokorrelation

Analyse des Nahrungsmittelkonsums in der Schweiz

Wir untersuchen die Bestimmungsgründe des Nahrungsmittelkonsums in der Schweiz. Die Daten umfassen den Zeitraum 1960-2008 und stammen aus der Volkswirtschaftlichen Gesamtrechnung der Schweiz. Sie sind im File Konsum.xls bereitgestellt, welches Sie in den vorherigen Übungen bereits in ein EViews-Workfile umgewandelt haben.

Versuchen Sie, die folgenden Schritte in EVIEWS nachzuvollziehen. Am Schluss der Übung sind dann ohne weitere Anleitungen drei Fragen zu beantworten.

Zusätzliche verwendete Variablen:

KNAHRN Konsum Nahrungsmittel nominell
PKNAHR Konsumentenpreise Nahrungsmittel

Die zu untersuchende Grundhypothese lautet, dass der Nahrungsmittelkonsum der Haushalte von ihrem Einkommen und dem relativen Nahrungsmittelpreis abhängt. Die Analyse dieses Zusammenhangs wird in realen pro-Kopf-Grössen durchgeführt.

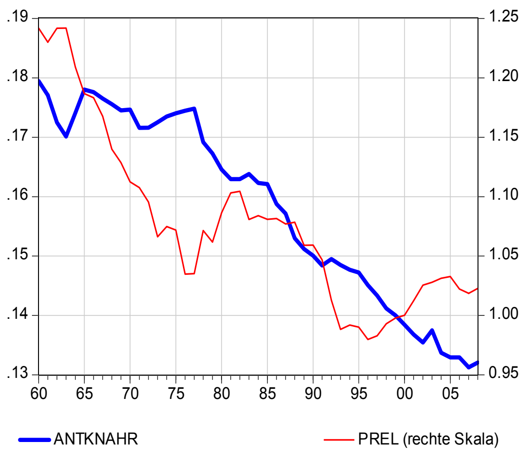
Der reale Nahrungsmittelkonsum pro Kopf und der relative Nahrungsmittelpreis sind zunächst mit GENR zu berechnen:

KNAHRK = KNAHRN/PKNAHR/WOBEV
PREL = PKNAHR/PKONS

Das reale pro-Kopf Einkommen ist bereits berechnet worden:

EINRK = EINN/PKONS/WOBEV

In einem ersten Schritt ist es sinnvoll, sich die verschiedenen Variablen grafisch anzusehen. Dadurch erhält man ein "Gefühl" für die Daten und es würden auch grobe Datenfehler zum Vorschein kommen. Man kann z.B. den Anteil des Nahrungsmittelkonsums am Gesamtkonsum ausrechnen, indem man im WORKFILE die Option GENR wählt und dort ANTKNAHR = KNAHRK/KONSRK eingibt. Die beiden Variablen ANTKNAHR und PREL lassen sich grafisch darstellen, indem man sie im WORKFILE markiert, dann SHOW anklickt und dort die Option GRAPH wählt. Die Grafik kann anschliessend angepasst werden, indem man sie anklickt und dann unter den eingeblendeten Optionen die passenden auswählt.



Der Anteil des Nahrungskonsums am Gesamtkonsum war von 1960 bis 1978 mit Werten zwischen 17% und 18% ziemlich konstant. Danach hat er sich bis 2008 auf noch rund 13% zurückgebildet.

Die Nahrungsmittel haben sich über den ganzen Zeitraum betrachtet deutlich verbilligt. Diese Tendenz war vor allem von 1960 bis 1978 ausgeprägt. Möglicherweise hat dies zur Konstanz des Anteils des Nahrungsmittelkonsums am Gesamtkonsum in diesem Zeitabschnitt beigetragen.

Wir schätzen nun in einem ersten Schritt die einfache Regression:

$$ln(KNAHRK_t) = \alpha + \beta \, ln(EINRK_t) + u_t$$

Da wir die Regression in natürlichen Logarithmen rechnen wollen, spezifizieren wir die Gleichung als

LOG(KNARK) C LOG(EINRK)

oder LOG(KNAHRK) = C(1) + C(2)*LOG(EINRK)

Gleichung 1:

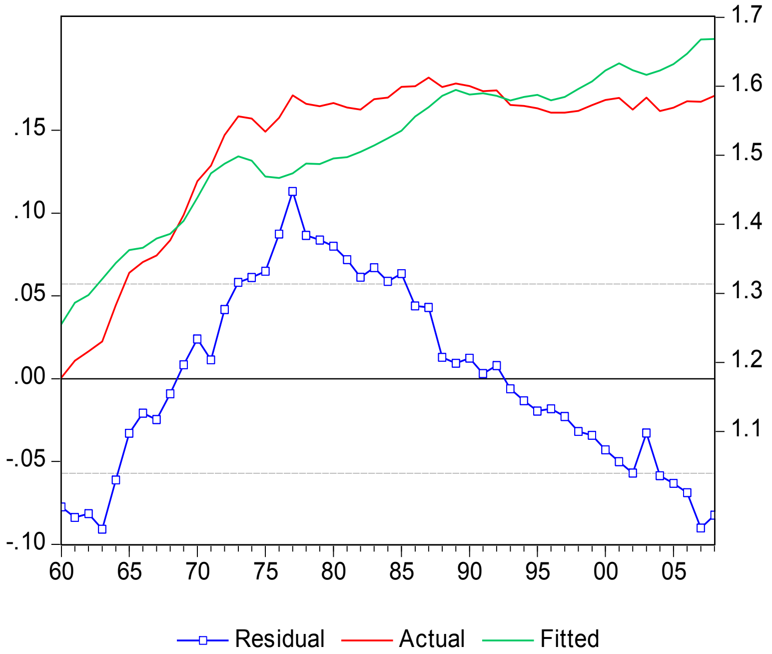
Sample: 1960 2008
Included observations: 49
LOG(KNAHRK) = C(1)+C(2)*LOG(EINRK)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.214566	0.133229	-1.610511	0.1140
C(2)	0.503848	0.038763	12.99806	0.0000

R-squared	0.782357	Mean dependent var	1.513891
Adjusted R-squared	0.777726	S.D. dependent var	0.121230
S.E. of regression	0.057155	Akaike info criterion	-2.846141
Sum squared resid	0.153534	Schwarz criterion	-2.768924
Log likelihood	71.73045	Hannan-Quinn criter.	-2.816845
F-statistic	168.9495	Durbin-Watson stat	0.066082
Prob(F-statistic)	0.000000		

Die geschätzte Einkommenselastizität des Nahrungsmittelkonsums beträgt 0.504. Die Residuen der Gleichung sind aber extrem autokorreliert (Durbin-Watson-Statistik = 0.066), wie auch die entsprechende Grafik zeigt. Von zufallsverteilten Störein-flüssen kann somit keine Rede sein.

Fit und Residuen von Gleichung 1



Berechnen Sie zum Vergleich die Einkommenselastizität der gesamten Konsumnachfrage, indem Sie

LOG(KONSRK) = C(1) + C(2)*LOG(EINRK)

eingeben! Es zeigt sich, dass die Einkommenselastizität mit einem Wert von 0.926 deutlich grösser ist. Die vergleichsweise kleine Einkommenselastizität des Nahrungsmittelkonsums entspricht seinem im Zeitverlauf sinkenden Anteil am Gesamtkonsum.

Gleichung 1 ist trotz relativ hohem R-Quadrat und hochsignifikanter Schätzung der Einkommenselastizität nicht befriedigend, weil die tiefe DW-Statistik eine ausgeprägte Autokorrelation der Residuen anzeigt und die Annahme eines zufällig normalverteilten Störeinflusses somit klar verletzt ist. Die Gleichung ist fehlspezifiziert. Die Autokorrelation der Residuen kann verschieden gedeutet werden:

A) In der Gleichung wurden wichtige Einflussfaktoren vernachlässigt.

B) Der Nahrungsmittelkonsum passt sich nur verzögert an das Einkommen an.

C) Der Störterm des "wahren Modells" ist autokorreliert.

In den Fällen *A* und *B* sollte eine entsprechende Erweiterung der Gleichung die Autokorrelation zum Verschwinden bringen. Wenn keine solchen ***ökonomisch begründbaren*** Erweiterungen gefunden werden können, bleibt nur die weniger befriedigende, weil rein ***statistisch-technische*** Lösung gemäss *C*. In diesem Fall ist die Gleichung so zu transformieren, dass für den resultierenden Störterm die Annahme der Unkorreliertheit zulässig ist.

Diese Übung dient zur Illustration von *A*, *B* und *C*.

Ansatz *A:* Einbezug weiterer erklärender Variablen

Als bislang vernachlässigte Einflussfaktoren (*A*) kommen in Betracht:

- der relative Preis der Nahrungsmittel PREL = PKNAHR/PKONS
- die Arbeitslosigkeit ARBLOS
- ein Zeittrend (eine Variable, die mit Werten 1,2,3,4,… kontinuierlich über die Zeit zunimmt und somit eine trendmässige Zu- oder Abnahme der Linkhand-Variablen auffangen kann.

Gleichung 2:

Sample: 1960 2008
Included observations: 49
LOG(KNAHRK) = C(1)+C(2)*LOG(EINRK)+C(3)*LOG(PREL)+C(4)*ARBLOS +C(5)*@TREND

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.493998	0.424305	-1.164251	0.2506
C(2)	0.650339	0.130114	4.998226	0.0000
C(3)	-1.031008	0.208546	-4.943786	0.0000
C(4)	-0.000633	0.000190	-3.329519	0.0018
C(5)	-0.004836	0.001868	-2.589061	0.0130

R-squared	0.936920	Mean dependent var	1.513891
Adjusted R-squared	0.931186	S.D. dependent var	0.121230
S.E. of regression	0.031802	Akaike info criterion	-3.962149
Sum squared resid	0.044499	Schwarz criterion	-3.769106
Log likelihood	102.0727	Hannan-Quinn criter.	-3.888909
F-statistic	163.3823	Durbin-Watson stat	0.377919
Prob(F-statistic)	0.000000		

Die zusätzlichen Variablen PREL und ARBLOS haben den theoretisch erwarteten negativen Einfluss. Der Zeittrend @TREND wirkt sich erwartungsgemäss ebenfalls negativ aus. Das Problem autokorrelierter Residuen hat sich aber durch den Einbezug der drei Variablen nur wenig entschärft. Noch immer zeigt die Durbin-Watson-Statistik ausgeprägte Autokorrelation an. In der hier untersuchten Form hat sich Ansatz *A* zur Verminderung der Autokorrelation somit nicht bewährt.

Ansatz *B:* "Partial adjustment"-Modell

Um die Möglichkeit einer verzögerten Anpassung zu berücksichtigen, führen wir die abhängige Linkhand-Variable um eine Periode verzögert als erklärende Variable auf der rechten Gleichungsseite ein. Die diesem Ansatz zugrundeliegende Vorstellung ist die folgende:

(1) $y_t^* = \alpha + \beta_1 x_{1_t} + \beta_2 x_{2_t} + \ldots + \varepsilon_t$ langfristige Beziehung

(2) $y_t = \lambda y_t^* + (1-\lambda) y_{t-1}$ mit $0 \le \lambda \le 1$ kurzfristige Anpassung

Substitution von (1) in (2) ergibt:

(3) $y_t = \lambda\alpha + \lambda\beta_1 x_{1_t} + \lambda\beta_2 x_{2_t} + \ldots + (1-\lambda) y_{t-1} + u_t$

Man kann eine solche Gleichung in EVIEWS in verschiedener Form eingeben (vgl. Multiple lineare Regression: Übung 1), z.B.

LOG(KNAHRK) = C(1)+C(2)*LOG(EINRK)+C(3)*LOG(PREL)
+C(4)*LOG(ARBLOS)+C(5)*LOG(KNAHRK(-1))

C(5) hat dann die Interpretation C(5) = $1-\lambda$, d.h. man kann aus C(5) den Anpassungsparameter λ ableiten. C(2) hat die Interpretation C(2) = $\lambda\beta_1$, d.h. man kann aus C(2) und C(5) β_1 ableiten, usw.

Gleichung 3:

Sample (adjusted): 1961 2008
Included observations: 48 after adjustments
LOG(KNAHRK) = C(1)+C(2)*LOG(EINRK)+C(3)*LOG(PREL)+C(4)*ARBLOS +C(5)*@TREND +C(6)*LOG(KNAHRK(-1))

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	-0.090775	0.177786	-0.510588	0.6123
C(2)	0.153644	0.062604	2.454214	0.0183
C(3)	-0.224109	0.096811	-2.314910	0.0256
C(4)	-0.000150	8.12E-05	-1.848319	0.0716
C(5)	-0.001556	0.000793	-1.961300	0.0565
C(6)	0.756722	0.048304	15.66591	0.0000

R-squared	0.988987	Mean dependent var	1.520885
Adjusted R-squared	0.987676	S.D. dependent var	0.112078
S.E. of regression	0.012442	Akaike info criterion	-5.818957
Sum squared resid	0.006502	Schwarz criterion	-5.585057
Log likelihood	145.6550	Hannan-Quinn criter.	-5.730566
F-statistic	754.3165	Durbin-Watson stat	1.770039
Prob(F-statistic)	0.000000		

Die Berücksichtigung einer dynamischen Anpassungsverzögerung in *Gleichung 3* hat das Autokorrelationsproblem schlagartig entschärft. Die DW-Statistik erhöht sich auf 1.770. Die drei Kurzfrist-Elastizitäten C(2), C(3) und C(4) haben das erwartete Vorzeichen (positiver Einkommenseinfluss, negative Einflüsse von relativem Preis und Arbeitslosigkeit). Die Arbeitslosigkeit sowie auch der Zeittrend sind aber auf dem 5%-Niveau knapp nicht signifikant. Die Anpassungsgeschwindigkeit ist mit 1-C(6) = λ = 0.243 eher gering. Gemäss dieser Schätzung reagiert der Nahrungsmittelkonsum kurzfristig nur träge auf Veränderungen in den verschiedenen Erklärungsfaktoren. Die langfristige Einkommenselastizität ist durch C(2)/(1-C(6)) = 0.6316 gegeben. Die langfristige Preiselastizität liegt nahe bei -1.

Spezifikationstests

Autokorrelation: Durbin-Watson und Durbin-h Test

Im Folgenden soll *Gleichung 3* weiter untersucht werden. Zunächst wollen wir die Zulässigkeit der bei der Schätzung getroffenen Annahme testen, dass der Störterm gemäss $u_t \sim IN(0,\sigma^2)$ verteilt ist (unabhängig normalverteilt, d.h. insbesondere nicht autokorreliert).

Die ***DW-Statistik*** von 1.77 erlaubt keine eindeutigen Schlussfolgerungen. Für k' = 5 und n = 48 sind die kritischen Werte d_L = 1.32 und d_U = 1.77. Der berechnete DW-Wert liegt also genau am oberen Rand des Bereichs, wo der Test nicht schlüssig ist.

Dabei ist aber Folgendes zu beachten:

Durbin-h-Test

Wenn die abhängige Variable verzögert als erklärende Variable auf der rechten Gleichungsseite erscheint, ist der DW-Test gegen den Wert 2 verzerrt, weil ein Teil der Autokorrelation fälschlicherweise durch die verzögerte abhängige Variable aufgefangen wird. Somit tendiert der DW-Test dazu, die Null-Hypothese eines unkorrelierten Störterms nicht zu verwerfen, obwohl ein Autokorrelationsproblem vorhanden ist. In solchen Situationen muss der Durbin-h-Test angewendet werden.

$$\text{Durbin-h} = \hat{\rho}\sqrt{\frac{n}{1 - n\, s_\beta^2}}$$

Diese Teststatistik ist ***standard-normalverteilt***.

$\hat{\rho}$ ist der Autokorrelationskoeffizient erster Ordnung der berechneten Residuen, n ist der Stichprobenumfang und s_β^2 ist der quadrierte Standardfehler des Koeffizienten vor der verzögerten abhängigen Variablen. $\hat{\rho}$ kann approximativ als

$$\hat{\rho} = 1 - \frac{DW}{2}$$ berechnet werden.

Exakter erhält man den Autokorrelationskoeffizienten nach Schätzung der Gleichung über VIEWS / RESIDUAL TESTS / CORRELOGRAM:

	AC	PAC	Q-Stat	Prob
1	0.109	0.109	0.6067	0.436
2	-0.067	-0.080	0.8432	0.656
3	-0.078	-0.063	1.1684	0.761
4	-0.008	0.003	1.1719	0.883
5	-0.001	-0.010	1.1720	0.948
6	-0.086	-0.092	1.5934	0.953
7	0.053	0.074	1.7575	0.972
8	-0.008	-0.036	1.7611	0.987

Der Autokorrelationskoeffizient erster Ordnung beträgt 0.109.

(Mit Prob = 0.436 ist die Autokorrelation erster Ordnung auf dem 5%-Niveau nicht signifikant. Diese Aussage bezieht sich auf die Ljung-Box Q-Statistik, auf die wir in dieser Vorlesung nicht eingehen.)

Für die *Durbin-h*-Statistik ergibt sich somit:

$$h = 0.109\sqrt{\frac{48}{1-48\cdot 0.048304^2}} = 0.801$$

Die kritischen Werte gemäss Standard-Normalverteilung für einen einseitigen Test sind 1.645 auf dem 5%-Niveau und 2.326 auf dem 1%-Niveau, also grösser als der berechnete Wert. Die bei der Schätzung der Gleichung unterstellt Annahme eines nicht autokorrelierten Störterms ist folglich nicht abzulehnen, d.h. die Gleichung ist in dieser Hinsicht in Ordnung.

Stabilität

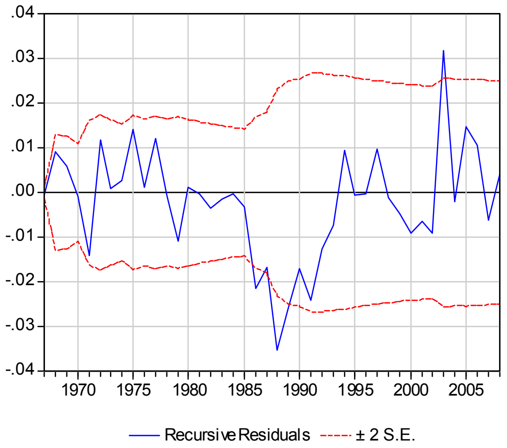
Eine andere Frage ist, ob die *Parameter* der Gleichung im Zeitverlauf stabil sind (was bei der Schätzung ebenfalls unterstellt ist) oder ob z.B. Verhaltensänderungen der Haushalte stattgefunden haben. Dieser Frage kann nachgegangen werden, indem man die Gleichung über verschiedene Perioden schätzt. Im ***Chow-Breakpoint-Test*** unterteilt man die Stichprobe (1961-2008) in zwei Teilstichproben (z.B. 1961-1989, 1990-2008) und vergleicht die Summe der beiden Residuenquadratsummen in einem F-Test mit der Residuenquadratsumme einer Schätzung über die gesamte Periode (1961-2008). Ein signifikanter F-Wert deutet auf nicht-stabile Parameter hin.

Chow Breakpoint Test: 1989
Null Hypothesis: No breaks at specified breakpoints
Equation Sample: 1961 2008
F-statistic = 5.2270 Prob. F(6,36) = 0.00059

Bei einer Unterteilung der Stichprobe in 1989/1990 ergibt sich ein F-Wert von 5.227, womit die Nullhypothese stabiler Parameter klar abgelehnt wird (Prob = 0.00059).

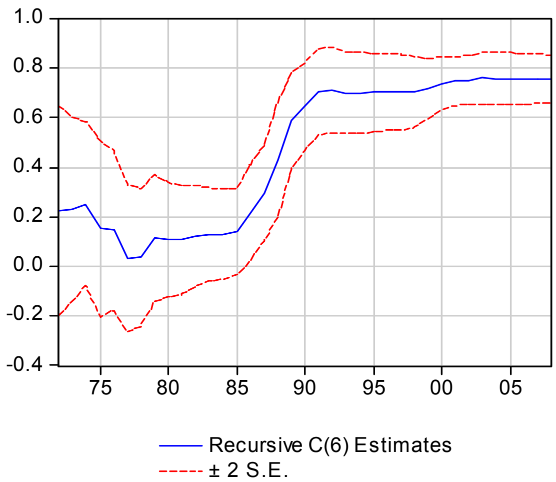
Genauer lässt sich das Problem nicht-stabiler Parameter anhand einer ***rekursiven Schätzung*** der Gleichung analysieren. Darunter versteht man eine schrittweise Vergrösserung der Stichprobe um je eine Beobachtung. Die ***rekursiven Residuen*** sind folgendermassen berechnet. Man schätzt die Gleichung mit Daten bis zur Periode t-1 und macht dann auf Basis dieser Gleichung eine "Prognose" für die abhängige Variable in Periode *t*. Die Abweichung der Beobachtung der von dieser Prognose ist das rekursive Residuum der Periode *t*. Dann fügt man einen Datenpunkt zur Schätzung hinzu und berechnet für die nachfolgende Periode das nächste rekursive Residuum. Rekursive Residuen, die ausserhalb des Vertrauensbereiches (± 2 Standardfehler der Gleichung) liegen, deuten auf nicht-stabile Parameter hin (Verhaltensänderungen, "Strukturbruch").

Für Gleichung 3 erhält man über VIEW / STABILITY TESTS / RECURSIVE ESTIMATES / RECURSIVE RESIDUALS die folgende Grafik:



Die Gleichung erscheint hinsichtlich der Parameterstabilität problematisch. Sie tendiert dazu, den Nahrungsmittelkonsum gegen Ende der 1980er Jahre zu überschätzen und im Jahre 2003 zu unterschätzen. Dies kommt in den rekursiven Residuen besser zum Ausdruck als in den üblichen, über den ganzen Schätzzeitraum berechneten Residuen, weil diese auf Parameterschätzungen basieren, die die das "mittlere" Verhalten reflektieren.

Die ***rekursiv geschätzten Parameter*** lassen genauer erkennen, bei welchen Einflussfaktoren das Problem der Nicht-Stabilität liegt. Man erhält die rekursiven Parameterschätzungen über VIEW / STABILITY TESTS / RECURSIVE ESTIMATES / RECURSIVE COEFFICIENTS. Für die Parameter C(2) (kurzfristige Einkommenselastizität) und C(6) (1 – Anpassungsgeschwindigkeit) ergibt sich:



Gegen Ende der 1980er und zu Beginn der 1990er Jahre zeigt sich eine markante Verhaltensänderung: Die Einkommenselastizität C(2) geht bei Einbezug weiterer Beobachtungspunkte zurück und die Nahrungsmittelnachfrage reagiert kurzfristig träger auf die Bestimmungsfaktoren (C(6) steigt stark an).

Hypothesentests

Hypothesentests (d.h. Test auf Zulässigkeit bestimmter Parameterrestriktionen) scheinen aufgrund der Ergebnisse der vorstehenden Spezifikationstests problematisch. Zuerst muss ein Modell gefunden werden, das die diversen Spezifikationstests besteht. Die Grundannahme eines unabhängig normalverteilten Störterms scheint in Gleichung 3 zwar nicht verletzt zu sein. Hingegen zeigt die rekursive Schätzung, dass verschiedene Parameter über die Zeit nicht konstant sind. Somit ist die Gleichung als fehlspezifiziert zu betrachten, denn bei der Schätzung wurden konstante Parameter vorausgesetzt. Zu Demonstrationszwecken wird trotz dieser Probleme auf Basis von Gleichung 3 ein Hypothesentests durchgeführt.

Hypothese: Der Nahrungsmittelkonsum ist nicht vom Einkommen abhängig

Für diese Hypothese, die der Parameterrestriktion C(2) = 0 entspricht, liefert bereits der Standard-Output von EVIEWS die benötigten Informationen. Die Hypothese wird mit einem t-Wert (beobachtetem Signifikanzniveau) von 2.454 (0.0183) auf dem 5%-Niveau abgelehnt. Zum gleichen Resultat führt ein sogenannter ***Wald-Test***. Man wählt dazu unter VIEWS die Option COEFFICIENT TESTS und dann WALD-COEFFICIENT RESTRICTIONS und gibt die Restriktion C(2)=0 ein.

Wald Test:
Null Hypothesis: C(2)=0
F-statistic = 6.023165 (1, 42) Probability = 0.0183

Das Ergebnis ist identisch. Der F-Wert ist mit einem beobachteten Signifikanzniveau von 0.0183 auf dem 5%-Niveau signifikant.

Im Falle einer linearen Gleichung entspricht der Wald-Test dem üblichen F-Test. Dies lässt sich wie folgt nachvollziehen: Zunächst notiert man sich die Residuenquadratsumme der "unrestringierten" Gleichung (URSS). Dann schätzt man die "restringierte" Gleichung unter Ausschluss des Terms C(2)*LOG(EINRK) und erhält so die (grössere) restringierte Residuenquadratsumme (RRSS). Eingesetzt in die Formel

$$F = \frac{(RRSS - URSS)/r}{URSS/(n-k-1)} \qquad \text{mit } r = 1 \text{ und } n - k - 1 = 42$$

erhält man F = 6.023. Der kritische auf dem 5%-Niveau für 1 und 42 Freiheitsgrade ist 4.07.

Mit dem Wald- bzw. F-Test kann man auch verbundene Hypothesen (z.B. C(4) = 0 und C(5) = 0, oder C(3) = -C(2)) testen.

Ansatz *C*: Modell mit autokorreliertem Störprozess (Maddala, S. 234 ff.)

Die Erweiterung von *Gleichung 1* durch Einbezug des relativen Preises, der Arbeitslosigkeit und eines Zeittrend als zusätzliche Erklärungsfaktoren in *Gleichung 2* hat das Autokorrelationsproblem nicht behoben. Hingegen erwies sich die Berücksichtigung einer Anpassungsverzögerung in Gleichung 3 hinsichtlich des Autokorrelationsproblems als erfolgreich. Gleichung 3 kann aber trotzdem nicht überzeugen, weil ihre Parameter über die Zeit nicht stabil sind. Als weitere Alternative betrachten wir nun Ansatz *C*, der das Autokorrelationsproblem durch eine Transformation der Ausgangsgleichung zu lösen versucht.

Wenn in einer Gleichung (wie z.B. Gleichung 2) die Annahme $u_t \sim IN(0,\sigma^2)$ verletzt ist, kann möglicherweise davon ausgegangen werden, dass der Störterm einem autoregressiven Prozess erster Ordnung folgt. Wenn diese Annahme zutrifft, kann die Ausgangsgleichung in einfacher Weise so transformiert werden, dass das Autokorrelationsproblem verschwindet.

Gleichungstransformation nach Cochrane-Orcutt

In der Gleichung

(1) $y_t = \alpha + \beta_1 x_{1_t} + \beta_2 x_{2_t} + \ldots + u_t$

wird für den Störterm unterstellt:

(2) $u_t = \rho u_{t-1} + \varepsilon_t$ mit $-1 < \rho < 1$ und $\varepsilon_t \sim IN(0,\sigma^2)$

Gleichung (1) lässt sich in einfacher Weise so transformieren, dass eine Gleichung mit unkorreliertem Störterm entsteht. Schreibt man (1) um eine Periode verzögert und multipliziert mit ρ, so ergibt sich:

(3) $\rho y_{t-1} = \rho\alpha + \rho\beta_1 x_{1_{t-1}} + \rho\beta_2 x_{2_{t-1}} + \ldots + \rho u_{t-1}$

Subtrahiert man nun (3) von (1), so erhält man:

(4) $y_t - \rho y_{t-1} = (1-\rho)\alpha + \beta_1(x_{1_t} - \rho x_{1_{t-1}}) + \beta_2(x_{2_t} - \rho x_{2_{t-1}}) + \ldots + u_t - \rho u_{t-1}$

Gemäss Annahme (2) gilt $u_t - \rho u_{t-1} = \varepsilon_t$, d.h. Gleichung (4) hat einen Störterm mit der gewünschten Eigenschaft $\varepsilon_t \sim IN(0,\sigma^2)$. Man kann (4) auch wie folgt schreiben:

(4') $y_t - \rho y_{t-1} = (1-\rho)\alpha + \beta_1(x_{1_t} - \rho x_{1_{t-1}}) + \beta_2(x_{2_t} - \rho x_{2_{t-1}}) + \ldots + \varepsilon_t$

(4") $y_t = (1-\rho)\alpha + \rho y_{t-1} + \beta_1(x_{1_t} - \rho x_{1_{t-1}}) + \beta_2(x_{2_t} - \rho x_{2_{t-1}}) + \ldots + \varepsilon_t$

Diese Gleichungen sind nicht-linear in den Parametern $\alpha, \beta_1, \beta_2, \ldots$ und ρ. Zur Schätzung derartiger Gleichungen werden in der Literatur verschiedene Verfahren vorgeschlagen. Wir betrachten hier zwei dieser Verfahren:

1. Umgehen der Nicht-Linearität in einer ***zweistufigen Schätzung***: Man schätzt zunächst (1) und verwendet die Residuen $\hat{u}_t$ zur Berechnung des Autokorrelationskoeffizienten $\hat{\rho}$. Danach setzt man $\hat{\rho}$ in (4') ein, was auf eine lineare Regression in den transformierten Variablen $y_t - \hat{\rho} y_{t-1}$, $x_{1_t} - \hat{\rho} x_{1_{t-1}}$ und $x_{2_t} - \hat{\rho} x_{2_{t-1}}$ hinausläuft.

2. Schätzung mit ***"Nonlinear Least Squares"***: Man schätzt die Parameter $\alpha, \beta_1, \beta_2, \ldots$ in (4") simultan mit dem Parameter ρ, indem man die Residuenquadratsumme dieser (nicht-linearen) Gleichung in bezug auf alle Parameter minimiert. Dieses Verfahren ist in EVIEWS standardmässig implementiert.

Die beiden Schätzverfahren werden nun ausgehend von Gleichung 2 demonstriert.

1. Zweistufige Schätzung

Erste Stufe: Berechnung von $\hat{\rho}$

Wir bestimmen den Autokorrelationskoeffizienten anhand der Residuen von Gleichung 2, entweder über VIEWS / RESIDUALS TESTS / CORRELOGRAM, oder indem wir die Formel $\hat{\rho}$ = 1 - DW/2 = 1- 0.3779/2 = 0.811 anwenden.

Zweite Stufe: Schätzung der transformierten Gleichung (4')

Sample (adjusted): 1961 2008
Included observations: 48 after adjustments
LOG(KNAHRK)-0.811*LOG(KNAHRK(-1))= C(1)*(1-0.811)
+C(2)*(LOG(EINRK)-0.811*LOG(EINRK(-1)))
+C(3)*(LOG(PREL)-0.811*LOG(PREL(-1)))
+C(4)*(ARBLOS-0.811*ARBLOS(-1))
+C(5)*(@TREND-0.811*(@TREND-1))

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	0.498852	0.368356	1.354264	0.1827
C(2)	0.329783	0.114826	2.872038	0.0063
C(3)	-0.616989	0.156279	-3.948007	0.0003
C(4)	-0.000195	0.000126	-1.552280	0.1279
C(5)	-0.001964	0.001700	-1.155272	0.2544

R-squared	0.526106	Mean dependent var	0.294341
Adjusted R-squared	0.482023	S.D. dependent var	0.018718
S.E. of regression	0.013472	Akaike info criterion	-5.678111
Sum squared resid	0.007804	Schwarz criterion	-5.483194
Log likelihood	141.2747	Hannan-Quinn criter.	-5.604452
F-statistic	11.93440	Durbin-Watson stat	1.435645
Prob(F-statistic)	0.000001		

2. Nonlinear Least Squares

Schätzt man Gleichung 4" mit "Nonlinear Least Squares", so erhält man:

Sample (adjusted): 1961 2008
Included observations: 48 after adjustments
Convergence achieved after 6 iterations
LOG(KNAHRK)=C(6)*LOG(KNAHRK(-1))+C(1)*(1-C(6))
+C(2)*(LOG(EINRK)-C(6)*LOG(EINRK(-1)))
+C(3)*(LOG(PREL)-C(6)*LOG(PREL(-1)))
+C(4)*(ARBLOS-C(6)*ARBLOS(-1))
+C(5)*(@TREND-C(6)*(@TREND-1))

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C(6)	0.941209	0.031490	29.88869	0.0000
C(1)	0.946769	0.395850	2.391737	0.0213
C(2)	0.261546	0.097290	2.688324	0.0103
C(3)	-0.479907	0.137360	-3.493792	0.0011
C(4)	-0.000152	0.000106	-1.431588	0.1597
C(5)	-0.006077	0.004611	-1.318056	0.1946

R-squared	0.990331	Mean dependent var	1.520885
Adjusted R-squared	0.989180	S.D. dependent var	0.112078
S.E. of regression	0.011658	Akaike info criterion	-5.949131
Sum squared resid	0.005708	Schwarz criterion	-5.715231
Log likelihood	148.7791	Hannan-Quinn criter.	-5.860740
Durbin-Watson stat	2.112065		

Der simultan geschätzte Autokorrelationsparameter ist C(6) = $\hat{\rho}$ = 0.941, gegenüber $\hat{\rho}$ = 0.811 in der 2-stufigen Schätzung. Mit $\hat{\rho}$ = 0.941 wird die Autokorrelation der Ausgangsgleichung etwas "überkorrigiert" (DW = 2.11), während in der 2-stufigen Schätzung eine gewisse Autokorrelation verbleibt (DW = 1.44). In der simultanen Schätzung ist die Einkommenselastizität C(2) = 0.262 und für die Preiselastizität ergibt sich C(3) = -0.480.

Man kann sich in EViews etwas Arbeit ersparen, wenn man die Gleichung in nicht-expliziter Form eingibt und am Schluss einfach AR(1) hinzufügt:

LOG(KNAHRK) C LOG(EINRK) LOG(PREL) ARBLOS @TREND AR(1)

Sample (adjusted): 1961 2008
Included observations: 48 after adjustments
Convergence achieved after 8 iterations

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.946768	0.395859	2.391682	0.0213
LOG(EINRK)	0.261545	0.097289	2.688323	0.0103
LOG(PREL)	-0.479907	0.137360	-3.493789	0.0011
ARBLOS	-0.000152	0.000106	-1.431587	0.1597
@TREND	-0.006077	0.004611	-1.317992	0.1947
AR(1)	0.941209	0.031491	29.88852	0.0000

R-squared	0.990331	Mean dependent var	1.520885
Adjusted R-squared	0.989180	S.D. dependent var	0.112078
S.E. of regression	0.011658	Akaike info criterion	-5.949131
Sum squared resid	0.005708	Schwarz criterion	-5.715231
Log likelihood	148.7791	Hannan-Quinn criter.	-5.860740
F-statistic	860.3544	Durbin-Watson stat	2.112064
Prob(F-statistic)	0.000000		

Inverted AR Roots	.94

Aufgabe:

Sie erhalten von einem Grossverteiler den Auftrag, die Einkommenselastizität und die Preiselastizität des Nahrungsmittelkonsums zu schätzen.

Die Ausgangsgleichung der Form

LOG(KNAHRK) = C(1)+C(2)*LOG(EINRK)+C(3)*LOG(PREL)

kann keine zuverlässige Antwort liefern, da ihre Residuen extrem autokorreliert und die Parameter zeitlich instabil sind. Obwohl wir letztlich nur an der Einkommens- und Preiselastizität, C(2) und C(3), interessiert sind, kann es sinnvoll sein, die Gleichung mit Kontrollvariablen zu erweitern (ARBLOS, @TREND). Zudem ist eine rein statische Gleichung kaum in der Lage, den Nahrungsmittelkonsum gut zu erklären. Entweder muss man zu einen "Partial adjustment"-Ansatz (Ansatz B: PA-Modell) übergehen, oder man muss in der Gleichung explizit einen autokorrelierten Störprozess zulassen (Ansatz C: AR-Modell).

Vergleichen Sie die folgenden Modellspezifikationen (alle logarithmisch):

PA-Modell

endogen	exogen			
KNAHRK	EINRK	PREL	ARBLOS	@TREND

(entspricht Gleichung 3)

AR-Modell

endogen	exogen			
KNAHRK	EINRK	PREL	ARBLOS	@TREND

(entspricht den Gleichungen auf Seiten 11-13)

DIFF-Modell (Modell in ersten Differenzen der logarithmierten Variablen)

endogen	exogen			
KNAHRK	EINRK	PREL	ARBLOS	@TREND

(entspricht dem AR-Modell für den Grenzfall $\hat{\rho}$ = 1.0)

Die Variablen ARBLOS und @TREND sind in allen Ansätzen als Kontrollvariablen zu betrachten, d.h. wenn die Qualität der Gleichungen (Autokorrelation der Residuen, Gleichungsfit, zeitliche Stabilität der Parameter) darunter nicht leidet, lässt man sie besser weg. Beantworten Sie unter Beachtung dieses Punktes die folgenden Fragen.

Fragen:

1. Wie gross sind in der gemäss Ihrer Analyse besten Version des PA-Modells die kurz- und langfristigen Einkommens- und Preiselastizitäten?

2. Vergleichen Sie die drei Ansätze (in der gemäss Ihrer Analyse besten Version) anhand der folgenden Kriterien:

Gleichungsfit:	R^2, Standardfehler der Gleichung
Autokorrelation:	Durbin-Watson-Test, Durbin-h-Test
Parameterstabilität:	Rekursive Residuen, rekursive Parameterschätzungen Chow-Breakpoint-Test

3. Wie beantworten Sie - vor dem Hintergrund der Testergebnisse - die Anfrage des Grossverteilers nach der Einkommens- und Preiselastizität des Nahrungsmittelkonsums. Testen Sie in der von Ihnen bevorzugten Gleichung die Hypothese einer Einkommens- und Preiselastizität von Null (einzeln und gemeinsam mit t-Test bzw. F-Test).