

No. 63

Coïntegratie en foutencorrectiemodellen

A. ten Cate, D.A.G. Draper

Centraal Planbureau  
Van Stolkweg 14  
2585 JR 's-GRAVENHAGE  
telefoon: 070-3514151.  
telefax : 070-3505847

November 1989

## Abstract

Should econometric relations be specified in levels or in first differences? In this paper some recent literature about this subject is discussed, with an informal introduction to the theory of cointegration. Furthermore the Engle-Granger two step estimation procedure for cointegrated series is applied to an equation for employment in the construction industry.

INHOUDSOPGAVE	pag.
Samenvatting	1
1. Enige recente ontwikkelingen in de econometrie	2
1.1 Inleiding	2
1.2 Stationaire en geïntegreerde reeksen	2
1.3 Regressie op trendmatige reeksen: Granger en Newbold	4
1.4 Remedie: regressie op verschillen	5
1.5 Recent onderzoek: Phillips	6
1.6 Recent onderzoek: Stock	6
1.7 Recent onderzoek: Engle en Granger	9
1.8 Besluit	12
2. De werkgelegenheid in de bouw: een toepassing	13
2.1 Inleiding	13
2.2 De economische theorie	13
2.3 Toetsen op integratie	14
2.4 De lange termijn: coïntegratie	15
2.5 De korte termijn	19
Literatuurverwijzingen	22
Bijlage A Afleidingen behorende bij paragraaf 1	24
Bijlage B Het "autoregressive distributed lag" model (AD-model); enige kanttekeningen	27
Bijlage C Een meer formele beschrijving van integratie, coïntegratie en de Engle-Granger schattingsprocedure	32
Bijlage D Toetsen op integratie en coïntegratie	35

Samenvatting

Moeten econometrische schattingsrelaties in niveau of in (procentuele) mutatie worden geformuleerd? Over deze vraag is jarenlang gediscussieerd. Voorstanders van mutatie-vergelijkingen wezen op het gevaar van oneigenlijke ("spurious") correlatie, terwijl voorstanders van niveau-vergelijkingen betoogden dat mutatie-vergelijkingen over het algemeen de lange termijn relaties die de economische theorie suggereert, niet adequaat kunnen weergeven.

In dit onderzoeksmemorandum wordt enige recente literatuur op dit terrein besproken, op een informele manier. Een centrale plaats wordt daarbij ingenomen door het begrip coïntegratie: trendmatige niveau-grootheden waartussen een stabiele lange termijnrelatie bestaat, heten gecoïntegreerd.

Vervolgens wordt een in de literatuur voorgestelde schattingsprocedure toegepast op de werkgelegenheidsvergelijking in de bouwnijverheid. Het betreft hier een schattingsprocedure in twee stappen: eerst wordt de lange termijnrelatie geschat, vervolgens wordt daaraan het korte termijngedrag toegevoegd in de vorm van een foutencorrectiemodel.

Paul A. de Jongh voerde de regressie-berekeningen uit.

## 1. Enige recente ontwikkelingen in de econometrie

### 1.1 Inleiding

Onderstaand wordt enig recent theoretisch onderzoek besproken betreffende regressie-analyse met tijdreeksen. Allereerst wordt daartoe in paragraaf 1.2 kort ingegaan op stochastische tijdreeksen in het algemeen.

Vervolgens wordt in paragraaf 1.3 een niet meer zo recent stukje onderzoek besproken - betreffende "nonsens-correlatie" met niveau-reeksen - dat echter van groot belang is voor het meer recente onderzoek, dat daarna, in de paragrafen 1.5 en verder aan de orde komt, nadat in paragraaf 1.4 regressie op eerste verschillen besproken wordt. In paragraaf 1.8 wordt een conclusie gegeven.

### 1.2 Stationaire en geïntegreerde reeksen

Alvorens op econometrische literatuur in te gaan, eerst nog enige algemene opmerkingen over stochastische reeksen. Een stationaire stochastische reeks slingert zich rond een evenwichtswaarde. Bijvoorbeeld, beschouw de reeks  $z_t$  die gegenereerd wordt volgens een stationair AR(1)-patroon:

$$z_t = \rho z_{t-1} + c + \epsilon_t \quad (1)$$

Hierbij zijn de  $\epsilon_t$  onafhankelijk en identiek verdeeld met verwachting nul ("witte ruis"); dit geldt overigens voor alle  $\epsilon_t$  in deze notitie, tenzij anders vermeld. De letters  $r$  en  $c$  stellen vaste getallen voor. We trekken links en rechts  $z_{t-1}$  af van (1), en krijgen:

$$\begin{aligned} \Delta z_t &= -(1-\rho)z_{t-1} + c + \epsilon_t = \\ &= -(1-\rho)[z_{t-1} - c/(1-\rho)] + \epsilon_t \end{aligned} \quad (2)$$

In woorden: de reeks  $z_t$  krijgt iedere periode een "duw" ter grootte van  $\epsilon_t$ , maar corrigeert zijn afwijking in de vorige periode van de evenwichtswaarde  $c/(1-\rho)$  met een factor  $1-\rho$ . Stationairiteit wordt verkregen door de eis  $0 < 1-\rho < 2$ , oftewel  $|\rho| < 1$ . Dan zal  $z_t$  min of meer rond de evenwichtswaarde slingeren.

Men noemt (2) wel een "foutencorrectie-representatie" van (1), een term die we verderop weer tegen zullen komen; de "fout" is de waarde  $z_{t-1}-c/(1-\rho)$  en de "foutencorrectie" is  $-(1-\rho)$  maal deze fout.

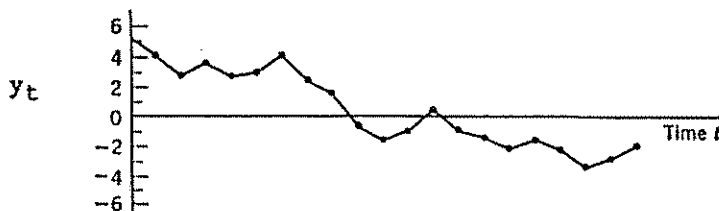
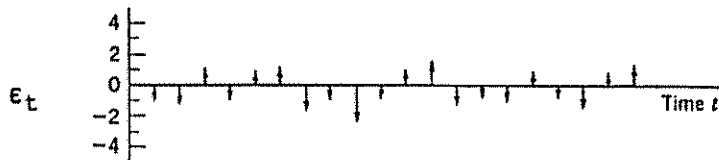
Een stochastische reeks  $y_t$  heet geïntegreerd als de mutatie van  $y_t$  stationair is (terwijl  $y_t$  zelf niet-stationair is). Ter verklaring van het woord "geïntegreerd": in continue tijd kan je zeggen dat een geïntegreerde reeks de integraal over de tijd is van een stationaire reeks.

De mutatie van een geïntegreerde reeks is onafhankelijk van het niveau van de reeks ("geheugenloos"). Daarom is de kans verwaarloosbaar klein dat zo'n reeks rond een evenwichtswaarde blijft slingeren, ook als  $c=0$ : een geïntegreerde reeks zal dus (bijna) altijd een soort "trendmatig" verloop vertonen, waarbij eventueel stijgende en dalende gedeelten elkaar kunnen afwisselen. Denk bij geïntegreerde reeksen aan de niveaus van economische grootheden.

Een reeks  $y_t$  waarvoor geldt dat

$$\Delta y_t = z_t, \tag{3}$$

waarbij  $z_t$  zich gedraagt zoals boven omschreven, is geïntegreerd, omdat  $z_t$  dan stationair is. We beschouwen nu het eenvoudige geval waarbij  $\rho = 0$  en  $c = 0$ . Dan geldt dus  $\Delta y_t = \epsilon_t$ , oftewel  $y_t = y_{t-1} + \epsilon_t$ : een z.g. "random walk", zonder drift ( $c = 0$ ). Als voorbeeld hiervan geven we onderstaande bewerking van figuur 6-2 uit Wonnacott en Wonnacott (1970), waarbij de bovenste reeks normaal verdeelde witte ruis is ( $\epsilon_t$ ) en de onderste reeks de daaruit berekende random walk  $y_t$ , gestart met  $y_0 = 5$ . We zien hierin duidelijk het "trendmatige" karakter van  $y_t$ . (Overgenomen met toestemming van de uitgever.)



### 1.3 Regressie op trendmatige reeksen: Granger en Newbold

In 1974 publiceerden Granger en Newbold een rekenoefening betreffende regressieanalyse, die veel aandacht heeft gekregen. Recentelijk zijn in dit verband weer enige publikaties verschenen, die we hier willen bespreken, met enig eigen puzzelwerk.

Granger en Newbold (1974) onderzochten het verschijnsel van "nonsenscorrelatie": het optreden van (schijnbaar) significante resultaten bij regressie van onafhankelijke, maar "trendmatig" verlopende, niveaureeksen op elkaar. Daartoe begonnen zij met het herhaald genereren, met een random-generator, van een tweetal reeksen,  $x$  en  $y$ , ieder 50 waarnemingen lang, en beide een "random walk" zonder drift.

Bij ieder aselekt gegenereerd reeksenpaar  $(x,y)$  voerden Granger en Newbold een gewone-kleinste-kwadraten-regressie van  $y$  op  $x$  uit, volgens  $y_t = \alpha + \beta x_t + \epsilon_t$ , en berekenden de  $t$ -ratio van de hellingscoëfficiënt. Het bleek dat in ruim drie kwart van de gevallen deze  $t$ -ratio in absolute waarde boven de 2 uit kwam: dus significant bij de gebruikelijke 95% betrouwbaarheid. In ruim de helft van de gevallen kwam de  $t$ -ratio in absolute waarde zelfs boven de 4 uit.

Per definitie mag bij een betrouwbaarheid van 95% slechts in 5% van de gevallen een significant resultaat optreden wanneer de werkelijke regressiecoëfficiënt nul is. Echter, de daarbij behorende kritieke  $t$ -waarde is afgeleid onder de veronderstelling van onafhankelijke storingen. Terwijl hier iets heel anders aan de hand is. Omdat hier in werkelijkheid  $\alpha = 0$  en  $\beta = 0$ , is de storingsterm  $\epsilon_t$  geheel gelijk aan  $y_t$  zelf en dus een geïntegreerde reeks.

Intuïtief is het resultaat van Granger en Newbold (1974) wellicht het beste te begrijpen, wanneer we ons het "trendmatige" karakter van een geïntegreerde reeks voor ogen halen. Het is voorstelbaar dat de regressie van twee zulke "trendmatige" reeksen op elkaar tot iets significantes leidt.

Het belang van een oefening als deze is gelegen in de waarschuwing die zij bevat: wees voorzichtig met de resultaten van de regressie van trendmatige tijdreeksen op elkaar; een naar de gebruikelijke maatstaven significante coëfficiënt behoeft niet te duiden op een causale samenhang tussen de reeksen.

Tenslotte zij hier gewezen op de bekende Durbin-Watson-test: het quotiënt van de kwadratensom van de eerste verschillen van het regressie-residu en de kwadratensom van de residuen zelf. Deze test kan een aanwijzing geven voor de aanwezigheid van een "nonsens-correlatie". In het geval van de rekenexercitie van Granger en Newbold (1974) geeft de DW-test een zeer lage waarde, corresponderend met hoge autocorrelatie. We komen later nog op de DW-test terug.

#### 1.4 Remedie: regressie op verschillen

Alvorens in te gaan op enige recente literatuur op het terrein van regressie met tijdreeksen - in het bijzonder geïntegreerde tijdreeksen - bespreken we een voor de hand liggende remedie tegen het probleem van "nonsens-correlatie": het regresseren op eerste (of hogere) verschillen. Deze remedie heeft echter een belangrijk nadeel: onder meer Davidson, Hendry, Srba en Yeo (1978, p.680) en Hendry, Pagan en Sargan (1984, p.1044) betogen dat door het nemen van eerste verschillen informatie verloren kan gaan die tot onderschatting van het verband tussen de reeksen kan leiden.

We zullen dit nadeel van het nemen van eerste verschillen illustreren aan de hand van de volgende regressievergelijking. (Bovengenoemde auteurs geven geen formele afleiding.) Deze regressievergelijking zal verderop ook als voorbeeld gebruikt worden; in bijlage B wordt nader op deze vergelijking ingegaan. De vergelijking luidt als volgt:

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + \epsilon_t \quad (4)$$

Het lange-termijn effect van x op y wordt verkregen door in (4) x en y constant te veronderstellen en de storing weg te laten:

$$y = ax + b \quad (5)$$

met

$$a = (\alpha_1 + \alpha_2) / (1 - \alpha_3) \quad (6)$$

en

$$b = \alpha_0 / (1 - \alpha_3) \quad (7)$$

Het argument van genoemde auteurs is gebaseerd op het feit dat onder bepaalde omstandigheden een regressie volgens vergelijking (5) leidt tot een asymptotisch raak resultaat voor de geschatte a en b, terwijl regressie van  $\Delta y$  op  $\Delta x$  leidt tot een coëfficiënt die (gemiddeld) kleiner is dan a. We gaan daar nu wat nader op in.



De regressie volgens (5) komt verderop aan de orde; zie paragrafen 1.6 en 1.7. De onderschatting van de coëfficiënt  $a$  bij de regressie met eerste verschillen is als volgt intuïtief in te zien. Vanwege het in de tijd gespreide karakter van de doorwerking van fluctuaties van  $x$  in fluctuaties in  $y$ , zal in het algemeen de variantie van  $\Delta y$  kleiner zijn dan die van  $a\Delta x$ . Dit leidt tot een schatter  $\hat{a}$  die kleiner is dan  $a$ . Immers

$$\hat{a} = \frac{\text{covar}(\Delta y, \Delta x)}{\text{var}(\Delta x)} = R \sqrt{\frac{\text{var}(\Delta y)}{\text{var}(\Delta x)}} < R \sqrt{\frac{\text{var}(a\Delta x)}{\text{var}(\Delta x)}} = Ra < a, \quad (8)$$

waarbij  $R$  de correlatiecoëfficiënt is tussen  $\Delta x$  en  $\Delta y$ . (We veronderstellen positieve  $R$  en  $a$ .)

Dus, als we in eerste verschillen schatten, missen we de mogelijkheid om met voorbijzien van de korte termijn dynamiek, het lange termijn gedrag goed te schatten.

#### 1.5 Recent onderzoek: Phillips

Zoals gezegd, recentelijk zijn er diverse onderzoeken gepubliceerd betreffende regressie met geïntegreerde reeksen. Phillips (1986) heeft o.a. analytisch het gedrag van regressieuitkomsten onder de in paragraaf 1.3 beschreven Granger-Newbold-omstandigheden afgeleid.

Het blijkt dat de correcte kritieke waarde voor de  $t$ -ratio hier inderdaad niet het getal 2 is, maar een getal dat groter wordt naarmate de reeksen langer worden: hoe langer de reeksen, des te "significanter" is het resultaat.

Voorts blijkt dat de Durbin-Watson-waarde naar nul gaat als de reekslengte naar oneindig gaat, terwijl de kansverdeling van de  $R^2$  naar een "nette" limietverdeling gaat (met alle massa tussen nul en één).

Conclusie: let niet alleen op  $t$ -ratio's en  $R^2$ , maar ook op de Durbin-Watson.

#### 1.6 Recent onderzoek: Stock

Stock (1987) heeft een analyse gemaakt van de regressie van geïntegreerde reeksen op elkaar, uitgaande van de veronderstelling dat een lineaire combinatie  $z_t$  van die reeksen stationair is. (Die  $z_t$  is wat in de regressie-terminologie de storing genoemd wordt.) Dergelijke reeksen heten gecoïntegreerd. Stock toonde aan dat met OLS de coëfficiënten van die lineaire combinatie consistent geschat kunnen worden.

Merk hierbij op dat dit resultaat onafhankelijk is van de keuze van wat er links en wat er rechts staat in de OLS-regressie-vergelijking. Bij de theorie van de regressie-analyse met gecoïntegreerde reeksen worden de betrokken reeksen symmetrisch behandeld, zonder onderscheid tussen verklarende en te verklaren variabelen.

Het is hierbij in zekere zin gunstig dat de variabelen geïntegreerd zijn: door hun trendmatig verloop vertonen zij een steeds grotere spreiding naarmate de reeksen langer zijn, en zoals bekend wordt de variantie van de schatting dan kleiner. Beschouw bijvoorbeeld de variantie van de hellingscoëfficiënt bij enkelvoudige regressie,

$$s^2 = \sigma^2 / \sum x_t^2, \quad (9)$$

waarbij  $x_t$  in afwijking van het gemiddelde is genomen. Hierin is  $\sigma^2$  de variantie van de storingsterm. (In matrixalgebra is dit de voor sommigen meer vertrouwde uitdrukking  $\sigma^2(X'X)^{-1}$ .) Als  $x$  stationair is, leidt een verdubbeling van de reekslengte tot ruwweg een verdubbeling van  $\sum x_t^2$  en dus tot een halvering van  $s^2$ . Echter, als  $x$  trendmatig verloopt, leidt een verdubbeling van de reekslengte tot een meer dan verdubbeling van  $\sum x_t^2$  omdat dan niet alleen het aantal termen in  $\sum x_t^2$  verdubbelt, maar het gemiddelde van  $x_t^2$  ook toeneemt.

Stock (1984) laat zien dat bij regressie met gecoïntegreerde reeksen de convergentiesnelheid van de orde  $T$  is, in plaats van van de orde  $\sqrt{T}$  bij regressie met stationaire reeksen ( $T$  = reekslengte).

Overigens: een wonderlijke consequentie van het voorgaande is dat met het voortschrijden van de jaren, de residuele variantie gedeeld door het gemiddelde niveau van de betrokken reeksen, steeds kleiner wordt.

Op basis van deze uitkomst kan worden aangetoond dat de parameters  $a$  en  $b$  van vergelijking (5) consistent geschat kunnen worden met OLS volgens (5), wanneer in feite (4) het data-genererend proces is, en  $x$  geïntegreerd is. We zullen daartoe eerst aantonen dat het residu van (5) stationair is. We definiëren daartoe, overeenkomstig (5):

$$z_t \equiv y_t - (ax_t + b), \quad (10)$$

waarbij  $a$  en  $b$  gegeven zijn in (6) en (7). Er geldt nu

$$z_t = \alpha_3 z_{t-1} - [(\alpha_2 + \alpha_1 \alpha_3)/(1 - \alpha_3)] \Delta x_t + \epsilon_t . \quad (11)$$

Zie bijlage A. Wanneer  $|\alpha_3| < 1$ , dan is  $z_t$  een stationaire reeks, met verwachtingswaarde nul. (Bedenk dat  $\Delta x_t$  en  $\epsilon_t$  stationair zijn, met verwachtingswaarde nul.) Uit het ongerijmde volgt nu dat  $y_t$  niet-stationair is; immers als  $y_t$  wel stationair zou zijn, dan zou uit (10) volgen dat  $x_t$  ook stationair is, hetgeen niet zo is. Nu rest nog te bewijzen dat  $\Delta y_t$  stationair is, als volgt. Het valt eenvoudig in te zien dat

$$\Delta y_t = \alpha_1 \Delta x_t - (1 - \alpha_3) z_{t-1} + \epsilon_t . \quad (12)$$

We zien dat  $\Delta y_t$  een som is van drie stationaire reeksen (met verwachting nul) en zelf dus ook stationair. Dus,  $y_t$  is een geïntegreerde reeks en  $y_t$  en  $x_t$  zijn gecoïntegreerd als (4) correct is en  $x$  geïntegreerd is. Dit zijn overigens toetsbare condities.

Een probleem wordt hierbij gevormd door het geval dat een deelverzameling van de gecoïntegreerde reeksen ook gecoïntegreerd is. (Zie de clause bij theorema 2 van Stock, 1987.) Beschouw bijvoorbeeld het drietal variabelen  $x_1$ ,  $x_2$  en  $x_3$ . Veronderstel dat deze variabelen gecoïntegreerd zijn: afzonderlijk zijn ze geïntegreerd en een lineaire combinatie is stationair. Echter, stel dat tevens  $x_2$  gelijk is aan twee maal  $x_1$ , plus een stationair storinkje. De coëfficiënten van de stationaire lineaire combinatie van de drie reeksen zijn dan onbepaald. OLS-regressie om die coëfficiënten te vinden levert in dat geval weinig zinvol op en wat daaruit komt is afhankelijk van de keuze van de linkerkant-variabele. Dit is het bekende probleem van de multi-colineariteit. Een voor de hand liggende remedie is hier het verwijderen van  $x_1$  of  $x_2$ .

Dus, we weten nu dat als  $x_t$  geïntegreerd is, dan kunnen we het lange termijn-effect in vergelijking (4) asymptotisch raak schatten met OLS-regressie op niveaus, met voorbijzien aan de korte termijn dynamiek, volgens vergelijking (5). Dit is van belang voor onderzoek dat zich speciaal richt op de lange termijn. Zie ook de discussie over regressie op niveaus versus eerste verschillen, in paragraaf 1.4.

### 1.7 Recent onderzoek: Engle en Granger

Tenslotte bespreken we het onderzoek van Granger (1986) en Engle en Granger (1987). Het foutencorrectiemodel speelt hierbij een belangrijke rol. We kwamen in paragraaf 1.2 reeds een zeer eenvoudig foutencorrectiemodel tegen. Een foutencorrectiemodel is in het algemeen een vergelijking met in het linkerlid een verschil en in het rechterlid zowel verschillen als vertraagde niveaus. Foutencorrectiemodellen zijn in de belangstelling gekomen door het werk van Davidson e.a. (1978).

Onze voorbeeldvergelijking (4) krijgt de vorm van een foutencorrectiemodel wanneer we definitie (10) invullen in (12). Dit geeft

$$\Delta y_t = \alpha_1 \Delta x_t - (1 - \alpha_2)(y_{t-1} - \alpha x_{t-1} - b) + \epsilon_t \quad (13)$$

Terzijde zij hier gewezen op het feit dat de tot nu toe gegeven formules afgeleid zijn onder veronderstelling dat de evenwichtswaarde van  $\Delta x_t$  gelijk is aan nul; vergelijk  $c=0$  in paragraaf 1.2. In Appendix A is echter van het meer algemene geval uitgegaan. Er ontstaat dan een evenwichtswaarde voor  $\Delta x_t$  die niet nul is, zeg  $\mu$ . In de formules (11), (12) en (13) wordt dan  $\Delta x_t$  vervangen door  $(\Delta x_t - \mu)$ . Voorts komt er in (12) en (13) een constante term bij, die evenredig is met  $\mu$ . Tevens wordt de formule voor  $b$  (vergelijking (7)) iets ingewikkelder. Zie verder de Appendix.

Engle en Granger (1987) stellen voor om een vergelijking zoals onze voorbeeldvergelijking (4) in twee stappen te schatten. De eerste stap is het schatten van de lange termijn relatie met OLS op de niveaus, volgens (5). We zagen dat dit bij geïntegreerde  $x_t$  en stationaire storing  $\epsilon_t$  een goede schatting oplevert van het lange termijn gedrag van  $y_t$ , omdat dan  $x_t$  en  $y_t$  gecoïntegreerd zijn.

De tweede stap die zij voorstellen is het daar aan toevoegen van het korte termijn gedrag door (12) te schatten met OLS, waarbij  $z_{t-1}$  vervangen wordt door het vertraagde schattingsresidu van de eerste stap, overeenkomstig (10).

Je zou kunnen zeggen dat op deze manier het dilemma tussen het schatten in niveaus en het schatten in verschillen is opgelost: het gebeurt allebei. Merk op dat de coëfficiënten  $\alpha_0$ ,  $\alpha_1$ ,  $\alpha_2$  en  $\alpha_3$  hierbij afgeleid kunnen worden uit de schatting van  $a$ ,  $b$ ,  $\alpha_1$  en  $1-\alpha_2$ , met behulp van (6) en (7).

Men zou zich kunnen afvragen: waarom deze ingewikkelde tweestapsprocedure, in plaats van direct (13) schatten, met regressoren  $\Delta x_t$ ,  $x_{t-1}$ ,  $y_{t-1}$  en de constante. Hiervoor zijn een aantal redenen. (Zie echter ook Boswijk, 1989.)

Ten eerste: met de tweestapsprocedure kan men het zoeken naar de beste specificatie opsplitsen, en daardoor wellicht eenvoudiger maken: onderzoek eerst de lange termijn (stap één) en dan de korte termijn (stap twee). De analyse van de korte termijn betreft de specificatie van de vertragingen en - in het algemeen - het toevoegen van stationaire verklaarders. (Bedenk dat zulke verklaarders niet in de eerste stap kunnen voorkomen, omdat ze per definitie niet deel kunnen uitmaken van de verzameling geïntegreerde reeksen.) Ook: men kan op deze manier een model voor de lange termijn en een model voor de korte termijn maken, die met elkaar corresponderen.

Ten tweede: het blijkt dat als  $x_t$  inderdaad geïntegreerd is ("trendmatig" verloopt) en de reekslengte naar oneindig gaat, dat dan deze tweestaps-procedure erg nauwkeurig is: de asymptotische verdeling van de geschatte coëfficiënten in de tweede stap is gelijk aan de asymptotische verdeling wanneer we  $a$  en  $b$  (en dus  $z_t$ ) zouden kennen. Merk hierbij op dat in de tweestapsprocedure alle coëfficiënten geschat worden zonder dat men last heeft van multicollineariteit, terwijl bij directe schatting van (13) multicollineariteit tussen  $y_{t-1}$  en  $x_{t-1}$  optreedt.

Ten derde: de parameterschattingen in de eerste stap zijn ook nog consistent als de te schatten vergelijking deel uitmaakt van een simultaan stelsel (en de reeksen geïntegreerd zijn, uiteraard). Zie Engle en Granger (1987), p.260. Dit is als volgt intuïtief in te zien. We schrijven in dezelfde notatie als vergelijking (9) de fout in OLS-schatter  $\hat{\beta}$  als volgt:

$$\beta - \hat{\beta} = \sum x_t \epsilon_t / \sum x_t^2 \quad (14)$$

(In matrixalgebra is het rechterlid hiervan de voor sommigen meer vertrouwde uitdrukking  $(X'X)^{-1}X'\epsilon$ .) We beschouwen de omstandigheid dat  $x_t$  geïntegreerd is en  $\epsilon_t$  stationair. Als de reekslengte naar oneindig gaat, gaat de noemer van het rechterlid van (14) sneller naar oneindig dan de teller, en het quotiënt gaat dus naar nul. Dit is ook het geval als  $x_t$  en  $\epsilon_t$  gecorreleerd zijn, zoals in het geval van simultaneïteitsbias. Dit laatste is iets beter te zien als we uit (14) het volgende afleiden:

$$|\beta - \hat{\beta}| = (s_\epsilon / s_x) R_{x\epsilon} < s_\epsilon / s_x \quad (15)$$

Hierin zijn  $s_\epsilon$  en  $s_x$  de standaardfout van resp.  $\epsilon_t$  en  $x_t$ ;  $R_{x\epsilon}$  is de absolute waarde van de correlatie tussen  $\epsilon_t$  en  $x_t$ . Merk op dat als de reekslengte naar oneindig gaat, dan gaat  $s_x$  ook naar oneindig, terwijl  $s_\epsilon$  dat niet doet.

Overigens moet worden opgemerkt dat de eindige-steekproef eigenschappen van de eerste stap in deze procedure niet zo goed zijn: wanneer inderdaad vergelijking (4) correct is, dan is vergelijking (5) dus niet correct en zijn de OLS-schattingen volgens (5) geen zuivere schatters van de  $a$  en  $b$  in resp. (6) en (7). Dit wordt weerspiegeld door de term met  $\Delta x_t$  in het residu  $z_t$  in (11). Voorts zien we aan de eerste term van  $z_t$  in (11) dat het residu autoregressief is. Tenslotte, nog afgezien van het voorgaande: wanneer we rekening houden met het stochastische karakter van  $x_t$ , dan kunnen we de gangbare theorie over regressie-analyse en  $t$ -ratio's niet meer toepassen, want de kansverdeling van de schatters wordt dan ook bepaald door de kansverdeling van  $x_t$ . Zie ook Stock (1987).

Ten vierde en tenslotte: de mogelijkheid om met de geschatte reeks  $\hat{z}_t$  (het schattingsresidu van de eerste stap) te testen of het in de eerste stap om een "nonsens"-regressie gaat (zie paragraaf 1.3), of niet. Dit kan gedaan worden door te testen of  $\hat{z}_t$  stationair is of geïntegreerd. Als we met een "nonsens"-regressie bezig zijn is  $\hat{z}_t$  geïntegreerd. Van de diverse tests op stationairiteit die Engle en Granger voorstellen, bespreken we er drie: de Durbin-Watson (DW), de Dickey-Fuller (DF) en de uitgebreide Dickey-Fuller (ADF). In paragraaf 2 worden deze testen toegepast.

De Durbin-Watson-test wordt berekend als  $\sum(\Delta\hat{z})^2/\sum\hat{z}^2$ . Als het om een "nonsens"-regressie gaat, dan is  $z_t$  niet stationair, maar geïntegreerd, en zal  $\hat{z}^2$  groot zijn in verhouding tot  $(\Delta\hat{z})^2$ . De DW zal dan dus laag zijn. We bespreken dit reeds in de paragrafen 1.3 en 1.5.

De test van Dickey en Fuller is in zijn eenvoudigste vorm een regressie van  $\Delta\hat{z}$  op  $\hat{z}_{-1}$ , zonder constante. Als  $z$  geïntegreerd is (nonsens-regressie) dan is  $\Delta\hat{z}$  onafhankelijk van  $\hat{z}$  en de regressie levert dan dus geen significante coëfficiënt op. Als model (4) echter wel waar is, dan schat deze regressie de coëfficiënt  $\alpha_3-1$ ; zie vergelijking (11), waarbij de  $\Delta x$ -term in het residu terecht komt.

De uitgebreide ("augmented") test van Dickey-Fuller (ADF) is bedoeld voor het geval dat  $y_t$  meer dan één periode vertraagd voorkomt in het rechterlid. (Bij

onze voorbeeldvergelijking (4) komt alleen  $y_{t-1}$  voor). In dat geval komt  $z_t$  in vergelijking (11) ook met meerdere periode vertraagd voor en dit wordt getest door  $\Delta z_t$  te regresseren op  $z_{t-1}$  en  $\Delta z_{t-1}, \Delta z_{t-2}, \dots, \Delta z_{t-(p-1)}$ , waarbij  $p$  de hoogste vertraging is van  $y_t$  in de oorspronkelijke regressievergelijking. De coëfficiënt van  $z_{t-1}$  in de testregressie is een schatter van één minus de som van de coëfficiënten van de vertraagde  $y_t$ 's in de oorspronkelijke vergelijking.

Overigens kunnen deze toetsen ook gebruikt worden om te onderzoeken of de  $x_t$  en  $y_t$  inderdaad geïntegreerd zijn - en dus niet stationair. Eventueel kan onderzocht worden of de betrokken reeksen geïntegreerd zijn van een hogere orde dan één: of ze pas na meer dan één keer het verschil nemen, stationair worden.

### 1.8 Besluit

In het bovenstaande is een overzicht gegeven van recente econometrische literatuur; voor het merendeel rond het begrip coïntegratie. Er is gepoogd een brug te slaan tussen het begrip coïntegratie en de gangbare regressie-analyse. Daartoe is uitgegaan van een regressievergelijking met veronderstellingen omtrent de  $x$ -variabele en de storing. Uit deze veronderstellingen is afgeleid dat de  $x$ -reeks en de  $y$ -reeks gecoïntegreerd zijn, en er is aangegeven wat daaruit volgt voor het gebruik van de schattingsmethode OLS.

De hier gepresenteerde analyse is niet erg precies. Bijvoorbeeld: de reeks  $\Delta y_t$  volgens (12) is een som van drie stationaire reeksen. De definitie van (co)integratie bij Engle en Granger eist echter dat deze reeks zelf stationair is. (Voor een analyse van een som van ARMA-reeksen, zie Granger en Morris (1976).) Wij meenden echter dat een wat informele benadering zoals hier gevolgd, geschikt is als introductie tot de hier besproken literatuur. Het derde nummer van het Oxford Bulletin of Economics and Statistics (1986) is volledig aan het hier besproken onderwerp gewijd, met o.a. inleidende artikelen van Hendry en Granger.

## 2. De werkgelegenheid in de bouw: een toepassing

### 2.1 Inleiding

In deze paragraaf wordt het voorgaande toegepast op het schatten van een relatie voor de werkgelegenheid in de bouwnijverheid in Nederland. Nadat de betreffende economische theorie is besproken, wordt de hierboven besproken tweestaps procedure uitgevoerd, met jaargegevens. Het is de bedoeling om deze relatie in het nieuwe FK-model op te nemen. Een enigszins aangepaste versie is opgenomen in het sectormodel ATHENA waarover naar verwachting nog in 1989 een publicatie zal verschijnen.

### 2.2 De economische theorie

Uitgangspunt is de veronderstelling dat bouwondernemers bij gegeven prijzen en productieomvang hun winst maximeren, waarbij ze enige vrijheid hebben in de keuze van produktietechniek: men kan meer arbeidsintensief of meer kapitaalintensief produceren. Een bouwondernemer wordt geacht produktieve uren te vragen. De werkgelegenheid in arbeidsjaren kan hieruit worden berekend indien het aantal produktieve uren per arbeidsjaar bekend is. Verder wordt verondersteld dat er arbeidsbesparende technische vooruitgang plaats vindt en dat de arbeidsproduktiviteit in de bouw fluctueert met de werkloosheid. Dit laatste verschijnsel heeft twee oorzaken. In tijden van krapte op de arbeidsmarkt houdt men eventueel overbodig personeel langer aan terwijl men bij een ontspannen arbeidsmarkt sneller geneigd zal zijn mensen te ontslaan. De tweede reden voor dit verschijnsel is, dat de meest produktieve mensen het laatst ontslagen worden. Het winstmaximeringsprobleem waar de bouwondernemer zich voor geplaatst ziet, kan onder de voorgaande veronderstellingen formeel worden weergegeven als:

$$\max_{a_h, k} p \cdot y - a_h \cdot (L/h) - k \cdot p_k$$

$$\text{o.d.v. } y = A_e \cdot [\delta \cdot a^{-\beta} + (1 - \delta)k^{-\beta}]^{-1/\beta} \quad (16)$$

$$a_e = a_h \cdot e^{\rho t} \cdot e^{\gamma u}$$

$$a_h = a \cdot h$$



waarin:

- y produktie (bruto toegevoegde waarde factorkosten)
- p prijs produktie
- a totaal arbeidsvolume, dus loontrekkers plus zelfstandigen (arbeidsjaren)
- a<sub>e</sub> totaal arbeidsvolume gemeten in efficiëncy eenheden
- h aantal produktieve uren per arbeidsjaar
- ℓ loonsom per loontrekker
- k kapitaal
- p<sub>k</sub> kosten kapitaal
- u werkloosheid in de bouwnijverheid in procenten van het arbeidsaanbod
- t tijd

De arbeidsvraagrelatie kan worden afgeleid uit de eerste orde voorwaarde voor een maximum: de opbrengst van de produktie van de marginaal ingezette hoeveelheid arbeid is gelijk aan de kosten van deze eenheid. Dus:

$$p \left( \frac{\partial y}{\partial a_h} \right) = \ell/h \quad (17)$$

Hieruit volgt voor de totale gevraagde hoeveelheid arbeid:

$$a = A \frac{y}{h} \left[ \frac{p}{\ell/h} \right]^{1/(1+\beta)} e^{-\rho t - \gamma u} \quad (18)$$

Dit theoretisch model suggereert dat er sprake moet zijn van coïntegratie tussen de logaritmen van de werkgelegenheid, produktie, produktieve uren, het reële uurloon, en het niveau van de werkloosheid en een tijdtrend. Verder zouden de elasticiteiten van zowel de produktie als de produktieve uren ongeveer 1 moeten zijn.

Alvorens te toetsen op coïntegratie dient eerst nagegaan te worden of de orde van integratie voor alle variabelen gelijk is.

### 2.3 Toetsen op integratie

In tabel 1 staan de uitkomsten voor de DW-toets en DF-toets op integratie. Een significant van nul verschillende waarde (positief (DW); resp. negatief (DF)) duidt op het verwerpen van de nulhypothese van integratie, ten gunste van de alternatieve hypothese van stationariteit. Voor de niveaureksen kan de hypothese van integratie niet verworpen worden. Voor de verschilreeksen wel. Dus: reeksen zijn geïntegreerd van orde één. (Notatie: I(1).) De kritieke waarde

voor de DF-toets is zeer onzeker. Meer in het algemeen valt op te merken dat alle kritieke waarden van de in deze paragraaf gebruikte toetsen afhangen van het data-genererend proces onder de nulhypothese, en van de reekslengte. Daarom moeten de formele toetsen in deze paragraaf gezien worden als niet meer dan een indicatie. Voorts zij met betrekking tot tabel 1 nog opgemerkt dat ook ADF-toetsen zijn uitgevoerd. Daarbij bleken de vertraagde  $\Delta z$ -termen echter niet significant. (Deze termen maken het verschil uit tussen DF en ADF).

Tabel 1 Toetsen op integratie: DW<sup>a)</sup> en DF<sup>b)</sup>

Variabele	toets toegepast op			
	niveau		eerste verschil	
	DW	DF	DW	DF
Werkgelegenheid (ln a)	0,1	-1,0	1,1	-3,2
Productie (ln y)	0,1	-1,5	2,5	-8,1
Productieve uren (ln h)	0,2	-1,7	2,2	-5,9
Reële lonen (ln $l/ph$ )	0,1	-0,6	1,8	-5,9
Werkloosheid (u)	0,2	1,1	1,3	-3,5
Kritieke waarde <sup>c)</sup>	0,7	-3,0	0,7	-3,0

a) Regressies over de periode 1951-1984

b) Regressies over de periode 1954-1984

c) Verkregen uit tabel 1 in bijlage D, door interpolatie over reekslengten.

#### 2.4 De lange termijn: coïntegratie

Bij de coïntegratievergelijking doet zich een probleem voor. Volgens de formele definitie van coïntegratie (bijlage D), mogen alleen niet-deterministische variabelen worden opgenomen. Een tijdtrend is deterministisch en zou dus niet in aanmerking komen. Bedenk echter dat de discussie rond (9) in paragraaf 1.6 evenzeer op een tijdtrend van toepassing is. Hall (1986) wijst bovendien op analogie met een constante. Strikt genomen zou deze ook niet mogen worden opgenomen, maar in iedere coïntegratie vergelijking komt een constante voor. Blijkbaar is de definitie van coïntegratie niet in overeenstemming met de gangbare praktijk.

Op economisch-theoretische gronden wordt er slechts één coïntegratie-vergelijking verwacht tussen de hiervoor genoemde variabelen, namelijk de evenwichtsrelatie, die de werkgelegenheid verklaart uit de produktie, produktieve uren, reële uurloon, tijdtrend en werkloosheid. De werkloosheid is immers afhankelijk van de werkgelegenheid en het arbeidsaanbod; de produktie is vooral bepaald door de vraag; de produktieve uren zijn afhankelijk van het arbeidsvoorwaardenbeleid. Bij de reële lonen ligt het iets ingewikkelder. De theoretische relatie kan ook worden herschreven tot een beloningsvergelijking. Het reële loon dat een ondernemer bereid is te betalen, is gelijk aan de marginale produktiviteit van de werknemer. Indien de werknemers ook invloed kunnen uitoefenen spelen andere factoren een rol zoals de consumptieprijs, de macro arbeidsmarktsituatie, etc.

Een en ander betekent dat bij de beperkte dataset die hier beschouwd wordt slechts één coïntegratie-vergelijking wordt verwacht. Wordt de dataset uitgebreid met de hiervoor genoemde variabelen dan zullen waarschijnlijk meerdere relaties worden gevonden.

Tenslotte nog een opmerking over simultaneïteitsproblemen. De werkgelegenheid en werkloosheid zijn simultaan. Dit betekent dat volgens de gangbare theorie niet met OLS mag worden geschat. Echter, zoals reeds besproken is in paragraaf 1.7: Engle en Granger hebben aangetoond dat als de variabelen gecoïntegreerd zijn, OLS een consistente schatter oplevert.

Indien alle variabelen worden opgenomen is het regressieresultaat voor de evenwichtsvergelijking:

$$\ln a = 0,81 \ln y - 0,76 \ln h - 0,23 \ln \ell / (p \cdot h) - 0,006 t - 0,004 u + 1,9$$

(8,2)            (3,9)            (3,0)            (1,5)            (1,6)            (2,2)

$$\bar{R}^2 = 0,968 \quad DF = -3,5 \quad ADF = -3,4 \quad DW = 1,61 \quad (19)$$

RCO : 0,09 0,05 -0,12 -0,13 0,10 0,11 -0,11 -0,15 -0,28 -0,16  
Steekproef: 1952-1984

De autocorrelaties van de residureeks (RCO) en de DW-toets duiden op coïntegratie. De waarden van de DF en ADF toetsgrootte zijn evenwel te laag (zie bijlage D tabel 1). Verder zijn de t-waarden gegeven, aangezien de DW redelijk hoog is. (Indien de DW laag is, geeft men meestal geen t-waarden bij de coïntegratievergelijking, aangezien ze dan niet zijn te interpreteren). De werk-

loosheid blijkt niet significant te zijn. Verder blijkt een regressie met  $u$  in het linkerlid niet dezelfde coëfficiënten op te leveren. Dit regressie-resultaat herschreven in dezelfde vorm als (5.4) is:

$$\ln a = -0,7 \ln y + 1,9 \ln h + 0,3 \ln \ell/(p.h) + 0,05 t - 0,05 u + 0,3 \quad (20)$$

$$\bar{R}^2 = 0,94 \quad DF = -2,8 \quad ADF = -2,4 \quad DW = 0,86$$

RCO : 0,55 0,20 0,06 -0,03 -0,23 -0,31 -0,20 -0,22 -0,25 -0,27  
Steekproef: 1952-1984

De autocorrelaties dempen in dit geval niet snel uit. De toetsen duiden niet op coïntegratie. Verder hebben de meeste variabelen een niet interpreteerbaar teken. De conclusie is dan ook dat er een deelverzameling gecoïntegreerd moet zijn.

Het schattingsresultaat zonder werkloosheid is:

$$\ln a = 0,95 \ln y - 1,01 \ln h - 0,28 \ln \ell/(p.h) - 0,012 t + 2,1$$
$$\quad (23,6) \quad (9,0) \quad (3,8) \quad (7,2) \quad (2,4) \quad (21)$$
$$\bar{R}^2 = 0,96 \quad DF = -4,4 \quad ADF = -4,5 \quad DW = 1,99$$

RCO : -0,09 -0,09 -0,18 -0,09 0,13 0,08 -0,05 -0,04 -0,18 -0,05  
Steekproef: 1952-1984

De toetsen duiden nu op coïntegratie. Het super-consistentie resultaat van Stock geldt als alle variabelen in de statische regressie  $I(1)$  zijn en er geen deelverzameling van variabelen bestaat die gecoïntegreerd zijn. (Zie: theorema 2 van Stock, 1987). Dit laatste is nagegaan (tabel 2).

Het resultaat verslechtert in alle gevallen aanmerkelijk, vooral indien de produktie wordt weggelaten. De conclusie is dat alle variabelen in de coïntegratie-vergelijking moeten worden opgenomen. Een en ander betekent, op grond van het theorema van Stock, dat de geschatte coëfficiënten uit vergelijking (21) betrouwbaarder zijn dan de coëfficiënten van vergelijking (19). Er is in de bouw dus waarschijnlijk nauwelijks sprake van Verdoorn-effecten. De volgens het theoretisch model veronderstelde elasticiteiten van 1 voor de produktie en produktieve uren worden ondersteund door deze laatste regressie-uitkomst.

Zoals hiervoor is opgemerkt moet regressie in alle richtingen bij voldoende grote steekproef hetzelfde resultaat geven, wil er sprake zijn van coïnte-

gratie (tabel 3).

Tabel 2 Toets op mogelijke uitsluiting van variabelen in de coïntegratievergelijking

	uitgesloten variabele			
	ln y	ln h	ln $\ell/(p.h)$	t
CRDW	0,19	0,92	1,31	0,87
DF	-0,6	-2,3	-2,9	-2,2
ADF	-1,2	-3,2	-3,2	-2,7
R <sup>2</sup>	0,27	0,86	0,95	0,90
RCO1	0,78	0,50	0,33	0,43
2	0,57	0,12	0,20	0,11
3	0,41	-0,16	-0,06	-0,04
4	0,28	-0,29	-0,08	-0,17
5	0,21	-0,24	-0,08	-0,07
6	0,13	-0,04	-0,06	-0,03
7	0,08	-0,03	-0,11	-0,09
8	0,03	-0,10	-0,09	-0,08
9	-0,06	-0,15	-0,21	-0,04
10	-0,19	-0,20	-0,19	-0,04

Tabel 3 De effecten op de evenwichtsrelatie bij verandering van afhankelijke variabele<sup>a)</sup>

afhankelijke variabele	coëfficiënten				c	R <sup>2</sup>	ADF	DF
	ln y	ln h	ln $\ell/(p.h)$	t				
ln a	0,95	-1,01	-0,28	-0,012	2,10	0,96	-4,5	-4,4
ln y	1,00	-1,03	-0,31	-0,012	1,58	0,98	-3,8	-4,8
ln h	0,98	-1,36	-0,38	-0,013	3,76	0,96	-4,3	-4,8
ln $\ell/(p.h)$	1,06	-1,39	-0,81	-0,005	-0,17	0,94	-2,7	-3,2
Standaardfout	0,04	0,11	0,07	0,002	0,875			

a) De coëfficiënten zijn genormeerd op ln a als te verklaren variabele. De gegeven standaardfouten hebben op de regressie van ln a betrekking.

Uit de tabel blijkt dat de schattingen met ln a en ln y in het linkerlid redelijk overeenstemmen. Over het geheel zijn de resultaten van tabel 3 echter niet onverdeeld gunstig.

## 2.5 De korte termijn

Vergelijking (21) geeft dus een schatting van de lange termijnevenwichtsrelatie. Nu kan worden overgegaan tot de tweede stap van de Granger-Engle procedure, namelijk het schatten van de korte termijnrelatie. Dit leverde het volgende op (met  $\hat{z}$  het residu uit vergelijking (21)):

$$\Delta \ln a = 0,68 \Delta \ln y - 0,46 \Delta \ln h - 0,0055 \Delta u_{-1} - 0,73 \hat{z}_{-1} - 0,0058 \quad (22)$$

(16,7)            (9,3)            (5,7)            (6,9)            (2,2)

$$R^2 = 0,92 \quad DW = 1,67 \quad \sqrt{LM1} = 1,1 \quad SSR = 0,0050$$

Steekproef: 1953-1984

De toetsgrootte LM1 (waarvan de vierkantswortel is gegeven) is de Lagrange Multipliertoets op eerste orde autocorrelatie. Er geldt dat  $\sqrt{LM1}$  een t-verdeling heeft met  $T-k-1=26$  vrijheidsgraden. Hierin is  $k$  het aantal verklarende variabelen inclusief de constante. Zie Kiviet (1986), p. 245.

We zien dat hier de mutatie van de werkloosheid een bijdrage levert die zeer significant van nul verschilt. De elasticiteit van  $\ln y$  en  $\ln h$  met betrekking tot  $\ln a$  zijn hier beide veel kleiner dan één. De coëfficiënt van de fouten correctie-term is zeer significant van nul verschillend.

Voor vergelijking (22) zijn de karakteristieken van de systeemdynamiek in tabel 4 samengevat. De manier waarop de elasticiteiten etc. zijn berekend is in bijlage B weergegeven.

Tabel 4 Korte en lange termijn elasticiteiten van de werkgelegenheidsvergelijking voor de bouwnijverheid

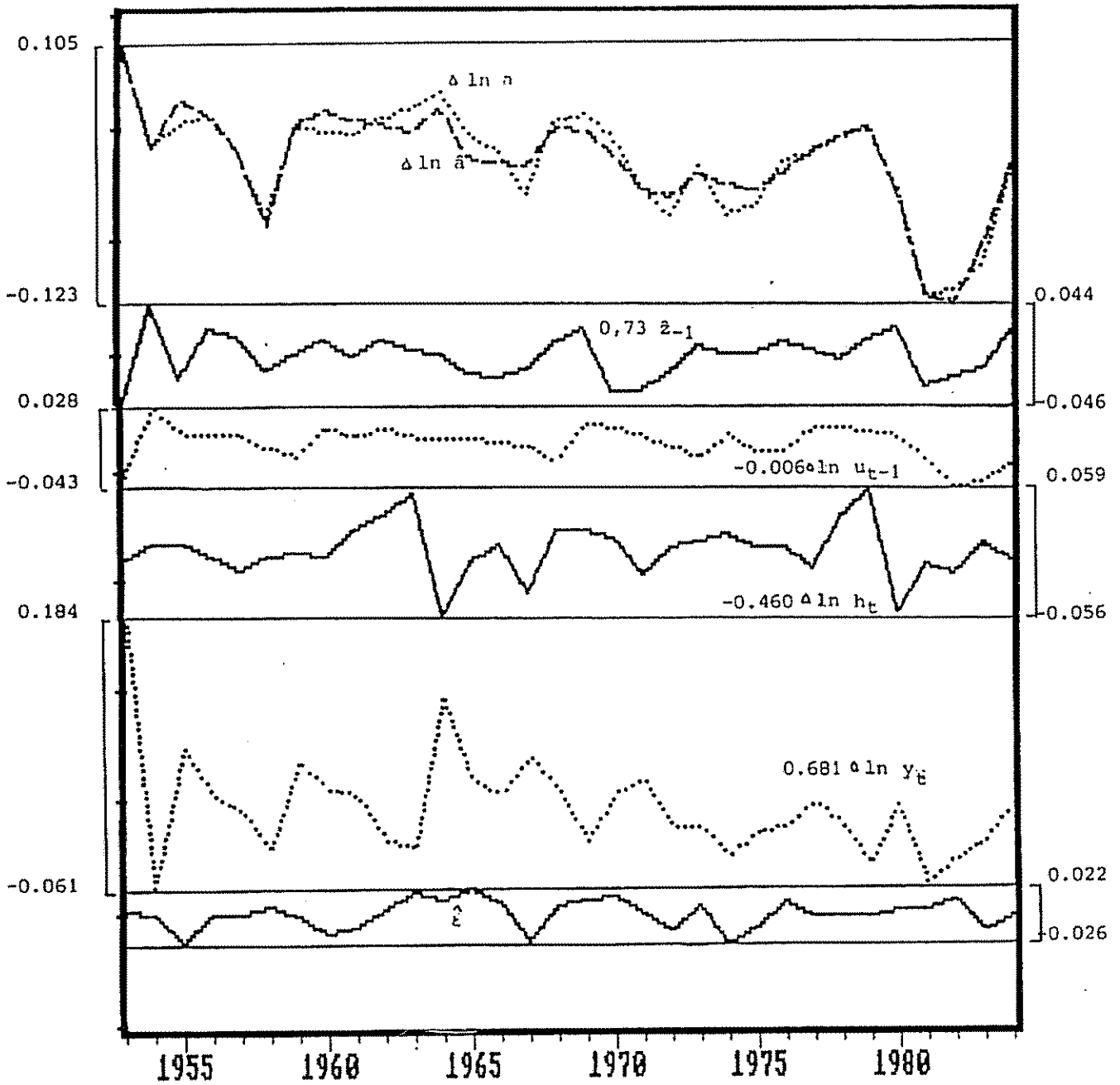
	Periode (jaren)				
	0	1	2	3	$\infty$
Productie (y)	0,68	0,86	0,91	0,93	0,95
Productieve uren (h)	-0,46	-0,86	-0,97	-1,00	-1,01
Reële uurloon ( $l/ph$ )	0	-0,20	-0,25	-0,27	-0,28
Werkloosheid <sup>a)</sup>	0	-0,55	-0,15	-0,04	0

a) semi-elasticiteit

Voor de werkloosheid worden de semi-elasticiteiten gegeven; deze geven weer met hoeveel procent de werkgelegenheid verandert als het werkloosheidspercentage met 1% toeneemt. De trendmatige arbeidsproduktiviteitsstijging is 1,2% per jaar. Het idee dat werkloosheid blijvend produktiviteitsverhogend werkt, vindt dus geen ondersteuning hier. De tijdelijke effecten van de werkloosheid kunnen voortvloeien uit het wel waargenomen verschijnsel dat men bij een ruime arbeidsmarkt tot ontslag overgaat, waarna tegen een lager loon weer mensen worden aangetrokken. Hierbij zullen vertragingen een rol spelen.

Tot slot is in figuur 1 het stapeldiagram behorend bij vergelijking (22) gepresenteerd.

Figuur 1 Stapeldiagram werkgelegenheidsvergelijking bouwnijverheid.





Literatuurverwijzingen

- Blangiewicz, M. en W. Charemza: "Cointegration in small samples: empirical percentiles, drifting moments and customized testing", University of Leicester, (1989).
- Boswijk, H.P.: "Estimation and testing for cointegration with trended variables: a comparison of static and dynamic regression procedures", Universiteit van Amsterdam (1989).
- Box, G.E.P. en G.M. Jenkins: Time Series Analysis, forecasting and control, San Francisco, 1970.
- Davidson, J.E.H, D.F. Hendry, F. Srba, en S. Yeo,: "Econometric Modelling of the Aggregate Time-Series Relationship Between Consumers' Expenditure and Income in the UK", Economic Journal, 88 (1978), pp. 661-692.
- Dickey, D.A. en W.A. Fuller: "Distribution of the Estimators for Autoregressive Time Series with a Unit Root", Journal of the American Statistical Association, 74 (1979), pp. 427-431.
- Dickey, D.A. en W.A. Fuller: "Likelihood Ratio Statistics for Autoregressive Time Series with a Unit Root", Econometrica, 49 (1981), pp. 1057-1072.
- Don, F.J.H., Linear Methods in Non-Linear Models, Den Haag, 1986.
- Dhrymes, P.J., Distributed Lags: Problems of Formulation and Estimation, San Francisco, 1970.
- Engle, R.F. en C.W.J. Granger: "Co-Integration and Error-Correction: Representation, Estimation and Testing", Econometrica, 55 (1987), pp. 251-276.
- Engle, R.F. en B.S. Yoo: "Forecasting and Testing in Co-integrated Systems", Journal of Econometrics, 35 (1987) pp. 143-159.
- Granger, C.W.J.: "Developments in the Study of Cointegrated Economic Variables", Oxford Bulletin of Economics and Statistics, 48 (1986), pp. 213-228.
- Granger, C.W.J. en M.J. Morris, "Time Series Modelling and Interpretation", Journal of the Royal Statistical Society, series A, 139, pp. 246-257
- Granger, C.W.J. en P. Newbold: "Spurious Regressions in Econometrics", Journal of Econometrics, 2 (1974), pp. 111-120.
- Hall, S.G.: "An Application of the Granger and Engle Two-Step Estimation Procedure to United Kingdom Aggregate Wage Data", Oxford Bulletin of Economics and Statistics, 48 (1986), pp. 229-239.
- Harvey, A.C.: The Econometric Analysis of Time Series, Oxford, 1982.

- Hendry, D.F.: "Econometric Modelling with Cointegrated Variables: an Overview", Oxford Bulletin of Economics and Statistics, 48 (1986), pp. 201-212.
- Hendry, D.F., A.R. Pagan en J.D. Sargan, "Dynamic Specification", in: Griliches, Z. en M.D. Intriligator, Handbook of econometrics, Amsterdam, 1984.
- Intriligator, M.D.: Econometric Models, Techniques and Applications, Amsterdam, 1978.
- Jenkinson, T.J.: "Testing Neo-Classical Theories of Labour Demand: an Application of Cointegration Techniques", Oxford Bulletin of Economics and Statistics, 48 (1986), pp. 241-251.
- Kiviet, J.F.: "On the Rigour of some Misspecification Tests for Modelling Dynamic Relationships", Review of Economic Studies, 53 (1986): pp. 245.
- Kloek, T., "Dynamic Adjustment when the Target is Nonstationary", International Economic Review, 25 (1984), pp. 315-326.
- Phillips, P.C.B.: "Understanding Spurious Regressions in Econometrics", Journal of Econometrics, 33 (1986), pp. 311-340.
- Sargan, J.D. en A. Bhargava: "Testing Residuals from Least Squares Regression for being generated by the Gaussian Random Walk", Econometrica, 51 (1983), pp. 153-174.
- Stock, J.H.: "Asymptotic Properties of Least Squares Estimators of Cointegrating Vectors", Econometrica, 55 (1987), pp. 1035-1056
- Wonnacott, R.J. en T.H. Wonnacott: Econometrics: Londen, 1970.

Bijlage A

Afleidingen behorende bij paragraaf 1

In deze bijlage worden de vergelijkingen (6), (7), (11) en (12) afgeleid voor het meer algemene geval dat  $\Delta x_t$  een verwachtingswaarde  $\mu$  heeft die niet nul is. Ten eerste leiden we af wat het lange termijnverband is tussen  $x_t$  en  $y_t$  uit vergelijking (4), die we hier herhalen:

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + \epsilon_t, \quad (\text{A.8})$$

met  $-1 < \alpha_3 < 1$ .

We veronderstellen dat  $\Delta x_t$  stationair is met een verwachtingswaarde  $\mu$  voor alle  $t$ . Wanneer  $\mu$  gelijk aan nul is, bevat  $x_t$  geen "drift" en wordt het lange termijnverband gedefinieerd door in de vergelijking (A.8) de storing  $\epsilon_t$  weg te laten en voor  $y_t$  en  $x_t$  respectievelijk de constanten  $y$  en  $x$  in te vullen, hetgeen, uitgaande van (5), direct de vergelijkingen (6) en (7) oplevert.

Als echter  $\mu$  ongelijk is aan nul, definiëren we het lange termijnverband als de relatie

$$y_t = ax_t + b \quad (\text{A.9})$$

die ontstaat wanneer we in (A.8) de storing weglaten en  $\Delta x_t$  constant veronderstellen (en dus gelijk aan  $\mu$ ). Dit laatste impliceert

$$x_{t-1} = x_t - \mu. \quad (\text{A.10})$$

We vullen dit in het rechterlid van (A.8) in. Tevens vullen we in:

$$\begin{aligned} y_{t-1} &= y_t + a(x_{t-1} - x_t) = \\ &= y_t - a\mu. \end{aligned} \quad (\text{A.11})$$

Dit levert op:

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_2 (x_t - \mu) + \alpha_3 (y_t - a\mu). \quad (\text{A.12})$$

Dus:

$$(1 - \alpha_3)y_t = \alpha_0 + (\alpha_1 + \alpha_2)x_t - \mu(\alpha_2 + \alpha_3 a). \quad (\text{A.13})$$

We delen (A.13) links en rechts door  $1-\alpha_3$  en vergelijken dan het rechterlid met het rechterlid van (A.9). Dit resulteert in:

$$a = (\alpha_1 + \alpha_2) / (1 - \alpha_3) \quad (\text{A.14})$$

en

$$b = (\alpha_0 - \mu(\alpha_2 + \alpha_3 a)) / (1 - \alpha_3) = \alpha_0 / (1 - \alpha_3) - \mu(\alpha_2 + \alpha_1 \alpha_3) / (1 - \alpha_3)^2 \quad (\text{A.15})$$

We zien dat de waarde van  $a$  niet verandert door  $\mu$ , maar de waarde van  $b$  wel. Bijvoorbeeld, bij een positieve  $\mu$  is de waarde van  $b$  kleiner. Intuïtief: doordat de aanpassing van  $y_t$  aan  $x_t$  niet onmiddellijk is, "blijft"  $y_t$  "achter" bij  $x_t$ . De aanpassing is onmiddellijk als  $\alpha_2 = \alpha_3 = 0$  (zie vergelijking (A.8)); de  $\mu$ -term in (A.15) is dan nul. Vergelijk het voorttrekken van een steen over de grond aan een elastiek (niet-onmiddellijke aanpassing), met het voorttrekken aan een touw (onmiddellijke aanpassing), waarbij in beide gevallen met een snelheid  $\mu$  wordt getrokken. De snelheid van de steen aan het elastiek is na enige tijd hetzelfde als de steen aan het touw (beide gelijk aan  $\mu$ ), maar de steen aan het elastiek blijft permanent iets achter bij de steen aan het touw. Deze achterblijving is groter naarmate de snelheid  $\mu$  groter is en naarmate het elastiek elastischer is.

Ten tweede leiden we vergelijking (11) af voor het geval  $\mu \neq 0$ . We vullen daartoe in de definitie

$$z_t \equiv y_t - ax_t - b \quad (\text{A.16})$$

vergelijking (A.8) in:

$$\begin{aligned} z_t &= \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + \epsilon_t - ax_t - b = \\ &= \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 (z_{t-1} + ax_{t-1} + b) + \epsilon_t - ax_t - b = \\ &= \alpha_3 z_{t-1} + (\alpha_1 - a)x_t + (\alpha_3 a + \alpha_2)x_{t-1} + \alpha_0 - b(1 - \alpha_3) + \epsilon_t . \end{aligned} \quad (\text{A.17})$$

Hierin vullen we (A.14) in, hetgeen oplevert:

$$z_t = \alpha_3 z_{t-1} - \Delta x_t (\alpha_2 + \alpha_1 \alpha_3) / (1 - \alpha_3) + \alpha_0 - b(1 - \alpha_3) + \epsilon_t \quad (\text{A.18})$$

Met vergelijking (A.15) resulteert:

$$z_t = \alpha_3 z_{t-1} - (\Delta x_t - \mu)(\alpha_2 + \alpha_1 \alpha_3) / (1 - \alpha_3) + \epsilon_t \quad (\text{A.19})$$

De Dickey-Fuller test (DF) is gebaseerd op de volgende herschrijving van (A.19):

$$\begin{aligned}\Delta z_t &= -(1-\alpha_3)z_{t-1} - (\Delta x_t - \mu)(\alpha_2 + \alpha_1\alpha_3)/(1-\alpha_3) + \epsilon_t \\ &= -(1-\alpha_3)z_{t-1} + \text{stationaire termen met verwachting nul.}\end{aligned}\quad (\text{A.20})$$

Merk op dat als in het rechterlid van (A.8) hogere vertragingen van  $y_t$  zouden voorkomen ( $y_{t-2}$ ,  $y_{t-3}$ , enz.) dat dan in het rechterlid van (A.20) diezelfde vertragingen in  $z_t$  voorkomen, via de toepassing van (A.16). Dit is te schrijven als extra termen in  $\Delta z_{t-1}$ ,  $\Delta z_{t-2}$ , enz. We hebben dan de "augmented" Dickey-Fuller test (ADF). Zie voor deze testen elders in deze notitie.

Ten slotte leiden we vergelijking (12) af voor het geval  $\mu \neq 0$ . Er geldt, met (A.8) en (A.16):

$$\begin{aligned}\Delta y_t &= y_t - y_{t-1} = \\ &= \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} + \alpha_3 y_{t-1} + \epsilon_t - y_{t-1} = \\ &= \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} - (1-\alpha_3)y_{t-1} + \epsilon_t = \\ &= \alpha_0 + \alpha_1 x_t + \alpha_2 x_{t-1} - (1-\alpha_3)(z_{t-1} + ax_{t-1} + b) + \epsilon_t = \\ &= \alpha_1 x_t + (\alpha_2 - (1-\alpha_3)a)x_{t-1} - (1-\alpha_3)z_{t-1} + \alpha_0 - b(1-\alpha_3) + \epsilon_t\end{aligned}\quad (\text{A.21})$$

Met vergelijking (A.14) en (A.15) volgt:

$$\begin{aligned}\Delta y_t &= \alpha_1 x_t + (\alpha_2 - \alpha_1 - \alpha_2)x_{t-1} - (1 - \alpha_3)z_{t-1} + \alpha_0 - \{\alpha_0 - \mu(\alpha_2 + a\alpha_3)\} + \epsilon_t = \\ &= \alpha_1 \Delta x_t - (1 - \alpha_3)z_{t-1} + \mu(\alpha_2 + a\alpha_3) + \epsilon_t = \\ &= \alpha_1 (\Delta x_t - \mu) - (1 - \alpha_3)z_{t-1} + \mu(\alpha_1 + \alpha_2 + a\alpha_3) + \epsilon_t = \\ &= \alpha_1 (\Delta x_t - \mu) - (1 - \alpha_3)z_{t-1} + \mu(a(1 - \alpha_3) + a\alpha_3) + \epsilon_t = \\ &= \alpha_1 (\Delta x_t - \mu) - (1-\alpha_3)z_{t-1} + a\mu + \epsilon_t\end{aligned}\quad (\text{A.22})$$

Bijlage B

Het "autoregressive distributed lag" model (AD-model)<sup>1</sup>;  
enige kanttekeningen

1. Vertragsoperator

Vertragingen in een model kunnen worden beschreven met de vertragsoperator L. We definiëren:

$$Lw_t = w_{t-1} \quad ; \quad L^i w_t = w_{t-i} .$$

Bovendien voeren we de vertragspolynoom van de orde  $m > 0$  in:

$$\mu(L) = \mu_0 + \mu_1 L + \dots + \mu_m L^m = \sum_{i=0}^m \mu_i L^i$$

Bij veel bewerkingen kan de vertragsoperator behandeld worden als een gewone getalsmatige variabele<sup>2</sup>. Vandaar dat het prettig is om de dynamiek hiermee tot uitdrukking te brengen.

2. Het AD-model

Men spreekt van een  $AD(k_0, k_1, \dots, k_n)$  model als:

$$y_t = \text{constante} + \text{dummies} + \sum_{i=1}^{k_0} \gamma_i y_{t-i} + \sum_{j=1}^n \sum_{i=0}^{k_j} \beta_i^{(j)} w_{t-i,j} + \epsilon_t$$

Meestal beperkt men zich tot het  $AD(1,1, \dots, 1)$  model. Dit is al zeer algemeen. Dit is in tabel 1 geïllustreerd<sup>3</sup> voor het geval van één verklarende variabele, dus voor het  $AD(1,1)$  model, dat we als volgt schrijven:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 x_{t-1} + \epsilon_t \quad (\text{B.1})$$

<sup>1</sup>Deze bijlage is onder andere gebaseerd op het collegedictaat voor de PAO-cursus uit 1982, Regressie van Tijdsreeksen, van J.F. Kiviet en G. Ridder.

<sup>2</sup>Zie: P.J. Dhrymes (1970).

<sup>3</sup>Zie: David F. Hendry, Adrian R. Pagan and J. Denis Sargan (1985).

Tabel 1 De bijzondere gevallen van het AD(1, 1) model

Modeltype	klasse	vergelijking	restricties <sup>a</sup>
<u>A</u> statisch	AD(0,0)	$y_t = \alpha_2 x_t + \epsilon_t$	$\alpha_3 = \alpha_1 = 0$
<u>B</u> univariaat tijdsreeksmodel	AD(1)	$y_t = \alpha_1 y_{t-1} + \epsilon_t$	$\alpha_2 = \alpha_3 = 0$
<u>C</u> eerste verschillen	AD(1,1)	$\Delta y_t = \alpha_2 \Delta x_t + \epsilon_t$	$\alpha_1 = 1; \alpha_3 = -\alpha_2$
<u>D</u> leading indicator	AD(0,1)	$y_t = \alpha_3 x_{t-1} + \epsilon_t$	$\alpha_2 = \alpha_1 = 0$
<u>E</u> partial adjustment	AD(1,0)	$y_t = \alpha_1 y_{t-1} + \alpha_2 x_t + \epsilon_t$	$\alpha_3 = 0$
<u>F</u> distributed lag	AD(0,1)	$y_t = \alpha_2 x_t + \alpha_3 x_{t-1} + \epsilon_t$	$\alpha_1 = 0$
<u>G</u> Common factor	AD(1,1)	$y_t = \alpha_2 x_t + u_t$ en $u_t = \alpha_1 u_{t-1} + \epsilon_t$	$\alpha_3 = -\alpha_2 \alpha_1$
<u>H</u> Error Correctie	AD(1,1)	$\Delta y_t = \alpha_2 \Delta x_t +$ $(1-\alpha_1)(x-y)_{t-1} + \epsilon_t$	$\sum \alpha_i = 1$
<u>I</u> Dead Start	AD(1,1)	$y_t = \alpha_1 y_{t-1} + \alpha_3 x_{t-1} + \epsilon_t$	$\alpha_2 = 0$

a) Restricties t.o.v.  $y_t = \alpha_1 y_{t-1} + \alpha_2 x_t + \alpha_3 x_{t-1} + \epsilon_t$ .

Door van een algemene specificatie als (B.1) uit te gaan kan door middel van toetsen nagegaan worden welke dynamische vorm het meest geëigend is. Indien het AD(1,1) model misspecificatietoetsen niet doorstaat, kan alsnog naar meer algemenere vormen worden uitgekeken<sup>4</sup>. De manier waarop het beste getoetst kan worden staat bij Hendry, Pagan en Sargan beschreven.

### 3. Karakteristieken van de systeemdynamiek<sup>5</sup>

We gaan weer uit van één verklarende variabele en abstraheren van storingen:

$$Y(L) y_t = \beta(L) w_t \quad \text{en} \quad \mu(L) = Y(L)/\beta(L) \quad (\text{B.2})$$

We nemen aan dat  $Y(L)$  zodanig is, dat (B.2) stabiel is. De wortels van dit verdragingspolynoom moeten dan buiten de eenheidscirkel liggen. Er is dan een evenwichtsooplossing mogelijk. Dit is een voorbeeld van het rekenen met  $L$  en  $Y(L)$  alsof het getallen zijn: we eisen dat de oplossing van  $Y(L) = 0$  voldoet aan  $|L| > 1$ . Bijvoorbeeld bij vergelijking (B.1) is  $Y(L) = 1 - \alpha_1 L$  en dus is  $L = 1/\alpha_1$  hier de gezochte oplossing. Dan impliceert  $|L| > 1$  dat  $|\alpha_1| < 1$ . Dit

<sup>4</sup>Zie bijv.: T. Kloek, (1984) hij geeft een argumentatie voor het gebruik van AD(2,2) modellen.

<sup>5</sup>Zie: A.C. Harvey (1981), pp. 229-235.

laatste herkennen we onmiddellijk als een voorwaarde voor stabiliteit. In het algemene geval bestaan de oplossingen van  $\Upsilon(L) = 0$  uit complexe getallen, en moeten we, zoals gezegd, dus  $|L| > 1$  opvatten als buiten de eenheidscirkel in het complexe vlak.

Indien  $w_t$  constant is, met  $w_t = \bar{w}$ , dan zal dus gelden  $y_t = \bar{y}$  zodat

$$\bar{y} = \frac{\beta(L)}{\Upsilon(L)} w_t = \frac{\beta(1)}{\Upsilon(1)} \bar{w} = \mu(1) \bar{w}$$

waarbij:  $\beta(1) = \sum \beta_j$ ;  $\Upsilon(1) = \sum \Upsilon_j$  en  $\mu(1) = \sum \mu_j$

We noemen nu:  $TM = \frac{\beta(1)}{\Upsilon(1)} = \mu(1) = \sum_{j=0}^{\infty} \mu_j$  de total multiplier. Hij geeft de verandering in  $y$  weer als  $w$  met 1 toeneemt. Als  $y$  en  $w$  in logaritmes gemeten zijn, dan is  $TM$  de lange termijnelasticiteit. Vergelijk ook met (6) in § 1.4.

We noemen:  $IM = \sum_{j=0}^J \mu_j$  de  $J$ -de interim-multiplier. Hij geeft de verandering in  $y$  weer na  $J$  perioden, indien  $w$  1 toeneemt. Indien  $y$  en  $w$  in logaritmen zijn gemeten, dan is  $IM$  de  $J$ -de korte termijnelasticiteit.

We noemen:  $\bar{\mu} = \sum j \mu_j / \sum \mu_j$  de gemiddelde vertraging. Deze is gedefinieerd als  $\mu_j \geq 0$  voor  $j = 0, \dots, \infty$ . Er geldt:

$$\bar{\mu} = \frac{\beta'(1)}{\beta(1)} - \frac{\Upsilon'(1)}{\Upsilon(1)}$$

waarbij:  $\beta'(L) = \frac{\partial}{\partial L} \beta(L)$   $\Upsilon'(L) = \frac{\partial}{\partial L} \Upsilon(L)$ .

Dit is eenvoudig af te leiden, als volgt. Bedenk dat  $\bar{\mu} = \mu'(1)/\mu(1)$ , omdat  $\mu(1) = \sum \mu_j$ . Voorts: met  $\mu(L) = \sum \mu_j L^j$  volgt  $\mu'(L) = \sum \mu_j j L^{j-1}$ , zodat  $\mu'(1) = \sum j \mu_j$ . Met  $\mu = \beta/\Upsilon$  volgt

$$\mu'(L) = \frac{\beta'(L)}{\Upsilon(L)} - \frac{\beta(L)\Upsilon'(L)}{\Upsilon(L)^2}$$

Wanneer we hierin  $L = 1$  invullen, en delen door  $\mu(1) = \beta(1)/\Upsilon(1)$ , volgt de gegeven formule.



#### 4. Voorbeeld

Het AD(1,1) model heeft de volgende algemene vorm:

$$y = \alpha_1 y_{-1} + \alpha_2 x + \alpha_3 x_{-1} \quad (\text{B.3})$$

Indien op deze wijze zou worden geschat, kunnen multicollineariteitsproblemen ontstaan. Vandaar dat de vergelijking vaak wordt geschreven als:

$$\Delta y = \alpha_2 \Delta x + (\alpha_1 - 1)y_{-1} + (\alpha_2 + \alpha_3)x_{-1}$$

Indien de variabelen in logaritmes zijn gemeten en men wil toetsen of de lange termijn elasticiteit van  $x$  gelijk aan 1 is, dan kan de relatie nog verder worden herschreven:

$$\Delta y = \alpha_2 \Delta x + (1 - \alpha_1)(x_{-1} - y_{-1}) + (\alpha_2 + \alpha_3 + \alpha_1 - 1)x_{-1} \quad (\text{B.4})$$

Gebruikmakend van de lag operator, krijgt men:

$$(1 - L)y = \alpha_2(1 - L)x + (1 - \alpha_1)L(x - y) + (\alpha_2 + \alpha_3 + \alpha_1 - 1)Lx \quad (\text{B.5})$$

Uitgaande van deze laatste vorm, die vaak wordt gehanteerd, zullen de karakteristieken van de systeemdynamiek worden bepaald.

##### 1) De total multiplier

De total multiplier volgt uit (B.3):

$$TM = \frac{\alpha_2 + \alpha_3}{1 - \alpha_1}$$

Zie ook vergelijking (6). In feite bepalen de niveauvariabelen de total multiplier.

##### 2) De interim multipliers

De interim multipliers zijn:  $IM = \sum_{j=0}^J \mu_j \quad (J = 0, 1, 2, \dots)$ .

Een eenvoudige manier om de  $\mu_j$  coëfficiënten te berekenen is de volgende:

Uit  $y = \frac{\alpha_2 + \alpha_3 L}{1 - \alpha_1 L} x = \sum_{j=0}^{\infty} \mu_j L^j x$  volgt:

$$(\alpha_2 + \alpha_3 L) = (1 - \alpha_1 L) \left( \sum_{j=0}^{\infty} \mu_j L^j x \right). \text{ of}$$

$$(\mu_0 + \mu_1 L + \mu_2 L^2 + \mu_3 L^3 + \dots + \\ - \alpha_1 \mu_0 L - \alpha_1 \mu_1 L^2 - \alpha_1 \mu_2 L^3 - \dots) = (\alpha_2 + \alpha_3 L)$$

Hieruit volgt:

$$\mu_0 = \alpha_2$$

$$\mu_1 - \alpha_1 \mu_0 = \alpha_3 \quad \rightarrow \quad \mu_1 = \alpha_3 + \alpha_1 \alpha_2$$

$$\mu_2 - \alpha_1 \mu_1 = 0 \quad \rightarrow \quad \mu_2 = \alpha_1 \mu_1$$

etc.

De gewichten vanaf  $\mu_1$  volgen een meetkundige reeks ( $\mu_j = \alpha_1 \mu_{j-1}$ ) en de extra effecten worden dus steeds kleiner als  $|\alpha_1| < 1$ .

### 3) De gemiddelde vertraging $\bar{L}$

$$\bar{\mu} = \frac{\beta'(1)}{\beta(1)} - \frac{\gamma'(1)}{\gamma(1)} = \frac{\alpha_3}{\alpha_1 + \alpha_2} + \frac{\alpha_1}{1 - \alpha_1}$$

We zien dat de vertraging onbeperkt lang kan worden als  $\alpha_1$  naar één gaat.

#### Opmerking

Indien men wil toetsen of de lange termijn elasticiteit (total multiplier) van  $x$  gelijk aan 1 is (ingeval in logaritmes wordt gemeten), dan moet men de volgende restrictie op de coëfficiënten toetsen:  $(\alpha_2 + \alpha_3) / (1 - \alpha_1) = 1$ . Indien de vergelijking in de vorm van (B.4) is geschreven, komt dit neer op het toetsen van de hypothese of de coëfficiënt van variabele  $x_{-1}$  significant van 0 afwijkt.

Bijlage C

Een meer formele beschrijving van integratie, coïntegratie  
en de Engle-Granger schattingsprocedure

1. Integratie, coïntegratie, foutencorrectiemodellen

De in de hoofdttekst geïntroduceerde begrippen zullen meer formeel worden beschreven.

1. Een vector  $X_t$  met  $N$  variabelen is geïntegreerd van de orde  $d$  (notatie  $X_t \sim I(d)$ ) als alle reeksen van de variabelen pas stationair zijn na  $d$  keer differentiëren (en dan kunnen worden weergegeven door een niet deterministisch ARMA-model).
2. De variabelen in de vector  $X_t$  noemt men gecoïntegreerd van de orde  $d, b$  (notatie  $X_t \sim CI(d, b)$ ) als:
  - de vector  $X_t$  met variabelen geïntegreerd is van de orde  $d$ , dus  $X_t \sim I(d)$
  - en er een vector  $\alpha$  ongelijk  $0$  bestaat, zodanig dat  $Z_t \equiv \alpha' X_t \sim I(d-b)$  met  $b > 0$ , waarin  $\alpha$  de coïntegratie vector wordt genoemd.

Het representatie-theorema van Granger (Engle en Granger, 1987, blz. 255).

Als  $X_t$ , een  $N \times 1$  vector van variabelen, gecooïntegreerd is van de orde  $d = 1$ ,  $b = 1$  en met coïntegratie-rang  $r$  (het aantal lineair onafhankelijke coïntegratie-vectoren dat kan worden gevonden), dan bestaat er een foutencorrectie-representatie  $A(L)(1-L)X_t = -\beta Z_{t-1} + f(L)\epsilon$  met  $Z_t = \alpha' X_t$  een  $r \times 1$  vector van stationaire random variabelen;  $A(0) = I_N$ ;  $\epsilon$  storingsterm; waarin de matrix  $A(L)$  een polynoom is in de lag operator en  $f(L)$  een scalair polynoom in de lag operator, zodat:

$$A(L) = I_N + A(1).L + A(2).L^2 + \dots + A(p)L^p;$$

$f(L) = f_0 + f_1.L + f_2.L^2 + \dots + f_q.L^q$  met  $A(j)$ ,  $N \times N$  matrices met coëfficiënten en  $f_j \in R$ ; en  $\alpha$  en  $\beta$ ,  $N \times r$  matrices van coëfficiënten.

## 2. De schattingsprocedure

Bij het schatten van de  $r$  niveaurelaties die beschreven worden door  $\alpha'X_t = Z_t$  zoeken we naar de kolommen  $\alpha_i$  in de matrix  $\alpha$  ( $i = 1..r$ ) die kleinste  $\sum z_{it}^2$  opleveren: OLS. Dit moet natuurlijk onderworpen worden aan een normalisatie-restrictie, omdat anders de triviale oplossing  $\alpha_i = 0$  resulteert. Zo'n normalisatie-restrictie geeft de elementen van  $\alpha_i$  als functie van een kleiner aantal onbekende coëfficiënten in een vector  $\theta_i$ :

$$\alpha_i = q_i + Q_i\theta_i \quad (C.1)$$

Hierin zijn  $q_i$  en  $Q_i$  respectievelijk een bekende vector en matrix. De  $i^e$  relatie  $\alpha_i'X_t = z_{it}$  kunnen we hiermee schrijven als

$$(q_i + Q_i\theta_i)'X_t = z_{it} \quad (C.2)$$

Dit is te herleiden tot

$$q_i'X_t = -\theta_i'Q_i'X_t + z_{it} \quad (C.3)$$

We hebben nu een regressievergelijking met  $q_i'X_t$  als te verklaren variabele en  $Q_i'X_t$  als verklarende variabele (en  $-\theta_i$  als coëfficiënten-vector). In de praktijk bestaan  $q_i$  en  $Q_i$  geheel uit nullen en énen, zodat  $q_i'X_t$  één element uit  $X_t$  is (de gekozen te verklaren variabele), terwijl  $Q_i'X_t$  bestaat uit de overige elementen van  $X_t$ . Zodoende is (C.3) direct met standaard regressie-software te schatten.

De formule die Engle en Granger (1987, blz. 261) geven voor de geschatte  $\theta_i$  ziet er op het eerste gezicht enigszins ingewikkeld uit:

$$\theta = -[Q'(I_r \otimes X'X)Q]^{-1} Q'(I_r \otimes X'X)q \quad (C.4)$$

Hierin is  $X$  de volledige datamatrix (met rijen  $X_t$ ) en  $I_r$  is de eenheidsmatrix van orde  $r$ . De matrix  $Q$  is blokdiagonaal, met  $Q_1, Q_2, \dots, Q_r$  als de blokken en de vector  $q$  bevat de "gestapelde"  $q_i$ 's. Vanwege het blokdiagonale karakter van  $Q$  en van  $I_r \otimes X'X$  is de matrix tussen vierkante haken in (C.4) ook blokdiagonaal, met de matrices  $Q_i'X'XQ_i$  op de diagonaal. Wanneer we ook rekening houden met het blokdiagonale karakter van de matrix  $Q'(I_r \otimes X'X)$  achter de geïnventeerde matrix in (C.4), volgt dat (C.4) neerkomt op

$$\theta_i = -(Q_i'X'XQ_i)^{-1} Q_i'X'Xq_i = -[(XQ_i)'(XQ_i)]^{-1} (XQ_i')Xq_i \quad (C.5)$$

voor  $i = 1 \dots r$ . Het is onmiddellijk in te zien dat deze laatste vorm resulteert uit regressie van  $Xq_i$  op  $XQ_i$ , overeenkomstig (C.3).

Deze schatter zal een zeer goede benadering van  $\alpha$  geven, aangezien de varianties van de residuen van de  $r$ -coïntegratie-vergelijkingen werden geminimeerd, terwijl alle lineaire combinaties van de variabelen uit  $X_t$  per definitie asymptotisch een oneindige steekproefvariantie hebben behalve de  $r$ -coïntegratie-vergelijkingen. Voor dit soort schatters geldt het super-consistentie theorema van Stock (zie Engle en Granger, 1987, blz. 262, en Stock, 1987, blz. 1042), dat stelt: indien  $X_t \sim CI(1,1)$  en  $\alpha$  wordt geschat volgens (C.5), dus met OLS, dan is de convergentie-snelheid van de orde  $T$ . Verder blijken de OLS schatters consistent te zijn ook al is er sprake van simultaneïteit. (Indien  $X_t \sim CI(0,0)$  dan is de convergentiesnelheid van de orde  $\sqrt{T}$  en zijn bij simultaneïteit de schatters niet consistent (Intriligator, 1978, blz. 375)).

De normalisatievector  $q_i$  kan willekeurig zijn (in zoverre de variabele voorkomt in de betreffende vergelijking). OLS zal evenwel niet exact dezelfde resultaten geven, maar naarmate de  $R^2$  hoger is, zullen de resultaten meer naar elkaar toe convergeren.

Engle en Granger (1987, blz. 262) bewijzen voor het met OLS schatten van de relaties uit het foutencorrectiemodel (zie representatie theorema, blz. 33) nadat  $Z_t$  vervangen is door  $\alpha'X_t$ , dat de schatters asymptotisch dezelfde verdeling zullen hebben als de maximum likelihood schatter die de juiste waarde van  $\alpha$  zou gebruiken. Kleinste kwadraten standaard fouten zijn consistente schatters van de echte standaard fouten. Dit bewijs is in de betreffende publicatie uitvoerig beschreven.

Bijlage D

Toetsen op integratie en cointegratie

1. Bepaling orde van integratie

In de literatuur worden diverse toetsen gebruikt om de orde van integratie te bepalen. Deze toetsen hebben dezelfde structuur. Hun nulhypothese is: de betrokken reeks, zeg  $x_t$ , is niet stationair. Als deze nulhypothese niet verworpen kan worden, dan wordt zij vervangen door:  $\Delta x_t$  is niet stationair; vervolgens  $\Delta^2 x_t$ , enz., tot een wel stationaire reeks gevonden wordt. Dan is per definitie de orde van integratie bepaald.

We bespreken in deze bijlage de drie formele toetsen die in paragraaf 1.7 reeds kort aan de orde kwamen. Kritieke waarden voor deze drie toetsen staan in de tabel 1 van deze bijlage, onder "aantal variabelen" = 1. Voor alle drie geldt dat gekeken wordt of de toetsgrootte significant van nul verschilt. Dus in alle gevallen: niet-stationariteit verwerpen als de toetsgrootte groter is dan de tabelwaarde.

Tevens wordt een vierde, meer informele toets besproken, die ook in paragraaf 2 gebruikt is: de analyse van de autocorrelaties van de betrokken reeks.

a. DW-toets (Sargan, Bhargava, 1983)

We gaan hier uit van een vergelijking als (1):

$$(1 - \rho L)x_t = c + \epsilon_t \quad (D.1)$$

De nulhypothese is:  $\rho = 1$  met  $c = 0$ ; oftewel:  $\Delta x_t = \epsilon_t$ . De alternatieve hypothese is:  $-1 < \rho < 1$ , met willekeurige  $c$ . Er geldt dan:

$$x_t = (c + \epsilon_t)/(1 - \rho L) = c/(1 - \rho) + \mu_t \quad (D.2)$$

met

$$\mu_t = \rho \mu_{t-1} + \epsilon_t \quad (D.3)$$

We toetsen met de bekende DW-toetsgrootte, toegepast op de regressie van  $x_t$  op een constante, overeenkomstig (D.2). Zoals bekend is de DW een schatter van  $2(1 - \rho)$ . Zij is door Durbin en Watson uitgevonden om de nulhypothese  $\rho = 0$  (geen residuele autocorrelatie) te toetsen en er wordt in dat geval dan ook gekeken of zij significant afwijkt van de waarde  $2(1 - 0) = 2$ . Hier toetsen we  $\rho = 1$ , en we moeten hier dus nagaan of zij significant afwijkt van de waarde  $2(1 - 1) = 0$ . Zie verder tabel 1.

Merk op dat onder onze nulhypothese ( $\rho = 1$ ) het laatste lid van (D.2) niet bestaat. Om in te zien dat de DW-grootte onder deze nulhypothese asymptotisch naar nul gaat, geven we hier nog eens het betoog weer uit paragraaf 1.7: onder de nulhypothese is  $x_t$  geïntegreerd en zal het gekwadraterde regressie-residu (in de noemer van de DW) bij toenemende reekslengte gemiddeld onbeperkt groot worden, terwijl de mutatie van het regressie-residu (in de teller) gelijk is aan  $\Delta x_t$ , die weer gelijk is aan  $\epsilon_t$ : de "witte ruis" met vaste constante variantie.

Deze toets heeft van de drie hier besproken toetsen de meeste "power". Dat wil zeggen dat de kans dat de nulhypothese ten onrechte niet wordt verworpen, bij de DW-toets het kleinst is. (Bij gelijke kans op ten onrechte verwerpen, uiteraard.)

b. DF-toets (Dickey en Fuller, 1979)

We gaan weer uit van (D.1), die we herschrijven als:

$$\Delta x_t = \beta x_{t-1} + c + \epsilon_t \quad (D.4)$$

met  $\beta \equiv \rho - 1$ . De nulhypothese is nu dus  $\beta = 0$ , weer met  $c = 0$ . De alternatieve hypothese is  $-2 < \beta < 0$ , met willekeurige  $c$ . We toetsen dit eenvoudigweg door (D.4) op te vatten als een regressievergelijking en na te gaan of de schatting van  $\beta$  significant kleiner is dan nul. De toetsgrootte is de t-ratio van deze schatting. (De schatting van  $c$  wordt niet gebruikt.) We kunnen hierbij echter niet de normale tabel met kritieke waarden toepassen, omdat de "verklarende" variabele  $x_{t-1}$  stochastisch is en bepaald wordt door de reeks storingen  $\epsilon_t$ . Tabellen met kritieke waarden zijn gegeven door Dickey en Fuller (1979) en door Engle en Yoo (1987). Zie verder tabel 1; daarin is de kritieke waarde exclusief minteken opgenomen.

Een nadeel van deze methode is, dat de verdeling van de toetsgrootte verandert als we een  $c$  ongelijk nul toelaten in de nulhypothese. Er ontstaat dan een "random walk met drift". Zie tabel 1 "met trend": de kritieke waarden worden bijna anderhalf maal zo groot. Zie ook Jenkinson (1986).

c. De ADF-toets (Dickey en Fuller, 1981)

We voegen termen met vertraagde  $\Delta x$  toe aan het rechterlid van (D.4):

$$\Delta x_t = \beta x_{t-1} + \sum_{i=1}^p v_i \Delta x_{t-1} + c + \epsilon_t \quad (D.5)$$

De nulhypothese is dezelfde als bij de DF-toets. De ADF verschilt slechts van de DF door de aanwezigheid van de  $\Delta x$ -termen in het rechterlid van de regressievergelijking. Zie verder tabel 1. Ook hier het probleem van de "drift", net als bij de DF-toets.

d. Het autocorrelogram

Het autocorrelogram van een tijdreeks, zeg  $x_t$ , is de serie autocorrelaties van die reeks: de correlatie tussen  $x_t$  en  $x_{t-1}$ , gevolgd door die tussen  $x_t$  en  $x_{t-2}$ , enzovoort. Het al of niet stationair zijn van een reeks kan worden afgelezen aan het autocorrelogram. Als een reeks stationair is, nemen de autocorrelaties snel af. Bij een geïntegreerde reeks is dit niet zo.

Is de reeks te vaak gedifferentieerd (nog eens gedifferentieerd nadat zij reeds stationair was, dan is de eerste autocorrelatie ongeveer minus een half. Dit is eenvoudig af te leiden door uit te gaan van een stationaire reeks  $x_t$  die slechts bestaat uit "witte ruis", en vervolgens de correlatie tussen  $\Delta x_t$  en  $\Delta x_{t-1}$  uit te schrijven.

Tenslotte noemen we hier nog de Box-Pierce toetsgrootte: de kwadratensom van de autocorrelaties. Hiermee kan de hypothese dat alle autocorrelaties nul zijn, getoetst worden.

2. Toetsen op coïntegratie

Een toets op coïntegratie is een toets op integratie op het residu van de regressie van de (verondersteld) gecoïntegreerde variabelen op elkaar: de eerste stap in de tweestapsprocedure van Engle en Granger. De coïntegratietheorie vereist dat dit residu stationair is. De kritieke waarden van de toetsgroottes zijn afhankelijk van het aantal gecoïntegreerde variabelen; zie tabel 1, bij "aantal variabelen" groter dan één.



Nog enige opmerkingen over welke toetsgrootheid het best kan worden gebruikt. Engle en Granger (1987) bevelen voor het toetsen op coïntegratie de ADF toets aan. De kritieke waarden van deze toets zijn veel minder gevoelig dan van de DW-toets voor het achterliggende data-generatieproces, terwijl de power nage- noeg identiek is. Indien blijkt dat de coëfficiënten  $\theta_1$  in de ADF toets niet significant zijn, kan de DF-toets het beste worden gebruikt. Zoals hiervoor reeds is gezegd pleiten Sargan en Bhargava (1983) voor het inbouwen van de Imhof-routine in standaard regressie pakketten, waardoor de DW-toets bruik- baarder wordt. Jenkinson (1986) waarschuwt voor het gebruik van de DF en ADF-toets in geval  $\beta$  in deze toetsen naar 1 nadert. De power is in dat geval zeer gering. Onderzoek naar het autocorrellogram van het residu en een directe schattings-methode van het foutencorrectiemodel, zoals toegepast door Jenkinson, zijn dan andere toetsmogelijkheden.

Hieronder volgen nog een paar opmerkingen over de afzonderlijke toetsen.

a. DW-toets

De kritieke waarden van standaard DW-toets op de afwezigheid van autocorrela- tie in de storing ( $\rho = 0$ ) worden gewoonlijk gegeven in de vorm van onder- en bovengrenzen. Welke waarde in een bepaald geval van toepassing is, hangt af van de regressoren. (Bij gegeven regressors kan men deze waarde berekenen met de z.g. Imhof-procedure).

In ons geval, bij de toets op  $\rho = 1$ , doet zich dit in beginsel ook voor. Echter, bij het model van (D.2) en (D.3) is de enige regressor een constante, en de kritieke waarden kunnen dan precies bepaald worden, zonder een interval op te moeten geven. Bij de toets op coïntegratie is dat niet het geval. De kritieke waarden voor de DW-toets in tabel 1 voor meerdere variabelen zijn dan ook berekend onder een veronderstelling omtrent de gecoïntegreerde variabelen, namelijk dat zij een allersimpelst patroon volgen: hun mutatie is "witte ruis".

b. De DF-toets en de ADF-toets

De regressie die bij de DF-toets en bij de ADF-toets op coïntegratie wordt uitgevoerd bevat geen constante term, aangezien de "te verklaren" variabele in deze regressie zelf een residu is, met gemiddelde nul. (De verwachtingswaarde is nul, per veronderstelling, en de gemiddelde waarde van de realisatie is ook nul, als er in de coïntegratievergelijking een constante term is opgenomen.) Om dezelfde reden hebben we hier niet het probleem dat we rekening moeten houden met kritieke waarden die bepaald zijn onder de nulhypothese met c ongelijk aan nul ("drift").

Tabel 1 Kritieke waarden van de (co)integratietoetsen<sup>a)</sup>

Toets	aantal variabelen	steekproefomvang				
		25	50	100	200	∞
DW	1 <sup>b)</sup>	0,77 <sup>f)</sup>	0,49	0,26		0
	2		0,78	0,39	0,20	0
	3 <sup>c)</sup>			0,37		0
DF	1 zonder trend <sup>d)</sup>	1,95	1,95	1,95	1,95	1,95
	1 met trend <sup>e)</sup>	3,00	2,93	2,89		2,86
	2		3,67	3,37	3,37	
	3		4,11	3,93	3,78	
	4		4,35	4,22	4,18	
	5		4,76	4,58	4,48	
ADFE)	2	3,2	3,25	3,17	3,25	
	3	3,6	3,65	3,62	3,78	
	4	3,9	3,95	4,02	4,13	
	5	4,1	4,25	4,36	4,43	

a) Indien niet anders aangegeven is de DF en ADF-toets voor het geval zonder trend. De ADF toets betreft het geval voor  $p = 4$ .  
Bron: Engle en Yoo (1987) tenzij anders aangegeven.  
Alle kritieke waarden: significantieniveau = 95%.  
Bij DF en ADF is het minteken weggelaten.

b) Bron: Sargan en Bhargava (1983).

c) Bron: Hall (1986).

d) Bron: Fuller (1976), tabel 8.5.2, eerste gedeelte.

e) Bron: Fuller (1976), tabel 8.5.2, tweede gedeelte.

f) Steekproefomvang = 30.

g) Voor steekproefomvang 25 en 50 is de bron: Blangiewicz en Charemza (1989), tabel 3. (iets anders gedefinieerd dan Engle en Yoo, 1987.)