Einfache Lineare Regression: Übung 1

Schätzung einer Konsumfunktion für die Schweiz

Wir untersuchen die Abhängigkeit der Konsumausgaben der Schweizer Haushalte von der Einkommensentwickklung. Die Daten umfassen den Zeitraum 1960-2008 und stammen aus der Volkswirtschaftlichen Gesamtrechnung der Schweiz. Sie sind im Excel-File Konsum.xls bereitgestellt. Versuchen Sie, die folgenden Schritte in EViews nachzuvollziehen. Am Schluss der Übung sind dann ohne weitere Anleitungen vier Fragen zu beantworten.

Nach dem Aufstarten von EViews wird aus dem Excel-File zunächst ein sogenanntes EViews-WORKFILE erstellt. Man wählt dazu unter FILE die Option NEW und dann WORKFILE. Im Excel-File stehen die Variablenbezeichnungen in der ersten Zeile und die Daten sind darunter spaltenweise angeordnet. Die erste Spalte enthält die Zeitachse (1960a, 1961a, …). Das Vorgehen zur Erstellung des Workfiles ist selbsterklärend. Das Workfile wird unter einem selbstgewählten Namen mit der Extension .wf1 durch Anklicken von SAVE gespeichert (z.B. Konsum.wf1), damit es in späteren Arbeiten weiter zur Verfügung steht. Von den im Excel-File enthaltenen Daten werden in dieser Übung die Folgenden benötigt:

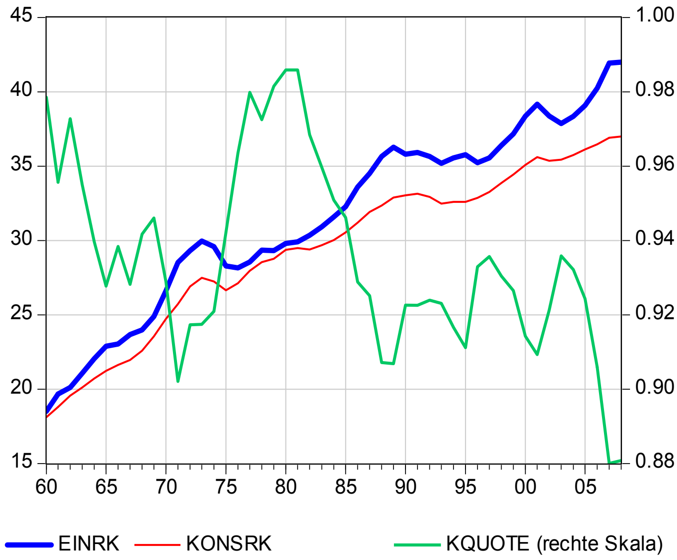
Variablen:

KONSN	Konsumausgaben in Mio Franken, nominell
EINN	Haushaltseinkommen in Mio Franken, nominell
PKONS	Konsumentenpreise, Index 2000 = 1
WOBEV	Wohnbevölkerung in 1000 Personen

Die zu untersuchende Grundhypothese lautet, dass der reale pro Kopf Konsum vom realen pro Kopf Einkommen abhängt.

In einem ersten Schritt werden der reale pro-Kopf Konsum und das reale pro-Kopf Einkommen durch Anklicken von GENR (für generate) berechnet (Division der nominellen Grössen durch den Preisindex und die Bevölkerung). Die derart berechneten Variablen nennt man z.B. KONSRK und EINRK. Sodann ist es sinnvoll, sich die Variablen grafisch anzusehen. Dadurch erhält man ein "Gefühl" für die Daten und es würden auch grobe Datenfehler zum Vorschein kommen. Man kann z.B. auch den Anteil des Konsums am Einkommen (durchschnittliche Konsumquote) ausrechnen, indem man im WORKFILE wiederum die Option GENR wählt und dort KQUOTE = KONSRK/EINRK eingibt. Die Variablen lassen sich mit EVIEWS grafisch darstellen, indem man sie im WORKFILE markiert, dann SHOW anklickt und dort die Option GRAPH wählt. Grafiken können anschliessend angepasst werden, indem man sie anklickt und unter den eingeblendeten Optionen die passenden auswählt. Für die gemeinsame Darstellung von KONSRK, EINRK und KQUOTE in einer Grafik muss z.B. KQUOTE auf eine separate Achse gezeichnet werden. Ist eine Grafik fertiggestellt, kann man sie über EDIT und COPY in ein CLIPBOARD und von dort z.B. in ein WORD-Dokument einfügen.

Einkommen und Konsum (real pro-Kopf), Konsumquote



Die Konsumquote bewegt sich zwischen 88% und 99%. Bei schwachen (starkem) Einkommenswachstum nimmt sie meist eher zu (ab). Darin kommt eine gewisse Trägkeit in der Anpassung des Konsums an die Einkommensentwicklung zum Ausdruck.

Wir schätzen nun in einem ersten Schritt die Gleichung

a) $KONSRK_t = \alpha + \beta\, EINRK_t + u_t$

Vorgehen in EViews. Über OBJECT => NEW OBJECT => EQUATION gelangen wir in ein Fenster, wo wir die Gleichung auf zwei Arten eingeben können:

KONSRK C EINRK

Klickt man OK an, so rechnet EVIEWS eine linerare Regression mit KONSRK als abhängiger und EINRK als erklärender Variablen und C als Achsenabschnitt (α).

Alternative kann man kann die Gleichung auch in exliziter Form eingeben:

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

C(1) ist in dieser Schreibweise α und C(2) ist β.

Eine grafische Darstellung der "Gleichungsfits"erhält man, indem man RESIDS anklickt. Die Grafik zeigt die Residuen sowie die aktuellen und berechneten (“fitted“) Werte von KONSRK. Bei den letzteren handelt es sich um die Werte auf der Regressionsgeraden, entsprechend dem Ausdruck C(1)+C(2)*EINRK.

Interpretieren Sie die Parameterschätzung für β und die Masszahl R^2! Berechnen Sie einen 95%-Vertrauensbereich für β. Testen Sie die Hypothese β = 1 auf dem 5%-Signifikanzniveau! Inwiefern werden diese Berechnungen durch die Verteilung der Residuen in Frage gestellt?

Die genauere Analyse der Gleichung a) zeigt, dass diese aufgrund eines stark autokorrelierten Störterms äusserst problematisch ist. Schätzen Sie die alternativen Regressionsmodelle:

b) $\ln(KONSRK_t) = \alpha + \beta \ln(EINRK_t) + u_t$

c) $\Delta \ln(KONSRK_t) = \alpha + \beta \Delta \ln(EINRK_t) + u_t$, was gleichbedeutend ist mit

$$\ln(KONSRK_t) - \ln(KONSRK_{t-1}) = \alpha + \beta\left(\ln(EINRK_t) - \ln(EINRK_{t-1})\right) + u_t$$

Die Operatoren in EViews für den natürlichen Logarithmus und die erste Differenz sind LOG bzw. D. Man schreibt also z.B. LOG(EINRK), D(LOG(EINRK)) oder kurz DLOG(EINRK). Manchmal ist es vorteilhaft, diese Transformationen mit GENR unter neuen Namen im WORKFILE abzulegen (z.B. LEINRK = LOG(EINRK)). Meist ist es aber besser, man schreibt die Transformationen direkt in die Gleichungsspezifikation hinein, also z.B. LOG(KONSRK) = C(1) + C(2)*LOG(EINRK)

Fragen:

1. Überlegen Sie sich theoretisch die Bedeutung der Ansätze a) bis c)! Beachten Sie insbesondere die unterschiedliche Interpretation des Parameters β.
2. Berechnen Sie in allen Gleichungen 95%-Vertrauensbereiche für β und testen Sie in b) und c) die Hypothese β = 1 !
3. Untersuchen Sie, ob die Verteilung der Residuen auf Autokorrelation hindeutet (DW-Test)!
4. Schätzen Sie die Gleichungen a) bis c) mit einer verkürzten Stichprobe von 1960 bis 1990! Diskutieren Sie die Veränderungen in den Schätzungen für β ! Verwenden Sie die geschätzten Gleichungen, um den realen pro Kopf Konsum in den Jahren 1991 bis 2008 zu "prognostizieren"! (Im Fenster mit den Regressionsresultaten FORECAST anwählen.) Welche der drei Gleichungen "prognostiziert" die tatsächliche Entwicklung des Konsums am besten?
5. Überprüfen Sie in den drei Gleichungen die zeitliche Stabilität der geschätzten Parameter!

Vergessen Sie nicht, die geschätzten Gleichungen im WORKFILE zu speichern. Nach Beendigung Ihrer Arbeit sollten Sie auch das WORKFILE selbst in Ihrem persönlichen Bereich ablegen. Somit können Sie später darauf zurückgreifen.

Ergänzende Bemerkungen zur Übung

1. Vergleich von R^2 zwischen verschiedenen Gleichungen

Ein Vergleich von R^2 zwischen verschiedenen Gleichungen ist nur sinnvoll, wenn die abhängige Variable die gleiche ist. Nur in diesem Fall ist *TSS* im Nenner von R^2 gleich gross, so dass sich der von R^2 angegebene “erklärte“ Varianzanteil auf die gleiche Gesamtvarianz bezieht.

Für die Gleichungen

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

wäre folglich ein Vergleich von R^2 problematisch.

Überhaupt nicht vergleichbar sind die R^2 zwischen einer “Niveaugleichung“ und einer “Differenzengleichung“, z.B:

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

(c) DLOG(KONSRK) = C(1) + C(2)*DLOG(EINRK)

Man kann jedoch (c) identisch umformulieren:

(c’) LOG(KONSRK) = LOG(KONSRK(-1)) + C(1) + C(2)*DLOG(EINRK)

In der Form (c’) ist die Gleichung hinsichtlich R^2 mit (b) vergleichbar (sofern das Regressionsprogramm bei der Berechnung von R^2 den Term LOG(KONSRK(-1)) als “Erklärende“ auf der rechten Gleichungsseite belässt und nicht Variablen, die ohne zu schätzende Parameter auf der rechten Gleichungsseite erscheinen, automatisch auf die linke Seite schafft).

2. Stablilität der geschätzten Parameter

In einer Regressionanalyse wird davon ausgegangen, dass die zu schätzenden Parameter (zum Beispiel Einkommenselastizität des Konsums) über die Zeit *konstant* sind (die Haushalte also immer gleich stark auf Einkommensänderungen reagieren).

Ob diese Voraussetzung erfüllt ist, kann mit Hilfe einer "rekursiven Schätzung" überprüft werden: Ausgehend von einer verkürzten Stichprobe werden sukzessive weitere Beobachtungen zum Schätzzeitraum hinzugefügt. Dadurch wird ersichtlich, ob sich die interessierenden Parameter über die Zeit systematisch verändern (z.B. im Sinne von Verhaltensänderungen). Ist dies der Fall, so ist die Gleichung fehlspezifiziert, weil fälschlicherweise davon ausgegangen wird, dass die Parameter konstant sind.

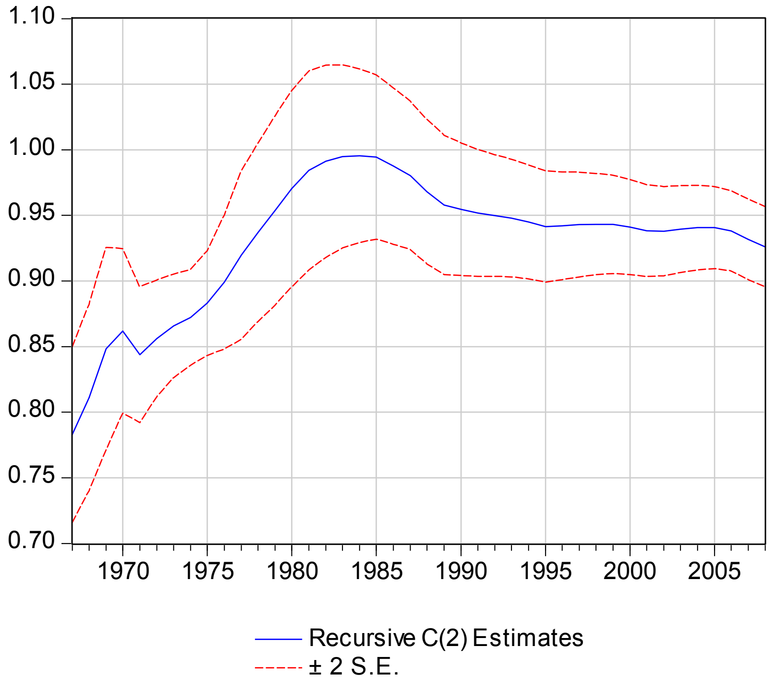
In EViews:

Nach Schätzung der Gleichung gelangt man über

View => Stability Tests => Recursive Estimates => Recursive Coefficients

zur gewünschten Analyse.

In der logarithmischen Konsumgleichung ergibt sich z.B: für C(2), die Einkommenselastizität, das folgende – sehr ungünstige - Bild:



3. Out-of-sample Forecasts

Eine weitere Möglichkeit für die Beurteilung der Regressionsgleichung besteht in folgendem Vorgehen:

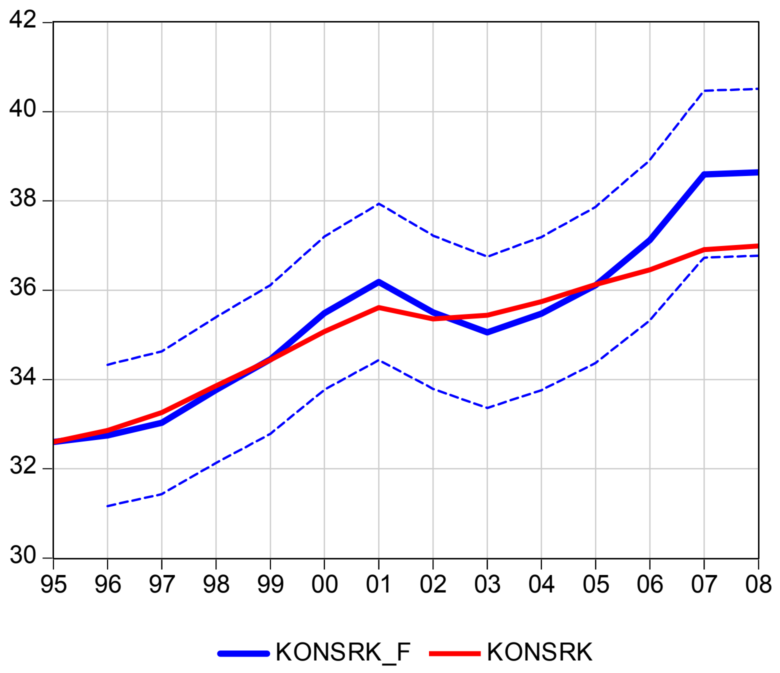
Man schätzt die Gleichung über einen verkürzten Zeitraum (z.B. 1960-1995) und "prognostiziert" den Konsum für die folgenden Jahre (z.B. 1996-2008) unter Verwendung der tatsächlichen Einkommensentwicklung. Danach vergleicht man diese "Prognose" mit den tatsächlichen Realisationen des Konsum. Signifikante Abweichungen von der Prognose deuten darauf hin, dass die Gleichung fehlspezifiziert ist (z.B. im Sinne von Verhaltensänderungen).

Für die logarithmische Konsumgleichung ergibt sich das folgende Bild:

Logarithmische Konsumgleichung:

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

KONSRK_F: Forecast (mit 95%-Vertrauensintervall)
KONSRK Tatsächliche Entwicklung



Das Ergebnis ist ziemlich zufriedenstellend: Die tatsächliche Entwicklung bleibt im 95%-Vertrausensintervall der Prognose.