

Einheitswurzeltest (unit-root test) und Kointegrationstest Eine erste Einführung

1. Stochastischer und deterministischer Trend

Die Zeitreihe $t = 1, 2, \dots, T$

$$(1) \quad y_t = \beta_0 + \beta_1 t + u_t$$

ist trendstationär (= hat einen deterministischen Trend [hier linear]). Ein trendstationärer Prozess (TSP) ist nach Elimination des Trends (= nach "detrending") stationär. Dagegen ist

$$(2) \quad y_t = y_{t-1} + u_t$$

ein differenz(en) stationärer Prozess (DSP), d.h. er hat einen stochastischen Trend, der durch Differenzbildung (differencing) eliminiert wird².

$$(3) \quad \Delta y_t = y_t - y_{t-1} = u_t \quad (u_t \text{ ist ein Zufallsprozess [= white noise]})$$

Gl. 2 ist ein random walk *ohne* drift mit $E(y_t) = 0$, dagegen ist

$$(2a) \quad y_t = y_{t-1} + c + v_t \quad (v_t \text{ ist white noise})$$

ein random walk *mit* drift (constant c mit $E(y_t) \neq 0$).

Wegen $E(y_t) = 0$ und $V(y_t) = t\sigma_u^2$ bei Gl. 2 (wenn $E(u_t) = 0$ und $V(u_t) = \sigma_u^2$ für alle t) ist der DSP gem. Gl. 2 mittelwertstationär, aber nicht varianzstationär (bei Gl. 2a gilt dagegen $E(y_t) = ct$ und $V(y_t) = c^2 t^2 + t\sigma_u^2$).

Ein Schock (eine "innovation") kann *bei einem random walk* wegen $y_t = y_0 + \sum u_\tau$ ($\tau \leq t$) *permanent* nachwirken, während er (sie) *bei einem TSP nur transitorisch* wirkt und sich bei einem autoregressiven Prozess erster Ordnung (= ein AR (1) Prozess) bei $\beta_1 < 0$ mit der Zeit ausdünnt (fading out). Ein AR(1) Prozess ist gegeben mit

$$(4) \quad y_t = \rho y_{t-1} + u_t$$

Er ist stationär mit $|\rho| < 1$ und es gilt $E(y_t) = 0$, $V(y_t) = \sigma_u^2 / (1 - \rho^2)$. Dagegen ist der random walk (Gl. 2, Fall $\rho=1$ =unit root) nicht stationär, vielmehr hat dieser DSP einen stochastischen Trend³. Man sagt dann er ist "integriert vom Grade 1" (abgekürzt I(1)), weil der differenzierte Prozess (Gl. 3) ein stationärer (oder I(0)) Prozess ist.

2. Spurious regression

Die Koeffizienten eines TSP können bei Geltung der Annahmen B1 – B3 (nach von Auer) für u_t konsistent geschätzt werden. Der t-Test, der zentrale Grenzwertsatz (central limit theorem) sowie die stochastische Konvergenz (law of large numbers) gelten. Das ist bei einem DSP nicht generell der Fall.

¹ Für eine kritische Durchsicht und wertvolle Hinweise und Ergänzungen danke ich Frau Dipl.-Ök. Nicole van de Locht, Frau Dipl. Kauffrau Janina Küter und besonders Herrn Claus Christian Breuer MSc. Und Herr Diplom Volkswirt Stephan Popp.

² Anders als bei Vogelvang sind Rückwärtsdifferenzen hier mit Δ und nicht mit ∇ (nabla) bezeichnet.

³ Optisch entsteht bei (2a) der Eindruck eines Trends, wie gilt $y_t = y_{t-1} + c + v_t = y_0 + ct + \sum v_i$ ($i = 1, 2, \dots, t$).

Weisen beide Variablen x_t und y_t in einer Regression einen stochastischen Trend (DSP) auf, so kann dies zu einem hohen R^2 führen, welches substantiell nicht gerechtfertigt ist. Der t-Test und F-Test sind nicht mehr valide, da sie fälschlich auf eine hohe Korrelation hinweisen und die Residuen häufig autokorreliert sind.

Man spricht von "Scheinregression" (spurious regression), weil die eigentlich vorhandene Unkorreliertheit (kein gemeinsamer Trend von x_t und y_t) wegen fehlerhafter (nicht valider) Tests nicht erkannt werden kann. Haben dagegen zwei I(1) Prozesse (mit stochastischem Trend) einen gemeinsamen Trend, dann sind sie "kointegriert". Es ist also wichtig die Existenz eines stochastischen Trends feststellen bzw. verwerfen zu können und danach gegebenenfalls auf Kointegration testen zu können.

3. Dickey-Fuller (DF) Test und Augmented Dickey-Fuller (ADF) Test

Von den Tests auf Einheitswurzel (unit-root) sind der DF und ADF Test die bekanntesten. Dickey und Fuller⁴ haben gezeigt, dass unter $H_0: \rho = 1$ die Teststatistik $t = (\hat{\rho} - 1) / \hat{\sigma}_{\hat{\rho}}$ nicht t-verteilt (auch nicht asymptotisch) ist⁵. Daher wird statt ρ in der Gleichung $y_t = \rho y_{t-1} + \beta_0 + \beta_1 t + u_t$ der Parameter $\theta = \rho - 1$ in der differenzierten Gl.

$$(5) \quad \Delta y_t = y_t - y_{t-1} = (\rho - 1) y_{t-1} + \beta_0 + \beta_1 t + u_t$$

geschätzt und die folgende Nullhypothese.

$$H_0: \theta = 0 \text{ (entspricht } \rho = 1) = \text{stochastischer Trend}^6 = \text{Einheitswurzel}^7 = \text{random walk}^8$$

gegen

$$H_1: \rho < 1 \text{ (oder } \theta < 0) = \text{kein stochastischer Trend} = \text{trendstationär}^9$$

getestet¹⁰,

($\rho > 1$ ist nicht sinnvoll, weil dies zu einer explosiven Entwicklung des AR (1) Prozesses führen würde. Ist u_t in Gl. 5 nicht white noise (reiner Zufallsprozess) empfiehlt es sich den ADF (Augmented DF) Test durchzuführen¹¹. In diesem Fall wird Gl. 5 erweitert um p Glieder $\alpha_i \Delta y_{t-i}$ (mit $i = 1, \dots, p$) zu

$$(5a) \quad \Delta y_t = \theta y_{t-1} + \beta_0 + \beta_1 t + \alpha_1 \Delta y_{t-1} + \alpha_2 \Delta y_{t-2} + \dots + \alpha_p \Delta y_{t-p} + v_t.$$

EViews bietet bei der Spezifikation folgende Optionen hinsichtlich β_0 und β_1 sowie was die maximale Lag-Länge p betrifft (diese Unterscheidungen sind notwendig, weil die Stichprobenverteilung der Prüfgröße in den drei Fällen unterschiedlich ist, was auch die Tab. 1 auf der nächsten Seite deutlich macht):

⁴ Dickey, D.; Fuller, W.: Distribution of the Estimators for Autoregressive Time Series with a Unit Root, Journal of the American Statistical Association, Vol. 74 (1979), S.427-431.

⁵ Wegen der Abweichung von der t-Verteilung würde die Nullhypothese bei Anwendung des herkömmlichen t-Tests (mit den kritischen Werten aus der t-Verteilung) zu oft abgelehnt. Dickey und Fuller lösten das Problem des zu häufigen Ablehnens der Nullhypothese anhand eines durch Monte-Carlo Simulation gewonnenen Satzes kritischer Werte (abhängig vom Signifikanzniveau und der Zahl T an Beobachtungen), vgl. Tab. 1.

⁶ oder DS, stationär nach Differenzenbildung

⁷ oder I(1) Prozess = DSP.

⁸ in Gl. 5 mit drift ($\beta_0 \neq 0$) und mit Trend ($\beta_1 \neq 0$).

⁹ In der Praxis kann es schwierig sein zwischen DS und TS zu differenzieren, es kann auch weder DS noch TS gelten, insbesondere bei langen Zeitreihen mit Strukturbrüchen, vgl. P. Newbold, S. Leybourne, M. E. Wohar, Trend-stationarity, difference stationarity, or neither: further diagnostic tests with an application to U.S. Real GNP, 1875 – 1993, Journal of Economics & Business, 53 (2001).

¹⁰ Man kann auch nach Dickey und Fuller mit einem modifizierten F-Test gemeinsam auf $\theta = \beta_0 = \beta_1 = 0$ testen.

¹¹ Wie ILS ist auch ADF für Flieger mit anderen Assoziationen verbunden; ADF (Automatic Direction Finder) ist der indicator bei einem NDB Navigationssystem (MW bzw. LW Funknavigation).

bei	Optionen		
Regressand	y_t : level	Δy_t : 1st difference	2nd difference (vgl. Gl. 6)
β Koeffizienten	intercept $\beta_0 \neq 0, \beta_1 = 0$ *	trend and intercept $\beta_0, \beta_1 \neq 0$	none $\beta_0 = \beta_1 = 0$
lag length (p)	automatic		user specified

* = random walk *with* drift $\beta_0 \neq 0$ ("none" = without drift = AR(1))

H_0 bedeutet bei unit-root tests: stochastischer Trend¹². Annahme von H_0 bedeutet, dass es *mindestens* eine Einheitswurzel gibt. Der Prozess könnte aber auch eine zweite Einheitswurzel haben, also $I(2)$ sein. In diesem Fall ist die mit OLS zu schätzende Prüfgleichung (Δ^2 heißt zweite Differenz)

$$(6) \quad \Delta^2 y_t = \tilde{\beta}_0 + \tilde{\theta} y_{t-1} + \sum_{i=1}^{p-1} \tilde{\alpha}_i \Delta^2 y_{t-1} + \tilde{u} \quad \text{mit } H_0: y_t \sim I(2) \text{ (oder } \tilde{\theta} = 0) \text{ und } H_1: y_t \sim I(1).$$

4. Durchführung des unit-root-tests nach Dickey und Fuller

a) **DF_t Tests** Unter der $H_0: \theta = 0$ (bzw. $\rho = 1$) existieren zwei Vorgehensweisen. Zum einen der Vergleich der t-Statistik $DF_t = \frac{\hat{\rho} - 1}{\hat{\sigma}_{\hat{\rho}}}$ mit den kritischen Werten der Tab. 1.

Tabelle 1: Kritische Werte des Dickey-Fuller DF_t Tests:
T = Zahl der Beobachtungen α = Signifikanzniveau

	T = 25	T = 50	T = 100	T $\rightarrow \infty$
α	AR Model ("none")			
0.01	-2.66	-2.62	-2.60	-2.58
0.025	-2.26	-2.25	-2.24	-2.23
0.05	-1.95	-1.95	-1.95	-1.95
0.10	-1.60	-1.61	-1.61	-1.62
0.975*	1.70	1.66	1.64	1.62
α	AR Model mit Achsenabschnitt (with intercept)			
0.01	-3.75	-3.59	-3.50	-3.42
0.025	-3.33	-3.23	-3.17	-3.12
0.05	-2.99	-2.93	-2.90	-2.86
0.10	-2.64	-2.60	-2.58	-2.57
0.975*	0.34	0.29	0.26	0.23
α	AR Model mit Achsenabschnitt und Trend (trend and intercept)			
0.01	-4.38	-4.15	-4.04	-3.96
0.025	-3.95	-3.80	-3.69	-3.66
0.05	-3.60	-3.50	-3.45	-3.41
0.10	-3.24	-3.18	-3.15	-3.13
0.975*	-0.50	-0.58	-0.62	-0.66

* positive und negative Vorzeichen ergeben sich aus der Schiefe der Stichprobenverteilung

Quelle: Fuller, W.: Introduction to Statistical Time Series (2nd ed.), New York (1996).

Wie man sieht sind die kritischen Werte im Allgemeinen größer als bei der t-Verteilung.

Die Nullhypothese des Vorhandenseins mindestens einer Einheitswurzel wird abgelehnt, wenn DF_t kleiner als der kritische Wert ist, die Gegenhypothese, dass der Prozess keinen stochastischen Trend

¹² Der ADF Test nicht sehr mächtig. Tendenz H_0 anzunehmen auch dann, wenn nicht gerechtfertigt ($1 - \beta$, die Macht [β bezieht sich hier auf den β -Fehler] ist gering). Ist ρ in Gl. 5a (zu) groß wird die Macht des Tests gering, ist ρ dagegen (zu) klein, sind die Tests wegen der Autokorrelation der Residuen nicht mehr korrekt anwendbar.

beinhaltet, wird dann angenommen. Für viele Anwendungen ist es nicht adäquat einen einfachen Random Walk Prozess als zugrunde liegende Regression der Zeitreihe anzunehmen. Wird z.B. die Zeitreihe einer Inflationsrate oder eines Wechselkurses betrachtet so müssen wir unter Annahme der Gültigkeit der Nullhypothese von einem Random Walk mit Drift (bzw. unter Gültigkeit der Gegenhypothese von einem AR Model mit Achsenabschnitt) ausgehen und ziehen die entsprechenden kritischen Werte aus Tabelle 1 zum Vergleich heran. Andere Zeitreihen könne unter der Nullhypothese trendstationär (bzw. unter der Gegenhypothese ein AR Model mit Achsenabschnitt und Zeittrend) seien. Ein Beispiel dafür wäre die Zeitreihe des Bruttoinlandsprodukts. In diesem Fall müssten die in der dritten Gruppe der Tabelle 1 aufgeführten kritischen Werte genutzt werden.

b) DF_ρ Tests Der zweite Vorschlag für eine DF-Teststatistik ist die Statistik $DF_\rho = T(\hat{\rho} - 1)$ mit den kritischen Werten der folgenden Tab. 2.

Tabelle 2: Kritische Werte des Dickey-Fuller DF_ρ Tests:

	T = 25	T = 50	T = 100	T $\rightarrow \infty$
α	AR Model			
0.01	-11.8	-12.8	-13.3	-13.8
0.025	-9.3	-9.9	-10.2	-10.5
0.05	-7.3	-7.7	-7.9	-8.1
0.10	-5.3	-5.5	-5.6	-5.7
α	AR Model mit Achsenabschnitt			
0.01	-17.2	-18.9	-19.8	-20.7
0.025	-14.6	-15.7	-16.3	-16.9
0.05	-12.5	-13.3	-13.7	-14.1
0.10	-10.2	-10.7	-11.0	-11.3
α	AR Model mit Achsenabschnitt und Trend			
0.01	-22.5	-25.8	-27.4	-29.4
0.025	-20.0	-22.4	-23.7	-24.4
0.05	-17.9	-19.7	-20.6	-21.7
0.10	-15.6	-16.8	-17.5	-18.3

Quelle: Fuller, W.: Introduction to Statistical Time Series (2nd ed.), New York (1996).

c) Ergänzungen

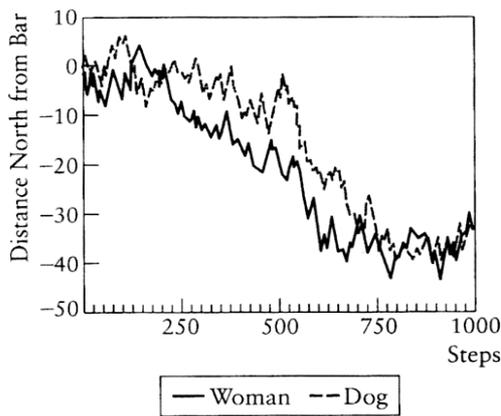
Im **Phillips-Perron-Test** (als Alternative zum DF-Test) wird die Prüfgröße t beim Test auf $H_0: \theta = 0$ bzw. F bei $H_0: (\beta_0, \beta_1, \theta) = (\beta_0, 0, 0)$ modifiziert mit der korrigierten (= gewichteten) Varianz der Störgröße u aus der Regression Gl. 5.

Im Unterschied zu Einheitswurzeltests ist bei **Stationaritätstests** die H_0 nicht $\theta = 0$ (oder äquivalent $\rho = 1$, d.h. ein [nichtstationärer] stochastischer Trend), sondern die Existenz eines (trend) stationären Prozesses.

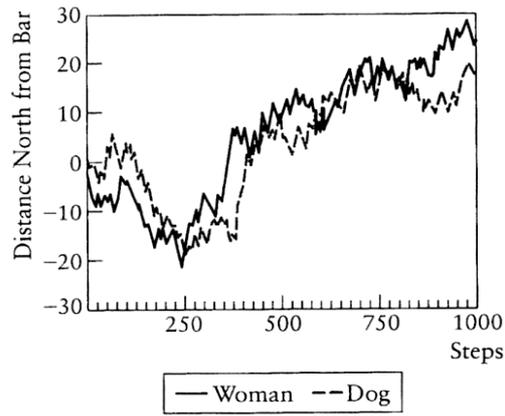
Die Ergebnisse des ADF- bzw. Phillips-Perron Test und die des Stationaritätstests können sich widersprechen (Kirchgässner/Wolters, S. 161 f). Die Widersprüche können sich auflösen, wenn nicht mit ganzzahligen Integrationsgraden d etwa $d = 0$ oder $d = 1$ in $I(0)$ und $I(1)$ gearbeitet wird, sondern mit $0 \leq d \leq 1$ (**fraktionale Integration**). Fraktional integrierte Prozesse (ARFIMA) sind für $d \geq 0,5$ nichtstationär.

5. Kointegration (Testverfahren nach Engle und Granger)

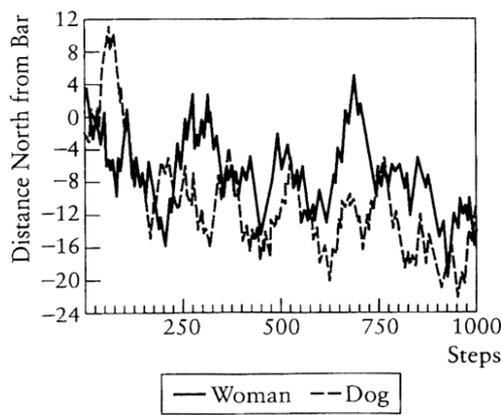
Zwei nichtstationäre (genauer $I(1)$ oder DSP) Zeitreihen x_t und y_t können kointegriert sein, d.h. sie haben langfristig einen gemeinsamen Trend. Den hierzu führenden Vorgang kann man als error correction model (ECM) beschreiben: Beispiel der betrunkenen Frau mit ihrem Hund (Murray, Econometrics, S. 790f).



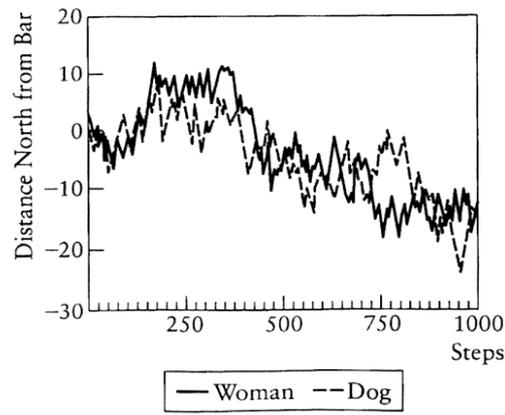
PANEL A



PANEL B



PANEL C



PANEL D

Woman (Y)
 $\Delta Y_t = Y_t - Y_{t-1}$
 $= \varepsilon_t + \alpha_1 D_{t-1}$

Dog (Z)
 $\Delta Z_t = Z_t - Z_{t-1}$
 $= \omega_t + \alpha_2 D_{t-1}$

Mit
 $D_{t-1} = Y_{t-1} - Z_{t-1}$

Four sample paths (A, ..., D) of 1,000 steps.

ECM for the changes in cointegrated variables Y and Z

"we see both the increasingly varying meandering of the drunk and dog..., and their propensity to stay roughly together"

"If you find the drunk, the dog is unlikely to be far away"

The distance between the two paths (Y, Z) contains no stochastic trend (distance or any linear combination of Y and Z)

Das allgemeinere **Fehlerkorrekturmodell** (ECM) nach der folgenden Gl. 7 lautet

$$(7) \quad \Delta y_t = \lambda_Y + \alpha_Y u_{t-1} + \varepsilon_t \quad \text{und} \quad \Delta x_t = \lambda_X + \alpha_X u_{t-1} + \omega_t \quad (\text{mit } u_t \text{ gem. Langfristbeziehung Gl. 8})$$

mit den Reaktionsgeschwindigkeiten a_Y und a_X bzw. a_1 und a_2 . Wie beim Übergang von DF zu ADF kann auch hier zur Verringerung der Autokorrelation die Gleichung ergänzt werden durch verzögerte endogene Variablen, etwa $\Delta y_t = \lambda_Y + \alpha_Y u_t + \lambda_1 \Delta y_{t-1} + \lambda_2 \Delta x_{t-1} + \varepsilon_t$ und Δx_t entsprechend. Nach dem Repräsentationstheorem von Engle u. Granger gilt: sind x und y kointegriert, dann können sie auch mit einem ECM dargestellt werden und umgekehrt gibt es eine ECM Darstellung liegt Kointegration vor.

Definition der **Kointegration**:

Ist $y_t \sim I(1)$ und $x_t \sim I(1)$, dann sind y_t und x_t kointegriert, wenn in

$$(8) \quad y_t = \beta x_t + u_t \quad \text{oder} \quad y_t = c + \beta x_t + u_t \quad (\text{"kointegrierenden Relation" oder static relation}^{13}) \quad \text{die Störgröße } u_t \text{ stationär ist (also } u_t \sim I(0)).$$

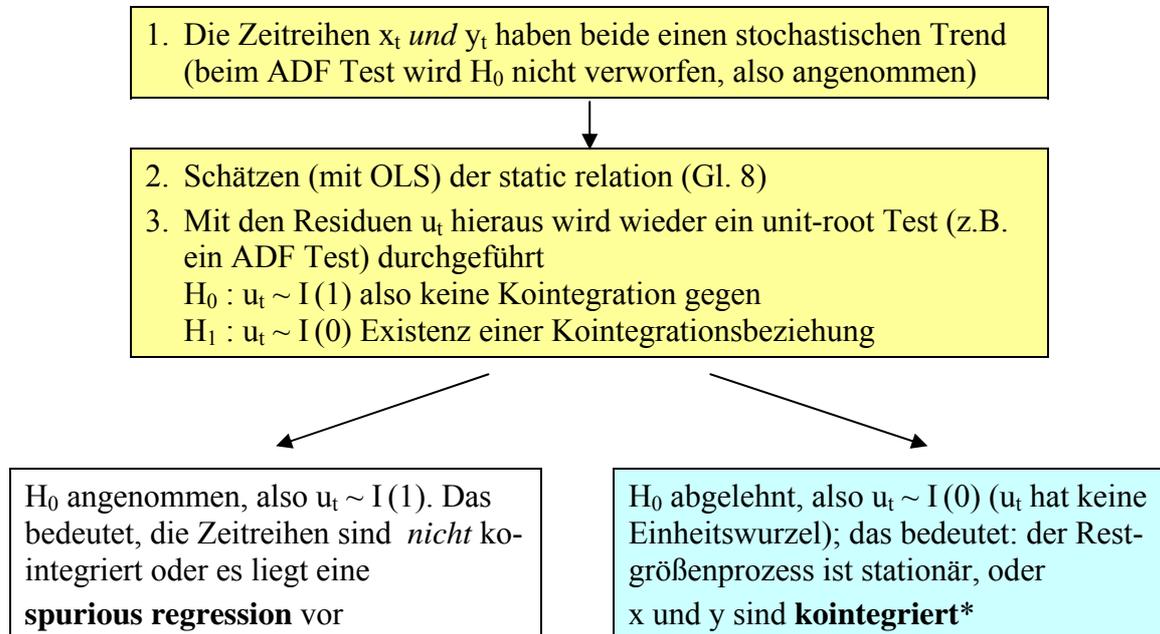
Kointegration bedeutet dass, es eine lineare Kombination von zwei nichtstationären Zeitreihen geben kann, die stationär ist, weil die Reihen einem gemeinsamen Trend folgen.

¹³ Oder long-run equilibrium relationship.

Allgemeiner Begriff:

Ist $y_t \sim I(d)$ und $x_t \sim I(d)$, dann sind x und y kointegriert, abgekürzt CI (d, b) , wenn $u_t = y_t - \beta x_t \sim I(d - b)$ mit $b > 0$.

Vorgehensweise in **drei Stufen nach Engle und Granger**



* there is a long-run equilibrium relationship among the variables, they share a common stochastic trend

Es werden also wiederholt (A)DF Tests auf Einheitswurzel durchgeführt, wobei in Stufe 1 die Annahme von H_0 und in Stufe 3 die Ablehnung von H_0 im Vordergrund steht.

Es ist üblich, Stufe 2 und 3 auch für die Residuen der umgekehrten Regression von (8), d.h.

$$(8a) \quad x_t = \tilde{c} + \tilde{\beta} y_t + \tilde{u}_t$$

zu betrachten und $u_t \sim I(0)$ und $\tilde{u}_t \sim I(0)$ für Kointegration zu fordern.

Bei mehr als zwei Variablen, etwa bei y_t , x_t und z_t ist der hier beschriebene Ansatz nur möglich, wenn zwischen den drei Variablen eine und nur eine Kointegrationsbeziehung besteht. Es ist dann eine andere Vorgehensweise angebracht, bei der von vektorautoregressiven (VAR) Modellen und ihnen entsprechenden Fehlerkorrekturmodellen (ECM) auszugehen ist.