

# Ökonomische Zeitreihenanalyse — Teil 2

Robert M. Kunst  
University of Vienna  
and  
Institute for Advanced Studies Vienna

November 24, 2006

## 1 Zeitreihen mit Trend

### 1.1 Kurze Geschichte der Trends in der Ökonomie

Es ist offensichtlich, dass viele in der Wirtschaftswissenschaft behandelte Variable, vom Bruttoinlandsprodukt bis zum Börsenindex, Trends in einer solchen Form aufweisen, dass eine Fortschreibung mit konstantem Mittel unplausibel ist. Andere, wie Inflationsraten und Arbeitslosenraten, weisen über längere Zeiträume hinweg steigende oder fallende Tendenzen auf, eine Extrapolation des Trendverhaltens scheint aber in diesen Fällen weniger sinnvoll.

Da die Methoden der Zeitreihenanalyse auf die Behandlung stationärer Variablen zugeschnitten sind, ist es eine nahe liegende Aufgabe, trendende Zeitreihen so zu behandeln (zu ‘transformieren’), dass Zeitreihen mit stationärem Aussehen entstehen. Vielfach wird behauptet, dass vor Erscheinen des Werks von BOX & JENKINS in der Ökonomie stets Funktionen der Zeit  $\tau(t)$  von trendenden Variablen  $X_t$  subtrahiert wurden, um stationäre Variable  $\tilde{X}_t = X_t - \tau(t)$  zu erhalten. In Wirklichkeit wurde etwa das RandomWalk-Modell des Aktienmarktes von BACHELIER bereits um 1900 entwickelt. Folgt aber  $X_t$  einem RandomWalk, dann ist  $\tilde{X}_t = X_t - \tau(t)$  für jede Funktion  $\tau(t)$  keine stationäre Größe, die Bereinigung durch Subtraktion eines Zeittrends einfacher Form versagt daher. Demgegenüber liefert für einen RandomWalk

$$X_t = X_{t-1} + \varepsilon_t$$

mit *white noise*  $\varepsilon_t$  die Anwendung des Operators ‘erste Differenzen’  $\Delta = 1 - B$

eine stationäre Variable

$$\Delta X_t = X_t - X_{t-1} = \varepsilon_t,$$

in diesem Fall sogar *white noise*. Der Operator  $\Delta = 1 - B$  ‘erste Differenzen’ verwendet hier die Notation  $BX_t = X_{t-1}$  für den Lagoperator ‘backshift’. Freilich wurden seit jeher erste Differenzen von logarithmierten makroökonomischen Größen gebildet, um Wachstumsraten zu erhalten. Es ist allerdings richtig, dass die Einführung der ARIMA-Modelle durch BOX&JENKINS zu einer verstärkten Anwendung des ‘Differenzierens’ geführt hat, während nichtlineare Zeitfunktionen  $\tau(t)$  nur mehr selten verwendet wurden. Hier sollte erinnert werden, dass BOX&JENKINS empfehlen, Korrelogramme der originalen Zeitreihe  $X_t$  neben solche der einmal differenzierten Daten  $\Delta X_t$  und der doppelt differenzierten Daten  $\Delta\Delta X_t = \Delta^2 X_t$  zu stellen. Jene Version, die das augenscheinlich einfachste Korrelogramm aufweist, sollte weiter verwendet und als ARMA modelliert werden.

In der Notation von BOX&JENKINS heißt ein Prozess  $\text{ARIMA}(p, d, q)$ , wenn nach  $d$ -maligem Differenzieren ein stationärer  $\text{ARMA}(p, q)$ -Prozess entsteht. Faktisch treten nur  $d = 0, 1, 2$  auf, wobei  $d = 0$  eben einen stationären ARMA-Prozess bezeichnet, daher gilt  $\text{ARIMA}(p, 0, q) = \text{ARMA}(p, q)$ . Später wurde das Konzept auch auf negative  $d$  und nicht-ganzzahlige  $d$  erweitert. Etwa seit Mitte der 1980er-Jahre spricht man statt von  $\text{ARIMA}(p, 1, q)$  lieber von  $I(1)$  und statt von  $\text{ARIMA}(p, 0, q)$  lieber von  $I(0)$ . Die genauen Definitionen von  $I(0)/I(1)$  schwanken etwas zwischen verschiedenen Quellen. Absicht dürfte aber sein, den Modellrahmen gegenüber reinen ARMA-Modellen zu erweitern, sodass etwa  $I(0)$  auch dann vorliegt, wenn keine endliche ARMA-Darstellung existiert.

Die verstärkte Verwendung erster Differenzen in der Ökonomie traf allerdings auf gewissen Widerstand. Die Vorstellung eines RandomWalks oder allgemeiner eines  $\text{ARIMA}(p, 1, q)$ -Modells

$$\Phi(B) \Delta X_t = \Theta(B) \varepsilon_t$$

impliziert nämlich, dass ein eventuell zum Zeitpunkt  $t$  auftretender ‘Schock’ das Niveau der Variable  $X$  dauerhaft ändert, wobei  $\varepsilon_t$  gerne als ein solcher ‘Schock’ interpretiert wird. Demgegenüber gilt für das ‘trend-stationäre’ Modell

$$\Phi(B) (X_t - a - bt) = \Theta(B) \varepsilon_t,$$

dass die Wirkung eines derartigen Schocks für  $t + h$  mit  $h \rightarrow \infty$  abstirbt, bei einem reinen MA-Modell nach  $q$  Beobachtungen überhaupt ‘vergessen’ ist. Viele ÖkonomInnen können sich nicht vorstellen, dass ein einmaliger

Schock eine Variable wie etwa realen Output *ad infinitum* beeinflusst. Mit einiger Fantasie wurde die Gültigkeit des ARIMA( $p, 1, q$ )–Modells mit der Dominanz von Angebotsschocks, die Gültigkeit des trendbereinigten ARMA–Modells mit der Dominanz von Nachfrageschocks identifiziert. Ein weiteres Argument gegen das Anwenden des  $\Delta$ –Operators waren die unsicheren Konsequenzen bei der *multivariaten* Modellierung durch Zerstörung langfristiger Information. Dieses Problem wurde später als ‘Kointegration’ formalisiert.

Während BOX&JENKINS deskriptive Grafiken, wie etwa Korrelogramme, empfehlen, um zu einer Entscheidung über ARIMA( $p, 1, q$ ) versus ARIMA( $p, 0, q$ ) = ARMA( $p, q$ ) zu gelangen, präsentierten DICKEY&FULLER (1979) einen statistischen Test zu diesem Problem. Der DICKEY–FULLER–Test (später ADF–Test) setzte sich schnell durch und wurde in der Wirtschaftswissenschaft allgemein eingesetzt. Es erwies sich, dass die meisten trendenden ökonomischen Variablen eher zu differenzieren sind, was im Jargon fortan I(1) genannt wurde. Es zeigte sich aber auch, dass die Stichproben der Ökonomie meist zu klein sind, um eine gesicherte Entscheidung zu ermöglichen. Zweifel an diesem I(1)–Prinzip stützten sich auch auf die Möglichkeit von Strukturbrüchen und auf Verallgemeinerungen des im ADF–Test unterlegten autoregressiven Rahmens.

Die Forschung späterer Jahre wurde durch Tests im multivariaten Rahmen (Kointegrationstests), durch Tests bei Vorhandensein von Strukturbrüchen, und durch nichtparametrische und semiparametrische Tests dominiert.

## 1.2 Der Test von Dickey und Fuller

In einem einfachen AR(1)–Modell

$$X_t = \phi X_{t-1} + \varepsilon_t \quad (1)$$

ist (asymptotische) Stationarität gleich bedeutend mit  $\phi \in (-1, 1)$ , während für  $\phi \geq 1$  nicht stationäre Prozesse auftreten. Der Fall  $\phi = 1$  charakterisiert den RandomWalk, also einen ARIMA(0, 1, 0)–Prozess, der durch erste Differenzen in Stationarität überführt wird. Ein Test für die Hypothesen

$$\begin{aligned} H_0 &: \phi = 1, \\ H_A &: \phi \in (-1, 1), \end{aligned}$$

wirkt scheinbar einfach. Die  $t$ –Statistik

$$DF_0 = \frac{\hat{\phi} - 1}{\sigma(\hat{\phi})}$$

bietet sich als Teststatistik an. Für  $\hat{\phi}$  kann ein üblicher Schätzer, wie OLS oder ML, eingesetzt werden, für den Nenner wäre die Wurzel aus  $\sigma_\varepsilon^2 / \sum X_{t-1}^2$  denkbar. Allerdings ist seit langem—nicht erst seit DICKEY&FULLER—bekannt, dass  $DF_0$  unter  $H_0$  *nicht asymptotisch normal verteilt* ist, womit auch die in Regressionsprogrammen dargestellte Annäherung durch  $t$ -Verteilung ungültig wird. Die asymptotische Verteilung unter  $H_0$  wurde bereits in FULLER (1976) tabelliert. Später wurde sie oft ‘Dickey-Fuller-Verteilung’ genannt und in der Form

$$DF_0 \sim \frac{\int B(\omega) dB(\omega)}{\{\int B(\omega)^2 d\omega\}^{1/2}} \quad (2)$$

dargestellt. Das Symbol  $B(\omega)$  steht für die so genannte *Brown’sche Bewegung* (*Brownian motion*), ein Gegenstück zu *white noise* in stetiger Zeit. Formal wird die  $Bm$ , ein reeller Prozess auf dem Intervall  $[0, 1]$  durch die Eigenschaften

$$\begin{aligned} B(0) &= 0, \\ B(t+h) - B(t) &\sim N(0, h), \\ E\{B(t+h) - B(t)\}B(t) &= 0, \end{aligned} \quad (3)$$

definiert. Aus der Definition folgt sofort die wichtige Eigenschaft  $B(t) \sim N(0, t)$  und auch  $B(1) \sim N(0, 1)$ . Prozesse, die sich nur durch die Skalierung unterscheiden, also  $\lambda B(t) \sim N(0, \lambda^2 t)$ , werden ebenso als Brown’sche Bewegungen bezeichnet. Zur Unterscheidung heißt der Prozess gemäß Definition (3) ‘standardisierte Brown’sche Bewegung’ (*standard Brownian motion*). In Bruchausdrücken wie (2) ist die Skalierung bedeutungslos.

Einen gewissen Eindruck von der  $Bm$  erhält man durch ein einfaches Zufallsexperiment, bei dem man skalierte RandomWalks der Form

$$X_{t,T} = \frac{1}{\sqrt{T}} \sum_{j=1}^t \varepsilon_j, \quad t = 0, \dots, T,$$

mit  $N(0, 1)$ -Zufallszahlen  $\varepsilon_j$  herstellt und die Trajektorie gegen  $0, 1/T, 2/T, \dots, 1$  anstatt gegen  $0, \dots, T$  aufträgt. Figur 1 zeigt 20 solcher Trajektorien. Lässt man  $T$  gegen  $\infty$  gehen, wird das Zittern immer mikroskopischer und  $Bm$  wird erreicht.

Die Integrale über  $[0, 1]$  wie in (2) existieren nicht, weil die Trajektorien von  $Bm$  zu stark ‘zittern’, können aber im Sinne von

$$T^{-1} \sum_{t=1}^T (X_t - X_{t-1}) X_{t-1} \rightarrow \int_0^1 B(\omega) dB(\omega),$$

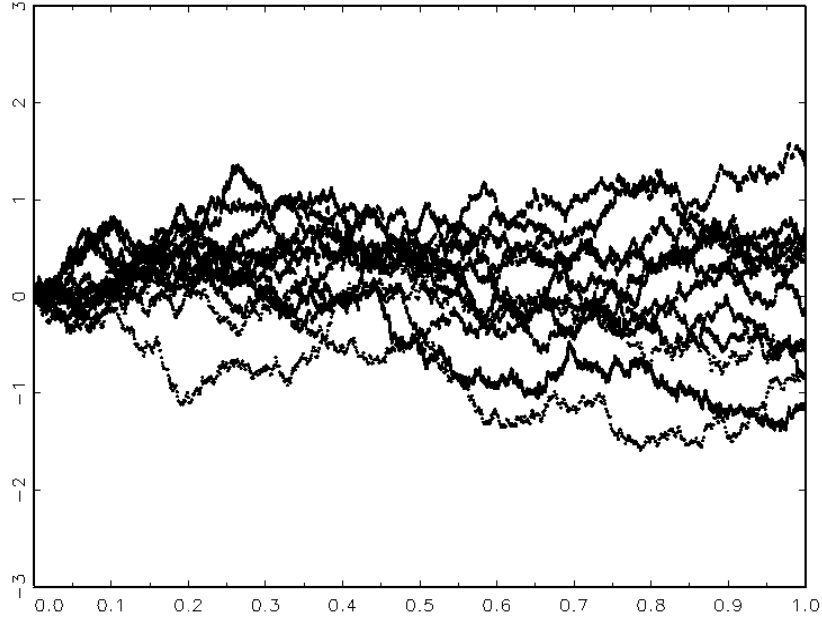


Figure 1: 20 Trajektorien des RandomWalks  $X_t = X_{t-1} + \varepsilon_t$  mit  $X_0 = 0$  und  $\varepsilon_t \sim iid N(0, 1)$ , für  $t = 0, \dots, 10^5$ .

$$T^{-2} \sum_{t=1}^T X_t^2 \rightarrow \int_0^1 B^2(\omega) d\omega$$

als Grenzwerte definiert werden. Diese Formeln sind insofern sinnvoll, weil sie ausgenützt werden können, um die  $DF_0$ -Verteilung aus leicht generierbaren RandomWalks  $X_t$  beliebig genau zu simulieren.

Meistens wird die  $DF_0$ -Statistik direkt als  $t$ -Statistik von  $\varphi$  in der Regression

$$\Delta X_t = \varphi X_{t-1} + \varepsilon_t \quad (4)$$

dargestellt. Natürlich ist  $\varphi = 0$  genau dann wenn  $\phi = 1$ , daher sind die Darstellungen (1) und (4) äquivalent. Neben der  $t$ -Statistik schlugen DICKEY&FULLER auch den Gebrauch des gewichteten Schätzwertes

$$T(\hat{\phi} - 1) = T\hat{\varphi}$$

als Teststatistik vor. Diese Idee wird heute nur mehr selten angewandt.

### 1.3 Konstante und Trends im Dickey-Fuller-Test

Das reine AR(1)-Modell (1) oder (4) ist in der Praxis nicht sehr interessant. Größere Relevanz hat das Modell mit Konstante

$$X_t = \mu + \phi X_{t-1} + \varepsilon_t. \quad (5)$$

Unter der Hypothese  $\phi = 1$  stellt es einen so genannten *Random Walk mit Drift* dar, der durch Einsetzen die Form

$$X_t = \mu t + \sum_{s=1}^t \varepsilon_s + X_0$$

gewinnt. Unter der Alternative  $\phi \in (-1, 1)$  ist es ein (zumindest asymptotisch) stationärer Prozess mit Erwartungswert

$$EX_t = \frac{\mu}{1 - \phi}.$$

Ein offensichtlicher Vorschlag für eine Teststatistik wäre wieder die  $t$ -Statistik

$$DF_\mu = \frac{\hat{\phi} - 1}{\sigma(\hat{\phi})}$$

in (5). Es zeigt sich jedoch, dass  $DF_\mu$  unter  $H_0 : \phi = 1$  eine andere (asymptotische) Verteilung als  $DF_0$  besitzt. Heute wird sie als

$$DF_\mu \sim \frac{\int \bar{B}(\omega) d\bar{B}(\omega)}{\{\int \bar{B}(\omega)^2 d\omega\}^{1/2}}$$

geschrieben, wobei  $\bar{B}(\omega)$  eine zentrierte *Brown'sche Bewegung* ist, die dadurch entsteht, dass das Mittel jeder Trajektorie über das Intervall  $(0, 1)$  gebildet und subtrahiert wird. Daher gilt zum Beispiel  $\bar{B}(0) = -\int B(\omega) d\omega$  im Gegensatz zu  $B(0) = 0$ . Diese  $DF_\mu$ -Verteilung der  $DF_\mu$ -Statistik bleibt gültig, wenn in (5)  $\mu = 0$  gilt. Auch diese Verteilung wurde von FULLER und von DICKEY&FULLER erstmals tabelliert.

Figur 2 gibt wieder eine Vorstellung von der zentrierten  $Bm$ . Es wurden die gleichen Zufallszahlen wie bei Figur 1 verwendet. Von der gewöhnlichen  $Bm$  unterscheiden sich die Trajektorien optisch dadurch, dass sie nicht bei 0 beginnen und sich durch die Zentrierung etwa in der Hälfte ihres Verlaufs um 0 verdichten.

Der  $DF_\mu$ -Test findet heute dort seine Anwendung, wo zwar Verdacht auf Nichtstationarität besteht, aber ein längerfristiges Trendverhalten entweder

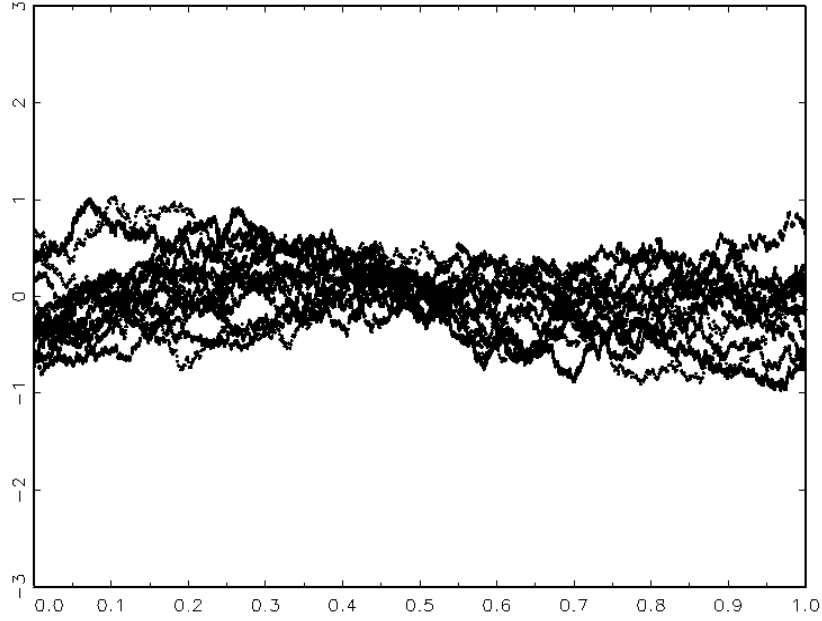


Figure 2: 20 Trajektorien des zentrierten RandomWalks  $\bar{X}_t = X_t - \frac{1}{T} \sum_{t=0}^T X_t$  mit  $X_t = X_{t-1} + \varepsilon_t$ ,  $X_0 = 0$ , und  $\varepsilon_t \sim iid N(0, 1)$ , für  $t = 0, \dots, 10^5$ .

nicht beobachtet wird oder nicht vorstellbar ist. Beispiele wären Zinssätze oder Inflationsraten. Treten echte Trends auf, dann enthält das Modell (5) zwar die Nullhypothese eines *RandomWalk mit Drift*, nicht aber die logische Alternative eines nach Subtraktion eines linearen Zeittrends stationären Prozesses. Für diesen Fall wird das Modell mit linearem Zeittrend

$$X_t = \mu + \tau t + \phi X_{t-1} + \varepsilon_t \quad (6)$$

empfohlen. Hier wurden zwei verschiedene Nullhypothesen erwogen. Die Hypothese

$$H_{0F} : \tau = 0, \phi = 1$$

impliziert eine  $F$ -artige Teststatistik, die allerdings unter  $H_{0F}$  nicht  $F$ -verteilt ist. Empirisch wird dieser Test seltener verwendet als ein auf einer  $t$ -Statistik beruhender Test für

$$H_{0T} : \phi = 1.$$

Dieser Test heißt auch  $DF_\tau$ -Test und ist der heute gebräuchlichste ‘Test auf Einheitswurzeln’ (*unit root test*). Die asymptotische Verteilung liegt ebenfalls in Tabellenform vor. Der Ausdruck ‘*unit root test*’ bezieht sich

auf die Nullstellen (Wurzeln, *roots*) des charakteristischen Polynoms des autoregressiven Modells  $\phi(z) = 1 - \phi z$ . Im einfachen AR(1)–Modell ist eine Einheitswurzel (*unit root*) gleich bedeutend mit  $\phi = 1$ .

Der Nachteil des  $DF_\tau$ –Tests gegenüber dem  $F$ –Test auf  $H_{0F}$  besteht darin, dass unter der Nullhypothese  $H_{0T}$  das Modell

$$X_t = \mu + \tau t + X_{t-1} + \varepsilon_t$$

auftritt, also ein *RandomWalk* mit überlagertem quadratischen Trend, wie man durch wiederholtes Einsetzen leicht zeigt (**Übung**). Dieses Modell ist für die meisten makroökonomischen Variablen nicht brauchbar und liefert schlechte Prognosen bei längerem Prognosehorizont. Es sollte daher nur als Hilfsmodell für den Test gesehen werden und nicht direkt bei Annahmen der Nullhypothese  $H_{0T}$  übernommen werden. Insbesondere sollten die geschätzten Werte  $\hat{\mu}$  und  $\hat{\tau}$  nicht weiter verwendet werden.

Die Verteilung der  $DF_\tau$ –Statistik beruht nun auf Integralen über enttendete  $Bm$ , die etwa wie die Trajektorien in Figur 3 aussehen. Durch die Einpassung linearer Trends erhalten sie optisch eine über das gesamte Intervall  $[0,1]$  gestreute Zitterform.

Figur 4 zeigt Box-Plots der Verteilungen der Statistiken  $DF_\mu$  und  $DF_\tau$  unter der Nullhypothese des RandomWalk gemäß einer simplen Simulation mit 100 Beobachtungen und 10,000 Wiederholungen. Es ist deutlich, dass die Werte um etwa  $-2$  zentriert sind, womit die Verteilung sich klar von einer Normal- oder  $t$ –Verteilung unterscheidet. Die Verteilungen sind allerdings kaum asymmetrisch und weisen Kurtosis von 3.32 und 3.48 auf, also nur wenig jenseits der Normalverteilung (Kurtosis 3). Positive Werte treten nur sehr selten auf. Daraus kann man schließen, dass ein RandomWalk, wenn er als AR(1)–Prozess geschätzt wird, fast immer geschätzte Koeffizienten liefert, die etwas kleiner als 1 sind und fast nie explosive Werte größer 1. Das ist eine Extremform des so genannten *Hurwicz-Bias*, der alle Zeitreihenmodelle betrifft. In der endlichen Stichprobe sind alle üblichen Schätzer für Zeitreihen-Koeffizienten biased (nicht erwartungstreu), wobei im Allgemeinen der Absolutwert des Koeffizienten systematisch unterschätzt wird.

## 1.4 Allgemeine autoregressive Modelle im Dickey-Fuller–Test

Unterstellt man statt AR(1) ein allgemeines AR( $p$ )–Modell, dann sind die einfachen Statistiken  $DF_0$ ,  $DF_\mu$ ,  $DF_\tau$  für Testentscheidungen nicht direkt geeignet, die asymptotischen Verteilungen stimmen nicht mehr. Für dieses Problem wurden zwei Lösungen entwickelt. Die ältere, einfachere stammt



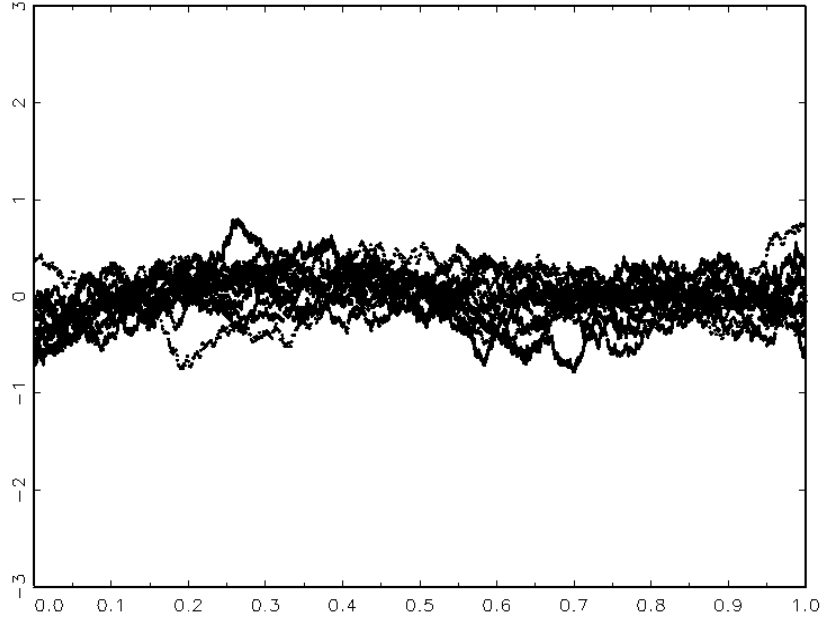


Figure 3: 20 Trajektorien des linear enttendeten RandomWalks  $\tilde{X}_t$ .  $\tilde{X}_t$  entsteht als Residuum der Regression von  $X_t$  auf  $\mu + \tau t$ , wobei  $X_t = X_{t-1} + \varepsilon_t$ ,  $X_0 = 0$ , und  $\varepsilon_t \sim iid N(0, 1)$ , für  $t = 0, \dots, 10^5$ .

von DICKEY&FULLER und wird mit ADF-Test (*augmented Dickey-Fuller*) bezeichnet. Die jüngere, semiparametrische Variante wird heute meist PHILLIPS& PERRON zugeschrieben.

Der  $AR(p)$ -Prozess

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t \quad (7)$$

ist (asymptotisch) stationär, wenn alle Wurzeln des Polynoms

$$\phi(z) = 1 - \phi_1 z - \phi_2 z^2 - \dots - \phi_p z^p$$

im Betrag größer eins sind. Ist genau eine Wurzel gleich 1, dann gilt

$$\phi(z) = (1 - z) \phi_0(z),$$

wobei alle Wurzeln von  $\phi_0(z)$  größer 1 sind. Das impliziert, dass  $\Delta X_t$  stationär ist,  $X_t$  aber nicht. Man spricht von einem *integrierten Prozess 1. Ordnung* oder  $I(1)$ -Prozess.

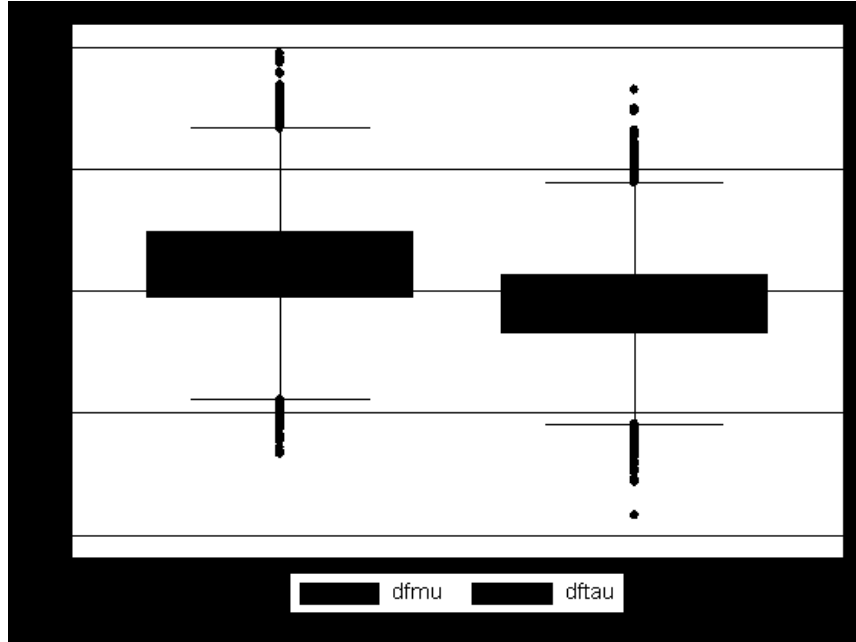


Figure 4: Box-Plots der empirischen Verteilungen der Statistiken  $DF_\mu$  und  $DF_\tau$ . Generierendes Modell sind RandomWalks mit 100 Beobachtungen. 10,000 Replikationen.

Damit ergibt sich logisch folgende Verallgemeinerung der Hypothesen eines UnitRoot-Tests:

$$\begin{aligned} H_0 &: \phi(1) = 0, \\ H_A &: \phi(1) \neq 0, \end{aligned}$$

wobei verschiedene Einschränkungen der allgemeinen Hypothese ( $H_0 \cup H_A$ , die *maintained hypothesis*) anzubringen sind. Mehrfache Nullstellen von 1 sind genau so ausgeschlossen wie Nullstellen mit Betrag 1 aber ungleich 1, meistens auch alle ‘explosiven’ Wurzeln kleiner 1. Üblicherweise werden alle Varianten des DF-Tests als *einseitige* Tests betrachtet, also mit Stationarität unter der Alternative.

DICKEY&FULLER schlugen das Regressionsmodell

$$\Delta X_t = \varphi X_{t-1} + \psi_1 \Delta X_{t-1} + \dots + \psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t \quad (8)$$

als Grundlage ihres Tests vor. Die  $t$ -Statistik für  $\varphi = 0$  hat unter der Nullhypothese asymptotisch die gleiche Verteilung wie die Statistik  $DF_0$  im AR(1)-Modell. Hierzu wären zumindest folgende Punkte zu zeigen:

1. Jedes  $AR(p)$ -Modell (7) lässt sich in die Form (8) bringen. Es gibt eine bijektive Beziehung zwischen den Parametern  $\phi_1, \dots, \phi_p$  und  $\psi_1, \dots, \psi_{p-1}; \varphi$ .
2.  $\varphi = 0$  genau dann, wenn  $\phi(1) = 0$ .
3. Die Regressionsterme  $\Delta X_{t-1}, \dots, \Delta X_{t-p+1}$  haben asymptotisch keinen Einfluss auf die Verteilung der Teststatistiken  $DF_0, DF_\mu, DF_\tau$ , wenn  $\varepsilon_t$  *white noise* ist.

Die erste Behauptung ist eine einfache algebraische Aufgabe und ist zur **Übung** empfohlen. Einfacher ist es, zunächst die Form (8) in die Form (7) zu bringen und  $\phi_1, \dots, \phi_p$  als Funktionen von  $\psi_1, \dots, \psi_{p-1}; \varphi$  auszudrücken. Das entstehende lineare Gleichungssystem lässt sich leicht invertieren.

Für die zweite Behauptung ist fest zu halten, dass

$$\phi(z) = 1 - z - \varphi z - \psi_1(1-z)z - \dots - \psi_{p-1}(1-z)z^{p-1},$$

falls Behauptung 1 gilt. Die Eigenschaften des Polynoms müssen ja von der Darstellungsform unabhängig sein. Hier wird der Lagoperator  $B^h X_t = X_{t-h}$ , der  $z^h$  entspricht, ausgenützt und auch  $\Delta = 1 - B$ , was  $1 - z$  entspricht. Man sieht sofort  $\phi(1) = -\varphi$  und die zweite Behauptung ist erwiesen.

Die dritte Behauptung wird hier nicht bewiesen. Sie bleibt gültig, wenn *white noise* durch andere, ähnliche Anforderungen ersetzt wird, wird aber ungültig, wenn ein ARMA-Prozess vorliegt oder wenn die Ordnung  $p$  zu niedrig spezifiziert wurde. In diesen Fällen ist  $\varepsilon_t$  autokorreliert. Später wurde allerdings gezeigt, dass es möglich ist, den ADF-Test auch auf ARMA-Fälle anzuwenden, wenn man bereit ist, die Ordnung  $p$  mit zunehmender Stichprobe zu vergrößern.

Wie der originale DF-Test wird auch der ADF-Test auf Modelle mit Konstanten und Trends angewandt, wobei wieder die Variante mit Trend  $ADF_\tau$  die gebräuchlichste ist.

Ein wichtiges praktisches Problem stellt die Bestimmung der Lagordnung für den ADF-Test dar. Es wurden hier unter anderem folgende Vorschläge gemacht:

1.  $p$  in Abhängigkeit von der Stichprobengröße fest setzen;
2. Informationskriterien (meist AIC) für AR-Modelle in  $X_t$  (nicht in  $\Delta X_t$ ) minimieren;
3.  $p$  vergrößern, bis die Q-Statistik für residuale Autokorrelation insignifikant wird.

Simulationen haben gezeigt, dass der dritte Vorschlag schlecht ist. Im Zweifelsfall sollte eher eine etwas größere Lagordnung gewählt werden.

**Beispiel.** Als Beispiel möge hier eine Untersuchung der Zeitreihe der österreichischen Industrieproduktion dienen, die im 1. Teil grafisch dargestellt ist. Nach saisonaler Filterung mittels  $0.25(1 + B + B^2 + B^3)$  ist ein Trend erkennbar. Einheitswurzeln mit Wert ungleich 1 sind unwahrscheinlich, da diese fast nur durch Saison generiert werden, welche aber ausgefiltert wurde. Mittels AIC wird für das autoregressive Modell mit Konstante und Trend ein Lag von  $p = 6$  fest gestellt. Daher wird ein ADF-Test mit **fünf** Lags in Differenzen verwendet. Die Teststatistik hat den Wert  $-2.334$ . Durch Vergleich mit dem tabellierten Signifikanzpunkt sieht man, dass die Nullhypothese nicht verworfen wird, daher scheint die Variable  $I(1)$  zu sein. Die Software EViews gibt hier einen kritischen Wert auf dem Niveau von 10% mit  $-3.143$  an, Stata den Wert von  $-3.141$ . Der tabellierte Wert wäre  $-3.15$ . EViews und Stata verwenden hier die Technik der *response surfaces*, um den kritischen Wert besonders genau der Stichprobengröße und anderen Spezifika anzupassen. Der erzielte Genauigkeitsgewinn ist nicht immer überzeugend.  $\square$

## 1.5 Der Test von Phillips und Perron

Die Testidee von PHILLIPS&PERRON fußt auf der Beobachtung, dass zwar die asymptotische Verteilung der Statistiken  $DF_0$ ,  $DF_\mu$ ,  $DF_\tau$  im Modell

$$\Delta X_t = d + \varphi X_{t-1} + u_t$$

(mit  $d = 0$ ,  $d = \mu$ ,  $d = \mu + \tau t$ ) bei autokorreliertem  $u_t$  nicht mit den originalen DF-Verteilungen überein stimmt, der Unterschied sich aber durch relativ einfache Korrekturfaktoren ausdrücken lässt. Die Korrekturfaktoren hängen von der ACF von  $u_t$  ab. Man könnte also diese Terme aus dem Korrelogramm der geschätzten  $\hat{u}_t$  berechnen, die Teststatistik selbst modifizieren, und erhält dann eine asymptotisch den bei DICKKY&FULLER tabellierten Verteilungen gehorchende Statistik. Dieser PP-Test wurde vor allem damit motiviert, dass er gültig bleibt, wenn  $X_t$  einem  $ARIMA(p, 1, q)$ -Modell folgt und nicht einem reinen AR-Modell. Auch wäre die Berechnung von der Wahl der Lagordnung unabhängig.

In Wahrheit wurde gefunden, dass der PP-Test bei ARIMA-Modellen gegenüber dem ADF-Test keineswegs die erwarteten Vorteile in Bezug auf *power* bietet. Zudem verlangt der Test eine Spezifikation eines Maximums, bis zu welchem Autokorrelationen in den Korrekturfaktor aufzunehmen sind. Heute ist der Test vor allem zur Kontrolle der Resultate der ADF-Entscheidung gebräuchlich.

## 1.6 Tests auf Stationarität statt UnitRoot-Tests

Für manche AnwenderInnen ist es unbefriedigend, dass die Nullhypothese des DF-Tests und seiner Varianten ein nicht-stationäres Modell ist und die Alternative ein stationäres Modell. Demgegenüber erscheint uns Stationarität als eine genau definierte Eigenschaft, die eher als Nullhypothese dienen sollte. Dieses Argument ist nicht überzeugend, denn die Nullhypothese des DF-Tests ist eben nicht Nichtstationarität, sondern mit  $I(1)$  oder  $ARIMA(*, 1, *)$  eine sehr spezielle Modellklasse.

Will man stationäre Modelle (inklusive ‘trend-stationäre’ Prozesse, die sich additiv aus Trendfunktion und stationärem Anteil zusammensetzen) zur Nullhypothese machen, so existieren zwei Strategien, um zu einem Test zu gelangen. Die erste beruht auf der Beobachtung, dass erste Differenzen stationärer Prozesse eine *unit root* im MA-Teil ihrer ARMA-Darstellung besitzen. Zum Beispiel haben erste Differenzen von *white noise*

$$\Delta \varepsilon_t = \varepsilon_t - \varepsilon_{t-1} = (1 - B) \varepsilon_t$$

eine solche MA-UnitRoot. Ist hingegen die beobachtete Variable  $I(1)$ , dann ist ihre erste Differenz stationär, ohne eine solche MA-UnitRoot zu besitzen. Daher könnte man ein MA-Modell für  $\Delta X_t$  schätzen und darauf testen, ob  $\Theta(1) = 0$  oder  $\Theta(1) \neq 0$ .

Eine alternative Strategie fußt auf dem Konzept der ‘strukturellen’ Zerlegungen in *unobserved components*. In dieser Denkwelt ist jedes beobachtete  $X_t$  als Summe eines *Random Walk* und eines stationären Rests darstellbar. Ist der durch den *Random Walk* erklärte Anteil nicht signifikant, wird  $X_t$  als stationär klassifiziert. Diese Denkwelt ist prinzipiell nicht-parametrisch. Ein typisches UC-Modell hat die Gestalt

$$\begin{aligned} X_t &= \mu + \tau_t + \gamma t + u_t, \\ \tau_t &= \tau_{t-1} + \eta_t, \end{aligned}$$

wobei  $u_t$  ein stationärer Restprozess ist,  $\eta_t$  hingegen *white noise* mit Varianz  $\sigma_\eta^2$ . Die Nullhypothese  $H_0 : \sigma_\eta^2 = 0$  beschreibt den trend-stationären Fall  $\tau_t = \tau_{t-1} = \tau$ , die Alternative  $H_A : \sigma_\eta^2 > 0$  einen  $I(1)$ -Prozess  $X_t$ .

Der gebräuchlichste Test dieser Klasse ist derjenige von KWIATKOWSKI & PHILLIPS & SCHMIDT & SHIN (KPSS-Test). Eine Schwierigkeit in der Praxis stellt das Auftreten widersprüchlicher Entscheidungen zwischen KPSS- und ADF-Test dar. Zum Beispiel können beide Tests verwerfen, womit  $X$  als  $I(0)$  nach ADF, aber als  $I(1)$  nach KPSS klassifiziert wird. Diese beiden Informationen unter einen Hut zu bringen, ist schwieriger als bei zwei Tests mit identischen Nullhypothesen, weil die Signifikanzniveaus nicht übereinstimmen.

Lösungsvorschläge, wie *joint confirmation analysis*, werden in der Literatur oft nicht akzeptiert, weil sie den gewohnten Rahmen der Hypothesentests verlassen.

**Beispiel.** Eine der in der Literatur verwendeten Empfehlungen für die Bestimmung der Laglänge für den KPSS-Test liegt bei  $(5/7)\sqrt{T}$ . Für den Fall der Industrieproduktion liegt dieser Wert etwas über 9. Die KPSS-Statistik ergibt sich nach Stata als 0.277. Damit wird der angegebene 1%-Signifikanzpunkt von 0.216 übertroffen und Trend-Stationarität verworfen. Es existiert auch eine dem  $DF_\mu$ -Test entsprechende KPSS-Statistik, die Stationarität und nicht Trend-Stationarität als Nullhypothese hat. Diese verwirft mit einem Wert von 1.74 noch überzeugender, der kritische 1%-Wert liegt hier bei 0.74. Für die ersten Differenzen, also die Wachstumsrate, ergibt sich ein Wert von 0.169, deutlich unter dem 10%-Wert von 0.347. Die KPSS-Tests bestätigen den bisherigen Eindruck, die Industrieproduktion entspreche einem  $I(1)$ -Modell.

## 2 Kointegration

Obwohl visuelle Hilfsmittel wie die ACF und auch Hypothesentests zur Verfügung stehen, um die Notwendigkeit des ‘Differenzierens’ zu überprüfen, zögern viele instinktiv, Differenzen einzelner Variablen zu bilden, wenn das Ziel multivariate Analyse, also das gemeinsame Modellieren mehrerer Variablen, ist. Während eine Beziehung wie

$$y_t = \gamma x_t + u_t \quad (9)$$

als Wirkung der Variable  $x$  auf  $y$  zu interpretieren ist, scheint die entsprechende Beziehung

$$\Delta y_t = \gamma \Delta x_t + u_t \quad (10)$$

eine eher marginale Reaktion auf geringfügige Zuwächse in  $x$  darzustellen. Tatsächlich enthält (9) Information über die gemeinsame längerfristige Entwicklung von  $x$  und  $y$ , die in (10) nicht enthalten ist. Dieser Umstand ist, nach früheren Vorarbeiten, in den 1980er-Jahren zufrieden stellend in der Zeitreihenanalyse behandelt worden. Als Bahn brechender Beitrag ist hier ENGLE&GRANGER (1987) zu nennen.

Grundsätzlich gilt, dass bereits stationäre Variablen ( $I(0)$ ) im multivariaten Modell nicht zu differenzieren sind. Integrierte Variable ( $I(1)$ ) können *kointegriert* sein oder nicht. Wenn sie nicht kointegriert sind, werden sie einfach differenziert und gehen so in multivariate Modelle ein. Wenn sie kointegriert sind, dann müssen Fehlerkorrektur-Modelle (*error correction*) gebildet

werden. Für den Fall der Kointegration zweier oder mehrerer  $I(1)$ -Variablen wird auch formal  $CI(1, 1)$  geschrieben.

Als archetypisches Beispiel für Kointegration wird in der Literatur gerne ein Set von zwei Zeitreihen gebracht, das aus dem aggregierten Konsum  $c$  und dem verfügbaren Haushaltseinkommen  $yd$  besteht. Diese Zeitreihen für Österreich sind als Figur 5 dargestellt. Beide Variablen zeigen augenscheinliche Trends, und zumindest für längere Beobachtungszeiträume lässt sich die UnitRoot-Hypothese bestätigen. Die Differenz der logarithmierten Variablen ist aber wahrscheinlich stationär, denn längerfristig ist die Sparneigung der Haushalte wahrscheinlich konstant. ‘Wahrscheinlich’ deshalb, weil sich die Annahme der Kointegration gerade für diese beiden Variablen statistisch keineswegs immer bestätigen lässt.

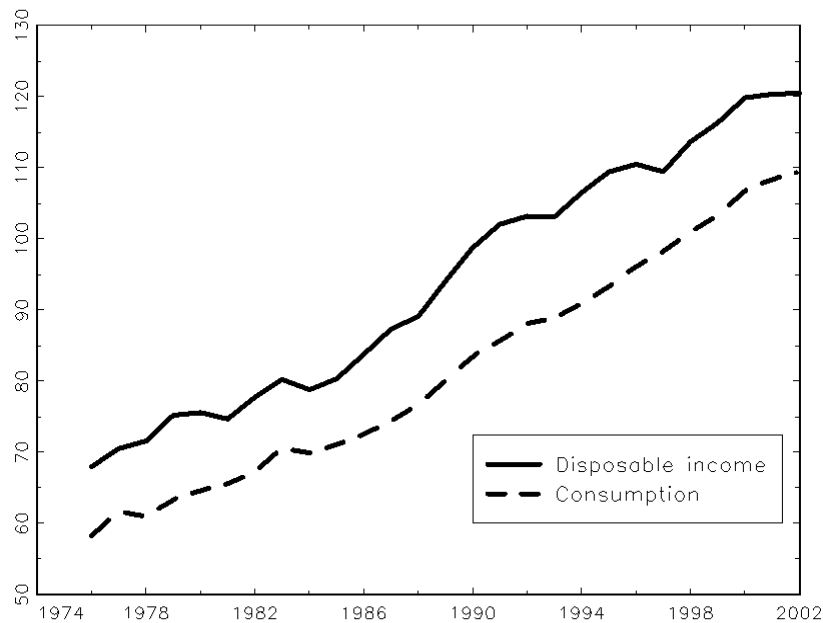


Figure 5: Aggregierter privater Konsum und verfügbares Haushaltseinkommen für Österreich, zu konstanten Preisen, in Logarithmen.

In diesem Beispiel wäre es eigentlich leicht, die Annahme der Stationarität von  $c - yd$  zu überprüfen, etwa durch einen ADF-Test auf diese Variable. Im allgemeineren Fall aber ist die ‘kointegrierende Beziehung’  $\beta$  in

$$\beta' X_t = \beta_1 X_{1t} + \dots + \beta_n X_{nt}$$

für  $n$  Variablen nicht vorgegeben, sondern muss geschätzt werden.

Als Definition der Kointegration soll eine modifizierte Form der von ENGLE&GRANGER gegebenen dienen:

**Definition.** Sind zwei Prozesse  $X_t, Y_t$  jeweils  $\text{ARIMA}(p, 1, q)$  und existiert eine lineare Kombination  $Z_t = \beta_1 X_t + \beta_2 Y_t$  mit  $\beta_1 \neq 0$  und  $\beta_2 \neq 0$ , sodass  $Z_t$  stationär ist, dann heißen  $X_t$  und  $Y_t$  *kointegriert* und  $(\beta_1, \beta_2)'$  heißt der *kointegrierende Vektor*.

Es ist sinnvoll, sich auf  $\text{ARIMA}(p, 1, q)$ -Prozesse zu beschränken. Alternativ könnte das anderswo übliche  $I(1)$  verwendet werden. Hier gilt die allgemeine Definition:

**Definition.** Ein Prozess  $X_t$  heißt *integriert der Ordnung  $k$* , in Zeichen  $I(k)$ , wenn  $\Delta^k X_t$  stationär ist, nicht jedoch  $\Delta^{k-1} X_t$ .

Auch hier könnte man sich auf  $\text{ARIMA}(p, k, q)$ -Prozesse beschränken. Die Behandlung deterministischer Anteile (Konstanten, Trends, Saisondummies) ist in der Literatur nicht ganz einheitlich. In der originalen Definition von ENGLE&GRANGER sind alle deterministischen Teile, inklusive Konstanten, ausgeschlossen. Die meisten Autoren würden  $X_t = \mu + \varepsilon_t$  als  $I(0)$  bezeichnen, womit dann auch  $X_{t-1} = X_t + \mu + \varepsilon_t$ , also ein Prozess mit überlagertem linearen Trend, als  $I(1)$  gilt. Es ist jedoch wenig sinnvoll,  $X_t$  und  $Y_t$  als kointegriert zu bezeichnen, wenn  $Z_t = \beta_1 X_t + \beta_2 Y_t$  ein stationärer Prozess mit überlagertem linearen Trend ist. Man sollte hier fordern, dass dieser lineare Trend in  $Z_t$  verschwindet.

### Beispiele:

1. Für jeden RandomWalk mit Drift  $X_t = a + X_{t-1} + \varepsilon_t$  ist  $Y_t = bX_t + \eta_t$  mit einem *white noise*  $\eta_t$ , der von  $\varepsilon_t$  ua. ist, mit  $X_t$  kointegriert, denn

$$Z_t = bX_t - Y_t = -\eta_t$$

ist stationär und  $(b, -1)$  ist ein kointegrierender Vektor.

2.  $X_t = \varepsilon_t - \varepsilon_{t-1}$  ist stationär, aber nicht  $I(0)$ , denn

$$\Delta^{-1} X_t = \sum_{s=1}^t X_s = \varepsilon_t - \varepsilon_0$$

ist ebenfalls stationär und die Definition ist nicht erfüllt. Weil nun  $\Delta^{-2} X_t$  ein nicht stationärer RandomWalk ist, heißen derartige Prozesse  $I(-1)$ . Hier tritt der zum Differenzenoperator  $\Delta$  inverse Operator  $\Delta^{-1}$  auf, also das Akkumulieren vom Anfangswert bis  $t$ . Man sieht leicht, dass  $\Delta \Delta^{-1} X_t = X_t$ .



3. Werden  $X_t$  und  $Y_t$  durch das bivariate AR-Modell

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$$

hergestellt, sind beide Komponenten Random Walks und damit  $I(1)$ . Sie sind auch dann nicht kointegriert, wenn  $\varepsilon_t$  und  $\eta_t$  mit einander korreliert sind.

4. Werden  $X_t$  und  $Y_t$  durch das bivariate AR-Modell

$$\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{bmatrix} 0.9 & 0.1 \\ 0 & 1 \end{bmatrix} \begin{pmatrix} X_{t-1} \\ Y_{t-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_t \\ \eta_t \end{pmatrix}$$

hergestellt, dann folgt wegen

$$X_t - Y_t = 0.9X_{t-1} - 0.9Y_{t-1} + \varepsilon_t - \eta_t$$

für  $Z_t = X_t - Y_t$ , dass

$$Z_t = 0.9Z_{t-1} + \varepsilon_t - \eta_t.$$

Also ist  $Z_t$  stationär,  $X_t$  und  $Y_t$  sind kointegriert mit Vektor  $(1, -1)'$ .  $Y_t$  ist ein RandomWalk und damit  $I(1)$ ,  $X_t$  ist als Summe eines RandomWalk und eines stationären Prozesses  $Z_t$  ebenfalls  $I(1)$ , die Bedingungen sind also erfüllt. Es wird sich noch zeigen, dass die Eigenschaften der Koeffizientenmatrix entscheidend sind.

Werden mehr als zwei Variable gemeinsam untersucht, dann reicht die oben gegebene Definition nicht mehr aus.

**Definition.** Gilt für einen  $n$ -dimensionalen Vektorprozess  $X_t = (X_{1t}, \dots, X_{nt})'$ , dass  $\Delta X_t = (\Delta X_{1t}, \dots, \Delta X_{nt})'$  stationär ist, während mindestens eine Komponente  $X_{jt}$   $I(1)$  ist, und dass eine Linearkombination mit mindestens einem  $\beta_k \neq 0$

$$Z_t = \beta_1 X_{1t} + \dots + \beta_n X_{nt} = (\beta_1, \dots, \beta_n)' X_t$$

stationär (oder  $I(0)$ ) ist, dann heißt  $X_t$  *kointegriert*, in Zeichen  $CI(1, 1)$ , und der Vektor  $\beta$  heißt ein *kointegrierender Vektor*.

### Beispiele:

1. Ist  $X_{1t}$  ein RandomWalk und  $X_{2t}$  ein unabhängiger *white noise*, dann ist  $(X_{1t}, X_{2t})'$  im Sinn der Definition kointegriert mit Vektor  $(0, 1)'$ . Dieser eher triviale Fall heißt *self cointegration* und wird von manchen Autoren ausgeschlossen. Es ist aber nicht sinnvoll zu fordern, dass *alle* Komponenten  $I(1)$  sind.

2. Ist  $X_{1t}$  ein RandomWalk und entstehen  $X_{2t}$  und  $X_{3t}$  durch Addieren von einander unabhängiger *white noise*'s zu  $X_{1t}$ , dann sind  $X_{2t} - X_{1t}$  ebenso wie  $X_{3t} - X_{1t}$  beide stationär. Man hat also die kointegrierenden Vektoren  $(1, -1, 0)'$  und  $(1, 0, -1)'$ . Auch alle Vielfachen und alle Linearkombinationen dieser Vektoren sind kointegrierende Vektoren. Der kointegrierende Vektor ist nicht eindeutig bestimmt.
3. Als ökonomisches Beispiel werden hier gerne Zinssätze verschiedener Fristigkeit genannt. Während die individuellen Zinssätze  $I(1)$  sein sollen, werden Differenzen als zeitlich konstante Risikoprämien (Fristprämien) interpretiert und sind stationär. Beide Annahmen sind nicht immer gesichert. Einige Autoren finden zeitvariante, nicht stationäre Risikoprämien, andere bezweifeln, dass Zinssätze  $I(1)$  sind.
4. Ein anderes Beispiel stellen die drei Variablen GDP, privater Konsum und aggregierte Investitionen, jeweils in Logarithmen, dar. Diese makroökonomischen Größen sind individuell alle  $I(1)$ , ihre Differenzen sind Logarithmen des Konsum- oder Investitionsanteil im Output und daher wahrscheinlich stationär. Dieses Beispiel ist durch Figuren 6 und 7 grafisch unterlegt.

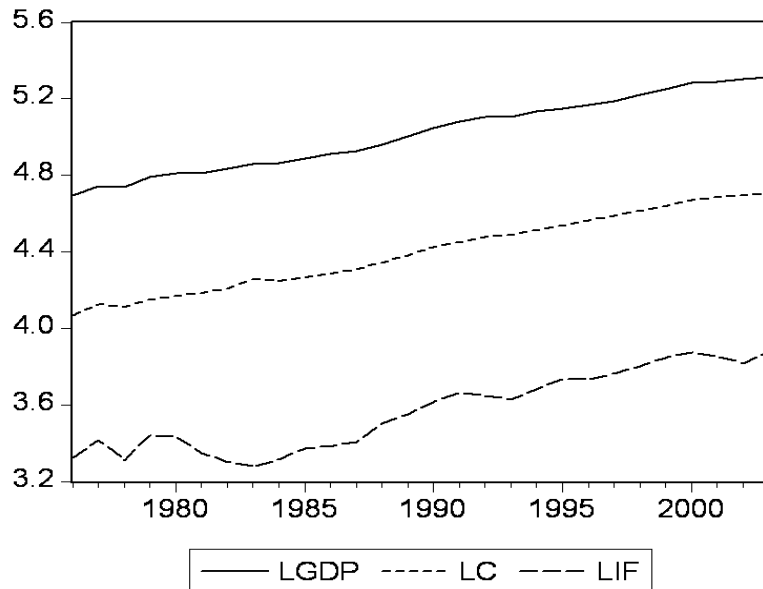


Figure 6: Austrian GDP (solid), private consumption (short dashes), and gross fixed investment (long dashes) series. Logarithmic transformations.

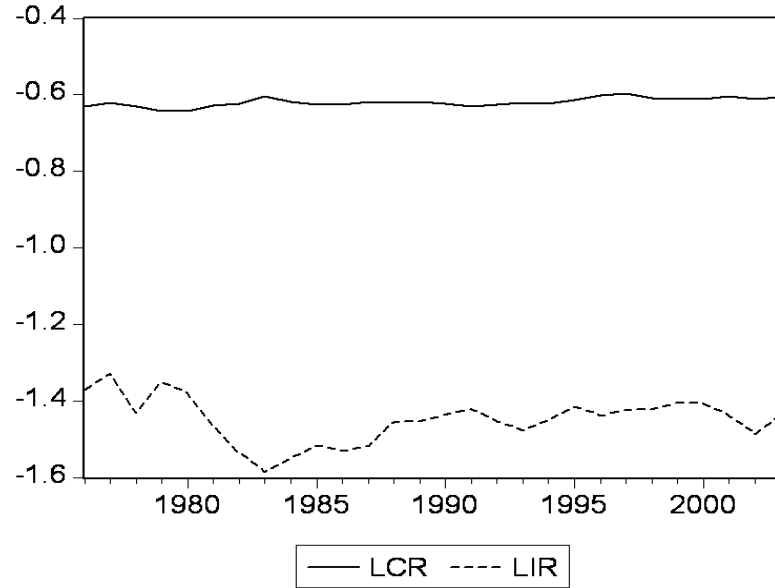


Figure 7: Logarithmic consumption-output ratio (solid) and investment-output ratio (dashes) for Austria.

## 2.1 Die kointegrierende Regression

Die älteste, von ENGLE&GRANGER vorgeschlagene Technik zum Schätzen kointegrierender Beziehungen ist die so genannte *kointegrierende Regression*. Idee ist, die angenommene kointegrierende Beziehung

$$Z_t = \beta_1 X_{1t} + \dots + \beta_n X_{nt}$$

nach Division durch  $\beta_1$ —wenn  $\beta_1 = 0$ , ist analog ein anderes  $\beta_j \neq 0$  zu wählen—auf die Form

$$\bar{Z}_t = X_{1t} + \bar{\beta}_2 X_{2t} + \dots + \bar{\beta}_n X_{nt}$$

zu bringen. Ist  $Z_t$  stationär, dann ist auch  $\bar{Z}_t = (1/\beta_1) Z_t$  stationär. Durch Umgruppierung erhält man

$$X_{1t} = -\bar{\beta}_2 X_{2t} - \dots - \bar{\beta}_n X_{nt} + \bar{Z}_t,$$

was einer Regression ähnelt. Der Fehler ist stationär, aber meist von *white noise* stark verschieden und zeitlich korreliert. Sind aber alle  $X_{jt}$   $I(1)$ -Variable, so lässt sich zeigen, dass trotzdem alle mit OLS geschätzten Koeffizienten  $\hat{\beta}_j$  gegen die wahren Werte konvergieren, und zwar ‘schneller’ als

in einer üblichen Regression mit stationären Regressoren und *white noise*-Fehlern. Wegen dieser schnellen Konvergenz können die Schätzwerte wie ‘wahre’ Werte weiter verwendet werden. Allerdings sind die in der OLS-Regression ausgewiesenen Standardfehler und  $t$ -Werte nicht gültig. Typisch in kointegrierenden Regressionen sind sehr hohe  $t$ -Werte.

Als Beispiel dient das einfache System aus Konsumausgaben und Einkommen

$$\begin{aligned} C_t &= \gamma Y_t + \zeta_t, \\ Y_t &= \delta + Y_{t-1} + \varepsilon_t. \end{aligned}$$

Hier folgt das Einkommen einem RandomWalk mit Drift, der Prozess  $\zeta_t$  ist irgend ein stationärer ARMA-Prozess. Die Variablen sind kointegriert, die marginale Konsumneigung  $\gamma$  wird aus der ersten Gleichung durch OLS geschätzt.

Ein sinnvolleres System dient als Prototyp der sogenannten Fehlerkorrektur-Modelle

$$\Delta C_t = \mu_1 + \alpha (C_{t-1} - \beta Y_{t-1}) + \gamma \Delta Y_{t-1} + \varepsilon_t, \quad (11)$$

$$\Delta Y_t = \mu_2 + \delta \Delta Y_{t-1} + \eta_t. \quad (12)$$

Hier kann man annehmen, dass  $\varepsilon_t$  und  $\eta_t$  *white noise* sind. Der negative Koeffizient  $\alpha$  dämpft die Konsumnachfrage, wenn sie in der vorigen Periode gegenüber der ‘Gleichgewichtsbeziehung’  $C = \beta Y$  zu hoch war, und verstärkt sie, wenn sie vorher niedrig war. Man kann zeigen, dass solche Fehlerkorrekturmodelle genau dann auftreten, wenn  $C$  und  $Y$  kointegriert sind mit Vektor  $(1, -\beta)'$ . Genau dieser Beweis ist für die Arbeit von ENGLE&GRANGER zentral gewesen. Auch dieses Modell lässt sich mit der kointegrierenden Regressionstechnik bewältigen. Zuerst wird die Regression

$$C_t = \beta_0 + \beta Y_t + \zeta_t$$

mit stationärem Fehler  $\zeta_t$  geschätzt, dann wird  $\hat{\beta}Y_{t-1}$  in (11) eingesetzt, als wäre es der wahre Wert, und beide Gleichungen (11) und (12) werden mit OLS geschätzt, um an die übrigen Koeffizienten heran zu kommen. In diesen Regressionen der zweiten Stufe sind die üblichen  $t$ -Werte und andere Statistiken wieder gültig.

Laut ENGLE&GRANGER ist diese Technik ‘effizient’, sie führt aber doch in kleineren Stichproben zu starken Bias-Effekten. Im akademischen Diskurs wird sie unter dem Namen EG-2 (ENGLE&GRANGER *two-step*) nur mehr selten oder als Benchmark verwendet. In der Praxis ist sie aber stark in Gebrauch, weil sie einfach zu handhaben ist. Ihr Vorteil liegt darin, dass sie

erlaubt, Gleichungen wie (11) isoliert zu behandeln, ohne das Gesamtsystem zu modellieren, also ohne (12). Empiriker verwenden auch gerne Varianten wie

$$\Delta C_t = \mu_1 + \alpha (C_{t-1} - \beta Y_{t-1}) + \gamma \Delta Y_t + \varepsilon_t,$$

also mit gleich laufendem Regressor  $\Delta Y_t$ , was ökonomisch leicht zu interpretieren ist, theoretisch aber Schwierigkeiten macht.

Typisch für die EG-2-Methode ist es auch, die Annahme, dass überhaupt Kointegration existiert, ebenfalls an Hand der kointegrierenden Regression zu überprüfen (*Test auf Kointegration*). Zunächst wurde hier vorgeschlagen, in der Regression

$$C_t = \beta_0 + \beta Y_{t-1} + \zeta_t$$

die Durbin-Watson-Statistik zu betrachten. Da die Fehler stationär, aber nicht unkorreliert sind, wurde erwogen, bei DW-Statistiken, die hinreichend von Null(!) verschieden sind, Kointegration zu diagnostizieren, also die *Nullhypothese der Nicht-Kointegration* zu verwerfen. Heute wird dieser DW-Test wegen seiner unbefriedigenden Eigenschaften nicht mehr verwendet. In allen gebräuchlichen Tests ist aber weiterhin Nicht-Kointegration die Nullhypothese und Kointegration die Alternative.

Der nächste Vorschlag war es, Kointegration durch ADF-Tests auf die Residuen  $\hat{\zeta}_t$  zu überprüfen. Bei Kointegration sollten ja die wahren  $\zeta_t$  stationär sein. Die Residuen  $\hat{\zeta}_t$  haben aber etwas andere Eigenschaften als die wahren Fehler  $\zeta_t$ , und daher sind die üblichen Signifikanzpunkte *ungültig*. Ein Verwenden der von der Software automatisch ausgewiesenen Signifikanzpunkte für diesen *residualen ADF-Test* ist ein häufiger Fehler in der Praxis. PHILLIPS & OULIARIS bieten für diesen Test adäquate Fraktile in Tabellenform an.

Die auch in Software-Dokumentationen dargestellte Vorgangsweise hat insgesamt folgende Punkte:

1. Die Variablen  $X_j$  werden durch ADF-Test oder andere UnitRoot-Tests auf ihre Integrationsordnung geprüft. Hier sind die Standardtabellen und automatisch erzeugte Signifikanzpunkte gültig. Verwirft der ADF-Test, dann ist die Variable möglicherweise bereits stationär ( $I(0)$ ) und kann in einer kointegrierenden Regression keine Rolle spielen. Ist keine oder nur eine Variable  $I(1)$ , kann keine (nicht triviale) Kointegration aufgefunden werden.
2. Variable, die in den Akzeptanzbereich der UnitRoot-Tests fallen, werden häufig auch noch auf  $I(2)$  getestet. Dies geschieht durch Anwendung der UnitRoot-Tests auf deren erste Differenzen. Jetzt sollte die

UnitRoot verworfen werden, ansonsten besteht Verdacht auf  $I(2)$  oder gar auf  $I(k)$  mit  $k > 2$ .  $I(2)$ -Variable können zwar mit anderen  $I(2)$ - und auch  $I(1)$ -Variablen kointegriert sein, jedoch macht die Behandlung solcher Fälle Schwierigkeiten. Echte  $I(2)$ -Variable treten in der Ökonomie selten auf, eher ist ausgewiesenes  $I(2)$  ein Hinweis auf Strukturbrüche o.ä.

3. Die kointegrierende Regression wird ausgeführt. Als Regressand wird meistens jene Variable gewählt, die aus theoretischen Gründen als von den anderen  $X_j$  abhängig gesehen werden könnte. In Wahrheit sollte die Auswahl des Regressanden keinen Einfluss haben, solange dieses  $X_j$  im kointegrierenden Vektor enthalten ist.
4. Auf die Residuen der kointegrierenden Regression wird ein ADF-Test angewandt. Hier sind die Standardtabellen ungültig. Verwirft der Test, so liegt Evidenz für Kointegration vor. Verwirft der Test nicht, dann kointegrieren die Variablen entweder nicht oder—bei mehr als zwei modellierten  $I(1)$ -Variablen—die als Regressand gewählte Variable kommt im kointegrierenden Vektor nicht vor. Als Kontrolle kann die Regression mit anderem Regressanden wiederholt werden.
5. Hat sich die Annahme der Kointegration bestätigt, gilt der geschätzte Vektor  $(1, \beta_1, \dots, \beta_k)'$  als kointegrierender Vektor. Die Residuen der kointegrierenden Regression können als Fehlerkorrektur-Variable in anderen Gleichungen weiter verwendet werden.

**Beispiel.** Wir untersuchen die Jahresdaten für Konsum und Einkommen in logarithmierter Form, wie sie in Figur 1 dargestellt sind.

1. Die Variable  $c$  wird einem ADF-Test unterzogen. Eine AIC-Suche ergibt eine Lagordnung von  $p = 3$ , also zwei Differenzen von  $\Delta c_t$  sind als Regressoren zu verwenden. Die entsprechende  $DF_\tau$ -Statistik hat den Wert  $-2.16$  und ist insignifikant. Für das Einkommen wird  $p = 5$  und vier Lags empfohlen. Die  $DF_\tau$ -Statistik ist mit  $-2.67$  wieder insignifikant. Das Gesamtbild bestätigt die  $I(1)$ -Hypothese für beide Variable.
2.  $DF_\mu$ -Tests auf  $\Delta c$  und  $\Delta yd$  liefern die signifikanten Werte  $-7.26$  und  $-4.75$ .  $I(2)$  kommt daher nicht in Betracht, wäre aber eigentlich *a priori* unrealistisch gewesen. Hier ist nur der  $DF_\mu$ -Test sinnvoll, denn  $DF_\tau$  würde lineare Trends in  $\Delta c$  und daher quadratische Trends in  $c$  zulassen, was nicht plausibel ist.

### 3. Die kointegrierende Regression

$$c_t = -0.306 + 1.036yd_t + \hat{u}_t$$

passt in das theoretische Bild. Die Elastizität ist mit 1.036 nur wenig von eins verschieden. Die negative Konstante impliziert, dass  $c$  im Gleichgewicht kleiner als  $yd$  ist. Die DW-Statistik ist mit 0.63 hinreichend von Null unterschiedlich, weist aber natürlich auf Autokorrelation hin. Die  $t$ -Statistik von 49.8 beim Koeffizienten 1.036 ist typisch überhöht.

4. Ein  $DF_0$ -Test auf die Residuen  $\hat{u}$  ergibt den Wert von  $-2.93$ , der mit den PHILLIPS&OULIARIS-Werten verglichen werden sollte. Ökonometrische Software berechnet die Statistik korrekt, gibt aber falsche Signifikanzpunkte. So gibt zum Beispiel EViews an dieser Stelle als 5%-Punkt  $-1.96$  an. Aus der Tabelle ergibt sich der korrekte 5%-Punkt mit  $-2.76$ . Die Kointegrationsbeziehung wäre also auf dem 5%-Niveau signifikant (Nicht-Kointegration wird verworfen), das Resultat ist aber knapper als es bei naiver Anwendung erscheint.
5. Der Vektor  $(1, -1.036)$  kann als kointegrierender Vektor weiter verwendet werden. Es zeigt sich allerdings, dass in einer Gleichung wie in (11)

$$\Delta c_t = \mu + \alpha(c_{t-1} - 1.036yd_{t-1}) + \gamma\Delta yd_{t-1} + \varepsilon_t$$

der Koeffizient  $\alpha$  nur mäßige  $t$ -Werte rund um  $-1.7$  erreicht, und dies verbessert sich auch nicht, wenn weitere gelagte Differenzen von  $\Delta c$  und  $\Delta yd$  als Regressoren eingefügt werden. Hier ist die klassische Näherung der Normalverteilung für die  $t$ -Statistik von  $\alpha$  gültig. Eine Lösung aller derartiger Widersprüche bietet nur die im nächsten Abschnitt vorgestellte Systemschätzung. Ein völlig klares Bild ist bei der geringen Stichprobengröße allerdings nicht zu erwarten.

## 2.2 Johansen's Systemschätzung

Es gab verschiedene Vorschläge in der Literatur, die EG-2-Methode zu verbessern. Die meisten von ihnen lassen sich in zwei Gruppen einordnen. Die erste konzentriert sich auf eine Verbesserung des Schätzverfahrens in der kointegrierenden Regression. In dieser Gruppe wird weiterhin eine Regression geschätzt. Eigentlich ist aber die kointegrierende Beziehung keine Regressionsbeziehung, welche eine Variable als reagierend auf andere unterstellt, sondern eine Gleichgewichtsbeziehung zwischen mehreren gleich berechtigten

Variablen. Daher rückt eine andere Gruppe von Verfahren von der Regressionsmethode ab und verwendet multivariate Verallgemeinerungen der Korrelationsanalyse. Das gebräuchlichste dieser Verfahren ist die Methode von JOHANSEN, die auch in die Software-Produkte RATS und EViews integriert wurde.

Die JOHANSEN-Methode sieht Kointegrationstests als multivariate Verallgemeinerungen des ADF-Tests. Im ADF-Test wird eine Regression

$$\Delta X_t = \mu + \varphi X_{t-1} + \psi_1 \Delta X_{t-1} + \dots + \psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t$$

geschätzt. Der  $t$ -Wert von  $\varphi$  gilt als Kriterium für die Nullhypothese  $\varphi = 0$  und damit für *unit roots*. Ist  $X_t$  ein Vektor von  $n$  Variablen, dann lässt sich das System

$$\Delta X_t = \mu + \Pi X_{t-1} + \Psi_1 \Delta X_{t-1} + \dots + \Psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t \quad (13)$$

schätzen, und zwar effizient Gleichung per Gleichung durch OLS (Kruskal's Theorem).  $\mu$  ist hier ein  $n$ -Vektor,  $\Pi$ ,  $\Psi_j$ ,  $j = 1, \dots, p-1$  sind  $n \times n$ -Matrizen.  $\varepsilon_t$  ist ein  $n$ -dimensionaler *white noise*, wobei  $E\varepsilon_t \varepsilon_t' = \Sigma$ , d.h. die Fehlerterme der einzelnen Gleichungen sind i.a. korreliert. Demgegenüber hat  $E\varepsilon_t \varepsilon_s' = \mathbf{0}$  für  $s \neq t$  zu gelten, d.h. keine Komponente von  $\varepsilon_t$  ist mit irgend einer Komponente von  $\varepsilon_s$  korreliert, außer  $s = t$ .

Die Demonstration, dass sich jede multivariate oder Vektor-Autoregression (*vector autoregression*, VAR)

$$X_t = \mu + \Phi_1 X_{t-1} + \dots + \Phi_p X_{t-p} + \varepsilon_t$$

in der Form (13) schreiben lässt, gelingt genau so wie im univariaten Fall. Die Zuordnung zwischen  $\mu; \Phi_1, \dots, \Phi_p$  und  $\mu; \Pi; \Psi_1, \dots, \Psi_{p-1}$  ist umkehrbar eindeutig, bijektiv. Auf die vielen und interessanten Details des VAR-Modellierens, wie die multivariate ACF, muss nicht eingegangen werden. Multivariate ARMA-Modelle sind übrigens nicht zufällig selten im Einsatz. Sie leiden unter theoretischen Identifikationsproblemen und numerischer Instabilität bei der Schätzung.

Während das univariate  $\varphi$  nur gleich oder ungleich Null sein kann, existieren für die Matrix  $\Pi$  Zwischenformen mit Rang kleiner  $n$ . Gerade diese sind für die Überprüfung von Kointegration interessant. Es lässt sich nämlich zeigen, dass der Rang von  $\Pi$ , eine Zahl  $\text{rk}\Pi$  mit  $0 \leq \text{rk}\Pi \leq n$ , gerade die *Anzahl* der (linear unabhängigen) kointegrierenden Vektoren *im System* angibt. Wichtig sind also vor allem die Fälle:

1.  $\text{rk}\Pi = 0$ : es gibt keine Kointegration. Alle Variable sind  $I(1)$ . Man kann sie differenzieren und erhält ein vereinfachtes Gesamtsystem.



2.  $\text{rk}\Pi = 1$ : der kointegrierende Vektor ist eindeutig, bis auf Vielfache.
3.  $\text{rk}\Pi = n$ : alle Vektoren kointegrieren und damit auch die Variablen selbst im Sinne von *self cointegration*. Der Vektor  $X$  enthält nur stationäre Variable.

Diese Eigenschaften gelten nur dann, wenn gewisse mögliche Konstellationen ausgeschlossen werden, wie saisonale Einheitswurzeln, Integration höherer Ordnung,  $I(-1)$ , explosives Verhalten. Diese Annahmen werden meistens innerhalb des Verfahrens nicht mehr überprüft. Gegebenenfalls kann man einzelne Variable auf  $I(2)$  testen und saisonbereinigte Reihen verwenden.

### 2.2.1 Die Bestimmung des kointegrierenden Ranges

Der Rang der Matrix  $\Pi$  wird durch Sequenzen von sogenannten *trace*-Statistiken fest gestellt. Diese ergeben sich aus den so genannten *kanonischen Korrelationen* von  $X_{t-1}$  und  $\Delta X_t$ . Für alle Details siehe das Buch von JOHANSEN (1995). Die Verteilung der *trace*-Statistiken  $J_r$  unter den Nullhypothesen  $H_{0r}$  im Testproblem

$$\begin{aligned} H_{0r} &: \text{rk}\Pi \leq r \\ H_{Ar} &: \text{rk}\Pi > r \end{aligned}$$

hängt nur von  $n-r$  ab und ist etwa bei JOHANSEN in Fraktilen tabelliert. Da die JOHANSEN-Prozedur eine multivariate Verallgemeinerung des DF-Tests ist, sind die Verteilungen aller Statistiken unter den Nullhypothesen wieder Funktionen von Integralen über Brown'sche Bewegungen.

In der Praxis geht man meist so vor, dass erst  $J_0$  untersucht wird. Ist  $J_0$  insignifikant, dann gibt es keine Kointegration und es braucht nicht weiter gesucht werden. Verwirft aber der Test  $H_{00}$ , betrachtet man  $J_1$ . Ist  $J_1$  insignifikant, existiert ein eindeutig bestimmter kointegrierender Vektor. Verwirft der Test  $H_{01}$ , betrachtet man  $J_2$ , etc.

Wie beim ADF-Test, so wird auch der JOHANSEN-Test durch die Spezifikation der deterministischen Teile, wie Konstanten oder Trends, beeinflusst. Interessant sind praktisch nur zwei Varianten:

1. ein Konstantenvektor  $\mu$  wird mitgeschätzt wie in (13) ohne weitere Einschränkung.
2. der Konstantenvektor unterliegt einer speziellen *no trend*-Restriktion.

In Fall #1 besitzt das *System* lineare Trends, d.h. mindestens eine Komponente von  $X_t$  trendet systematisch. In Fall #2 weist keine Komponentenvariable systematisches Trendverhalten auf. Man kann zwar zwischen beiden Varianten durch Hypothesentests entscheiden, sinnvoller ist es aber, die Wahl auf Grund von Vorinformation über das Verhalten der Variablen zu treffen. Zum Beispiel trenden makroökonomische Größen wie Output und Konsumausgaben offensichtlich, während Zinssätze oder Inflationsraten nicht trenden sollten.

### 2.2.2 Die Schätzung der Koeffizienten des Systems

Um Fälle #1 und #2 zu verstehen, ist auf die *Schätzung* der Parameter des Systems einzugehen. Hierzu zunächst das aus der linearen Algebra bekannte Resultat, dass jede  $n \times n$ -Matrix  $\mathbf{M}$  vom Rang  $r$  in der Form  $\mathbf{M} = \mathbf{A}\mathbf{B}'$  darstellbar ist, wobei die Matrizen  $\mathbf{A}$  und  $\mathbf{B}$  jeweils Dimension  $n \times r$  besitzen und *vollen Rang*  $r$ . Dies lässt sich auf die Matrix  $\Pi$  der Darstellung (13) anwenden in der Form

$$\Pi = \alpha\beta'$$

mit  $n \times r$ -Matrizen  $\alpha$  und  $\beta$ . Die Zerlegung ist nicht eindeutig, äquivalente Zerlegungen entstehen durch Linearkombinationen der Spalten von  $\alpha$  oder  $\beta$ .

Ist also  $\text{rk}\Pi = r$  bereits bekannt, so kann man statt (13) auch

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \Psi_1 \Delta X_{t-1} + \dots + \Psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t \quad (14)$$

schreiben. Es lässt sich zeigen, dass die Spalten von  $\beta$ , also die Zeilen von  $\beta'$ , die gesuchten *kointegrierenden Vektoren* sind. Auch diese sind also nicht eindeutig. Deshalb heißt  $\beta$  auch die *kointegrierende Matrix*. Die Matrix  $\alpha$  beschreibt, wie die einzelnen Variablen auf Abweichungen von den in  $\beta$  ausgedrückten Gleichgewichtsbeziehungen reagieren. Oft liest man Ausdrücke wie *loading coefficients* oder *adjustment coefficients* für diese Matrix  $\alpha$ , gelegentlich auch *speed of adjustment*.

Schätzwerte für  $\beta$  fallen als Nebenprodukte der Testprozedur ab. Da die Lösung des kanonischen Korrelationsproblems ein Eigenwertproblem voraussetzt, sind zur Berechnung der Statistiken  $J_r$  Eigenwerte nötig. Die zu den von Null verschiedenen Eigenwerten gehörigen Eigenvektoren sind schon die Schätzer für die Spalten von  $\beta$ . Ersetzt man  $\beta$  in (14) durch den Schätzer  $\hat{\beta}$ , dann lässt sich die Gleichung als *Regressionsgleichungssystem* auffassen, in dem die Komponenten von  $\Delta X_t$  auf eine Konstante und  $\hat{\beta}' X_{t-1}$ ,  $\Delta X_{t-1}, \dots, \Delta X_{t-p+1}$ , regressiert werden. Diese OLS-Regression liefert dann Schätzer für  $\mu$ ,  $\alpha$ ,  $\Psi_1, \dots, \Psi_{p-1}$ .

Ist der  $n$ -Vektor  $\mu$  der Matrix  $\alpha$  proportional, d.h. gibt es eine Darstellung  $\mu = \alpha\tilde{\mu}$ , so kann man das System in der Form

$$\Delta X_t = \alpha(\tilde{\mu} + \beta' X_{t-1}) + \Psi_1 \Delta X_{t-1} + \dots + \Psi_{p-1} \Delta X_{t-p+1} + \varepsilon_t$$

mit  $r$ -Vektor  $\tilde{\mu}$  anschreiben. Dies ist genau der Fall #2 der *no trend*-Restriktion. In diesem Fall liegt ein modifiziertes kanonisches Korrelationsproblem zu Grunde, in dem ein auf  $(1, X)'$  und nicht, wie im Fall #1, auf  $X$  operierender Eigenvektor gesucht wird. Dadurch schaut es so aus, als würde Kointegration zwischen  $X$  und einer Konstante gesucht. Das ist natürlich sinnlos, denn eine Konstante kann die Eigenschaft der Stationarität und damit der Kointegration nicht ändern. Leider sieht man doch in Software-Beschreibungen häufig "*constant in the cointegrating relation*", was immer das heißen soll.

Beide Fälle #1 und #2 sind multivariate Gegenstücke zu  $ADF_\mu$  im univariaten Fall. Man könnte auch den Fall ohne Konstante betrachten, als Gegenstück zu  $ADF_0$ , der ist aber praktisch irrelevant. Nur in Ausnahmefällen kann es sinnvoll sein, Gegenstücke zu  $ADF_\tau$  zu betrachten, bei denen lineare Trends verwendet werden. Diese implizieren ja meist quadratische Trends in den Variablen, diese aber sind unplausibel. Das Argument, trend-stationäre Prozesse als Gegenhypothese ins Modell zu integrieren, ist für die JOHANSEN-Prozedur nicht so zwingend, denn diese sind nicht wirklich im Zentrum der Aufmerksamkeit. Meistens wird diese Möglichkeit schon in Vortests mit ADF-Statistiken auf einzelne Variable ausgeschlossen.

### 2.2.3 Ein Beispiel

Wir betrachten das System mit den beiden Variablen  $c$  und  $yd$ . Mit einem geeigneten Programm können VAR-Modelle mit variierender Lagordnung vorgeschätzt werden. Für  $p = 2$  wird das niedrigste AIC-Kriterium gefunden. Da die beiden Variablen augenscheinlich trenden, kommt Fall #2 nicht in Betracht. Geschätzt wird also ein Modell der Form

$$\Delta X_t = \mu + \alpha\beta' X_{t-1} + \Psi_1 \Delta X_{t-1} + \varepsilon_t.$$

Der JOHANSEN-Test wird also mit *einem* Lag in Differenzen durchgeführt, nicht mit zwei, wie dies EViews automatisch empfehlen würde. Hier erhält man die beiden Statistiken  $J_0 = 8.127$  und  $J_1 = 0.027$ . Durch Vergleich mit Tabellen erkennt man, dass  $J_0$  *nicht signifikant* ist, die beiden Variablen sind also *nicht kointegriert*, weil  $H_{00}$  nicht verworfen wird. Grund dürfte die kleine Stichprobe sein.

Verwirft man versuchsweise  $H_{00}$  trotzdem, wird  $H_{01}$  sicher angenommen, denn  $J_1$  ist sehr klein. Setzt man entgegen der Testempfehlung  $r = 1$ , dann

erhält man aus dem Programm den Schätzer  $\hat{\beta}' = (-50.04, 54.35)$ , was einem Vielfachen von  $(1, -1)$  sehr nahe kommt. Die implizierte Elastizität ist mit 1.086 etwas größer eins, was einer langfristigen Zunahme der Konsumneigung entspräche. Die Absolutbeträge der Koeffizienten des Vektors  $\hat{\alpha}$  sind ebenfalls einander nahe.

Setzt man den Vektor  $\beta$  so, dass der erste Eintrag gleich eins ist—EViews erreicht dies durch die Normierung “B(1,1)=1”—erhält man  $\hat{\beta} = (1, -1.086)$  und  $\hat{\alpha}' = (-0.188, 0.153)$ . Allerdings ist der erste Koeffizient mit einem  $t$ -Wert von 1.57 ‘fast signifikant’, der zweite mit einem  $t$ -Wert von 0.81 aber nicht. Es bestätigt sich, dass die Existenz von Kointegration schlecht abgesichert ist, dass aber die Hauptlast der Korrektur von einer Anpassung des Konsums getragen wird und nicht so sehr von einer Anpassung des Einkommens. Die Konstanten in  $\mu$  und die Elemente von  $\Psi_1$  erreichen größtenteils Signifikanz auf dem Niveau von 5% oder 10%. In diesen Fällen gilt die übliche Näherung der Normalverteilung für die  $t$ -Statistiken.

Setzt man statt dem geschätzten  $\hat{\beta}$  den theoretisch begründeten Vektor  $\beta = (1, -1)$ , dann verschlechtert sich die Signifikanz der  $\alpha$ -Einträge. Die Restriktion wird auch statistisch abgelehnt. Für diese Testentscheidung verwenden Programme die Likelihood-Ratio-Statistik  $2(\log L_1 - \log L_0)$ , die asymptotisch  $\chi^2$ -verteilt ist mit Freiheitsgraden, die der Anzahl der Restriktionen entsprechen. Hier ist das eine  $\chi^2(1)$ -Verteilung. Das marginale Signifikanzniveau (der  $p$ -Wert) des Wertes der Teststatistik von 4.52 liegt knapp über 3%.□

## References

- [1] BOX, G.E.P., and JENKINS, G.C. (1976). *Time Series Analysis, Forecasting and Control*. Revised edition. Holden-Day.
- [2] BOX, G.E.P., JENKINS, G.M., and REINSEL, G.C. (1994). *Time Series Analysis, Forecasting and Control*. 3rd edition. Prentice-Hall.
- [3] DICKEY, D.A., AND FULLER, W.A. (1979). ‘Distribution of the Estimators for Autoregressive Time Series with a Unit Root’. *Journal of the American Statistical Association* **74**, 427–431.
- [4] ENGLE, R.F., AND GRANGER, C.W.J. (1987) ‘Co-integration and Error Correction: Representation, Estimation and Testing’. *Econometrica* **56**, 1333–1354.
- [5] FULLER, W.A. (1996). *Introduction to Statistical Time Series*. 2nd Edition, Wiley. (1976: first edition)

- [6] GOURIEROUX, C., and MONFORT, A. (1997). *Time series and dynamic models*. Cambridge University Press.
- [7] HAMILTON, J.D. (1994). *Time Series Analysis*. Princeton University Press.
- [8] JOHANSEN, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press.
- [9] KWIATKOSWKI, D., PHILLIPS, P.C.B., SCHMIDT, P., AND SHIN, Y. (1992) ‘Testing the null hypothesis of stationarity against the alternative of a unit root’. *Journal of Econometrics* **54**, 159–178.
- [10] PHILLIPS, P.C.B., AND PERRON, P. (1988) ‘Testing for a Unit Root in Time Series Regression’. *Biometrika* **75**, 335–346.
- [11] PHILLIPS, P.C.B, AND OULIARIS, S. (1990) ‘Asymptotic Properties of Residual Based Tests for Cointegration’. *Econometrica* **58**, 165–193.