Pleidooi voor een nieuw criterium voor het selecteren van regressoren

Pieter H.F.M. van Casteren*

Samenvatting

De literatuur bevat tal van model selectie criteria. Aangezien deze nogal verschillend kunnen uitwerken, is de vraag welk criterium men het beste kan toepassen. Dit artikel beantwoordt deze vraag voor het geval van univariate lineaire regressie modellen met witte ruis storingen. Er worden diverse criteria vergeleken en beoordeeld, en bovendien wordt een potentiële verbetering van de criteria van Mallows (C_P) en Amemiya (PC) tot stand gebracht in de vorm van de criteria $SDMEP_\omega$ en $SIMEP_\omega$. Hierin stelt $\omega \in (0,1]$ een zogenaamde krimp-parameter voor, die gekozen moet worden door de toepasser. De kwaliteit van de diverse criteria en van de mogelijke keuzen van ω wordt onderzocht middels analytische afleidingen, asymptotische analyse en een Monte Carlo experiment. Hieruit komt een duidelijke voorkeur voor $SIMEP_\omega$ naar voren, alsmede een helder en eenduidig beeld aangaande de optimale keuze van ω.

1. Inleiding

Het selecteren van een model is een oud en lastig probleem, waarvoor allerlei verschillende model selectie regels aangedragen zijn. Wat daarbij opvalt is dat de voorgestelde methoden nogal verschillen in uitwerking, zodat het van groot belang is te weten welke methode de voorkeur verdient. In dit artikel wordt het specifieke geval beschouwd, waarin men moet beslissen welke regressoren op te nemen in een univariaat lineair model met witte ruis storingen. Voor dit geval wordt een kader opgebouwd, waarbinnen verduidelijkt kan worden welke methode het meest geschikt is en waarmee

* Vakgroep AKE, Faculteit der Economische Wetenschappen en Econometrie, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam. Tel. 020-5254232. Fax 020-5254100. E-mail casteren@fee.uva.nl. De auteur bedankt een anonieme referent voor het nuttige commentaar op een eerdere versie van dit artikel.

verbeterde model selectie regels kunnen worden afgeleid. Om dit nader toe te lichten worden eerst de belangrijkste methoden kort besproken en vergeleken. Daarbij zal er vanuit worden gegaan dat de storingsterm normaal verdeeld is.

Traditioneel wordt beslist over het al of niet opnemen van regressoren door het uitvoeren van F-toetsen met een onbetrouwbaarheidsdrempel van 5%. Gaat het om het al of niet opnemen van één regressor (de F-toets reduceert dan tot een t-toets) en is het aantal waarnemingen (n) groot, dan komt deze methode neer op een kritieke F-waarde van ca. 4, dus een kritieke $|t|$-waarde van ca. $\sqrt{4} = 2$. De bekende kritiek op deze methode is dat onduidelijk is waarom precies het 5% significantieniveau genomen zou moeten worden. Bovendien kan het bij deze methode gebeuren dat "model a" verkozen wordt boven "model b", "model b" boven "model c" en "model c" boven "model a" (cycling), zodat deze methode niet tot een eenduidige model-voorkeur hoeft te leiden. Toch is deze methode ongetwijfeld nog steeds de meest gebruikte.

Een jongere methode is afkomstig van Theil (1961 blz. 212-214) en deze behelst het minimaliseren van de residuele variantie of, equivalent daarmee, het maximaliseren van de gecorrigeerde meervoudige determinatiecoëfficiënt $\bar{R}^2$. Deze methode is qua uitwerking ongeveer hetzelfde als F-toetsen met een kritieke waarde gelijk aan 1 (bij grote n). De motivatie voor deze methode is dat de verwachting van de residuele variantie minimaal is voor een waar model, d.w.z. een model waarin alle *relevante* regressoren (met een coëfficiënt ongelijk aan 0) zijn opgenomen. Hiermee wordt echter niet voorkomen dat ook *irrelevante* regressoren (met een coëfficiënt gelijk aan 0) worden opgenomen en om die reden is Theil's methode onbevredigend.

Een nog recentere methode is geïnitieerd door Mallows (1973) en later verfraaid door Amemiya (1980). Het uitgangspunt hierbij is dat een regressiemodel goed is als het kleine voorspelfouten genereert. Dit kan worden geformaliseerd binnen een beslissingstheoretisch kader en vervolgens kunnen selectie criteria worden afgeleid, zoals Mallows' C_p (ook wel aangeduid als MC) en Amemiya's "Prediction Criterion" (PC). Bij grote n komen deze criteria overeen met een kritieke F-waarde van ca. 2. Nu is wel een mechanisme ingebouwd om zoveel mogelijk de relevante regressoren wel en de irrelevante regressoren niet in het model op te nemen. Deze methode wordt in dit artikel uitgebreid aan de orde gesteld en daarbij zal blijken dat de afleiding van selectie criteria nog verbeterd kan worden.

Er zijn diverse methoden, die lijken op de voorgaande, doordat zij eveneens gericht zijn op het minimaliseren van de voorspelfout en qua uitwerking bij grote n ook corresponderen met een kritieke F-waarde van ca. 2. Genoemd kunnen worden het "Generalized Cross Validation criterion" (GCV) van Golub, Heath en Wahba (1979), Hocking's criterium S_p (Hocking 1976, Breiman en Freedman 1983) en voor grote n tevens "Akaike's Information Criterion" (AIC, Akaike 1974) en Sawa's informatie criterium BIC (Sawa 1978).

Tenslotte verdient ook de methode van Schwarz (1978) vermelding (zie ook Chow

1981). Deze behelst een Bayesiaanse aanpak, erop gericht om asymptotisch het model te selecteren dat de hoogste posteriori kans heeft een waar model van minimale omvang te zijn. Via een afleiding wordt dan het "Schwarz Bayesian Information Criterion" (SBIC) verkregen. Een ietwat gewijzigde versie hiervan, speciaal voor lineaire regressie modellen, is het "Bayesian Estimation Criterion" (BEC) van Geweke en Meese (1981). Beide criteria corresponderen bij grote n met een kritieke F-waarde ongeveer gelijk aan ln n (de natuurlijke logaritme van n). Een nadeel is dat deze methode is gebaseerd op de assumptie dat er een eindig waar model bestaat, hetgeen alleen zinvol is als er onder de alternatieve modellen ook minstens één is dat voor waar (of nagenoeg waar) kan doorgaan. In de praktijk is vaak onduidelijk of deze assumptie gerechtvaardigd is, zodat ook onduidelijk is of de afgeleide criteria wel zinvol zijn.

Samenvattend kunnen de bestaande methoden voor het selecteren van een regressie model geclusterd worden aan de hand van de kritieke F-waarden, waarmee zij voor grote n bij benadering corresponderen; zie tabel 1. Hierboven is fundamentele kritiek geleverd op alle methoden, behalve die in het tweede cluster. Binnen dit cluster is de beslissingstheoretische aanpak van Amemiya (1980) wel heel fraai en bovendien toegesneden op regressie modellen en het is dan ook niet toevallig dat deze aanpak de basis vormt van dit artikel.

Tabel 1. Clustering van model selectie regels op basis van kritieke F-waarden.

Clusters van model selectie regels	Benadering kritieke F-waarde (n groot)
residuele variantie (of $\bar{R}^2$)	1
C_P (of MC), PC, GCV, S_P, AIC, BIC	2
t-toets met 5% significantie niveau	4
SBIC, BEC	ln n

De werkwijze in dit artikel is Amemiya's aanpak verder uit te bouwen, ten eerste met een zogenaamde krimp-schatter, onder invoering van de krimp-parameter $0 < \omega \leq 1$, en ten tweede met het zogenaamde marginale substitutie verhouding principe. Daarbij wordt een familie van criteria verkregen, corresponderend (voor grote n) met een kritieke F-waarde van ca. $1+1/\omega$. Bij de keuze $\omega = 1$ verkrijgt men dus soortgelijke criteria als C_P en PC en door ω vrij te laten binnen $0 < \omega \leq 1$ ontstaat een familie van criteria, die correspondeert met alle mogelijke kritieke F-waarden van 2 tot oneindig (dit omvat 3 van de 4 clusters). Het zal duidelijk gemaakt worden dat elk lid van deze familie optimaal kan zijn onder bepaalde omstandigheden. Het eindresultaat is dat men inzicht krijgt in de relevantie van de bestaande selectie methoden en dat men in staat gesteld wordt een weloverwogen keuze te maken voor een bepaalde waarde van ω, corresponderend met een bepaalde kritieke F-waarde. Aangezien de gekozen kritieke F-waarde niet beperkt hoeft te

zijn tot de verzameling {2, 4, ln n}, ligt een verbetering van de bestaande methoden in het verschiet.

De opbouw van dit artikel is als volgt. In §2 wordt het model selectie probleem exact geformuleerd door het specificeren van de alternatieve modellen en de doelstelling. Vervolgens worden in §3 model selectie criteria afgeleid door uitbouw van Amemiya's aanpak. Daarbij wordt in §3.1 de krimp-parameter ω geïntroduceerd en wordt geanalyseerd wat de optimale waarde van ω is. Niettemin wordt de waarde van ω in principe vrij gelaten, zodat in §3.2, na toepassing van het marginale substitutie verhouding principe, een familie van selectie criteria verkregen wordt. In §4 worden de alternatieve criteria vergeleken wat betreft de neiging om een klein model te selecteren (zuinigheid) en in §5 worden ze omgezet in kritieke F-waarden met de bijbehorende onbetrouwbaarheidsdrempels. In §6 wordt aan de hand van twee typen asymptotische analyse besproken wat de asymptotisch optimale waarde van ω is. In §7 worden de resultaten van de voorgaande paragrafen geïllustreerd door een Monte Carlo experiment, met bijzondere aandacht voor de optimale keuze van ω. In §8 wordt kort uitgelegd dat men de afleiding van AIC op soortgelijke wijze kan uitbouwen m.b.v. krimp-schatting, waarbij men een familie van informatie criteria verkrijgt. Tenslotte worden in §9 de belangrijkste conclusies gegeven.

2. Formulering van het selectieprobleem

In dit artikel wordt verondersteld dat het volgende lineaire regressiemodel een foutloze weergave van de werkelijkheid is:

$$y_i = x_i'\beta + \varepsilon_i \qquad (i = 1,2,3\ldots), \tag{1}$$

waarin $\beta = (\beta_1,\ldots.\beta_k)'$ en x_i vectoren van constanten zijn en ε_i een stochastische variabele met $E[\varepsilon_i] = 0$, $E[\varepsilon_i^2] = \sigma^2 > 0$ en $E[\varepsilon_i\varepsilon_j] = 0$ voor $i \neq j$ (witte ruis). De foutloosheid van het bovenstaande model (men spreekt van een "waar model") impliceert dat er geen weggelaten regressoren zijn. Het sluit echter niet uit dat enkele parameters in β gelijk aan nul kunnen zijn. In dat geval zou weglating van alle irrelevante regressoren (met $\beta_j = 0$) leiden tot een waar model van kleinere omvang. Verder wordt verondersteld dat n waarnemingen $y = (y_1,\ldots,y_n)'$ en $X = (x_1,\ldots,x_n)'$ beschikbaar zijn en dat β, $\varepsilon = (\varepsilon_1,\ldots\varepsilon_n)'$ en σ^2 onbekend zijn.

Indien rang(X) = k, dan kan men kleinste kwadraten toepassen waarbij β geschat wordt door $\hat{\beta} = (X'X)^{-1}X'y$ en $X\beta$ door $X\hat{\beta}$. Deze methode heeft de beperking dat zij niet toepasbaar is als $k > n$ (en weinig zinvol als $k = n$), maar ook afgezien daarvan kan men zich afvragen of het wellicht beter is om kleinste kwadraten toe te passen onder weglating van de minder belangrijke regressoren. Beschouw daarom de matrix X_1, bestaande uit k_1 ($\leq$ k) kolommen van X. Indien rang(X_1) = k_1, dan vindt men $\hat{\beta}_1$ =

$(X_1'X_1)^{-1}X_1'y$ en de corresponderende schatter van β, zeg $\hat{\beta}_{(1)}$, wordt verkregen door juiste ordening van de k_1 elementen van $\hat{\beta}_1$ en $k-k_1$ nullen. De bijbehorende schatter van $X\beta$ is $X\hat{\beta}_{(1)} = X_1\hat{\beta}_1$. Het probleem is nu een matrix X_1 te kiezen uit een gegeven verzameling alternatieven, zeg $\{X_1\}$. Hierbij wordt verondersteld dat $\text{rang}(X_1) = k_1 < n$ voor alle $X_1 \in \{X_1\}$. Merk op dat $X \in \{X_1\}$ mogelijk is, mits $\text{rang}(X) = k < n$.

Om te kunnen beslissen welke keuze uit $\{X_1\}$ de beste is, dient men een doelstelling te kiezen. Stel dat het doel is $X\beta$ zo goed mogelijk te schatten, en dat de kwaliteit van de schatting $X\hat{\beta}_{(1)}$ kan worden gemeten door de gemiddelde kwadratische schattingsfout $(X\beta-X\hat{\beta}_{(1)})'(X\beta-X\hat{\beta}_{(1)})/n$. Gegeven deze verliesfunctie kan de kwaliteit van de schatter $X\hat{\beta}_{(1)}$ gemeten worden door het verwachte verlies (risico)

$$\rho_{ME} = E[(X\beta-X\hat{\beta}_{(1)})'(X\beta-X\hat{\beta}_{(1)})/n] = E[(\beta-\hat{\beta}_{(1)})'(n^{-1}X'X)(\beta-\hat{\beta}_{(1)})], \tag{2}$$

waarin het subscript ME staat voor "mean estimation". Minimalisering van ρ_{ME} over alle modellen in $\{X_1\}$ levert het beste model X_1 en de beste schatter $X\hat{\beta}_{(1)}$). Het probleem is echter dat ρ_{ME} afhankelijk is van onbekende parameters en daardoor als zodanig niet toepasbaar. Om daadwerkelijk een model X_1 te kunnen selecteren zal de doelfunctie ρ_{ME} dus geoperationaliseerd moeten worden (zie §3). Eerst wordt echter een tweede doelfunctie bekeken.

Stel dat de doelstelling is een nog niet waargenomen waarde van y_i, zeg y_p, zo goed mogelijk te voorspellen op basis van gegeven regressoren-waarden x_p. Stel voorts dat de kwaliteit van de voorspelling $x_p'\hat{\beta}_{(1)}$ kan worden gemeten door de verliesfunctie $(y_p-x_p'\hat{\beta}_{(1)})^2$, dan kan de kwaliteit van de voorspeller $x_p'\hat{\beta}_{(1)}$ worden gemeten door het risico

$$\rho_P = E[(y_p-x_p'\hat{\beta}_{(1)})^2]. \tag{3}$$

Hierin geldt het verwachtingsteken m.b.t. $\hat{\beta}_{(1)}$ (dus y) en y_p en ook m.b.t. x_p, wanneer x_p nog niet bekend is. Conditioneel op x_p geldt

$$\begin{aligned} E[(y_p-x_p'\hat{\beta}_{(1)})^2|x_p] &= E[(y_p-x_p'\beta)^2|x_p] + E[(x_p'\beta-x_p'\hat{\beta}_{(1)})^2|x_p] \\ &= \sigma^2 + E[(\beta-\hat{\beta}_{(1)})'x_px_p'(\beta-\hat{\beta}_{(1)})|x_p], \end{aligned}$$

en het nemen van de verwachting m.b.t. x_p leidt tot

$$\rho_P = \sigma^2 + E[(\beta-\hat{\beta}_{(1)})'x_px_p'(\beta-\hat{\beta}_{(1)})]. \tag{4}$$

Door te veronderstellen dat x_p onafhankelijk is van y (en dus van $\hat{\beta}_{(1)}$) en dat

$$E[x_px_p'] = n^{-1}X'X \tag{5}$$

krijgt men

$$\rho_P = \sigma^2 + \rho_{ME}. \tag{6}$$

Gezien (3) en (5) kan ρ_P geïnterpreteerd worden als een *theoretische maatstaf voor de voorspelkwaliteit* en minimalisatie hiervan over $\{X_1\}$ levert het beste model. Daar σ^2 een constante is, maakt het blijkens (6) niet uit of gestreefd wordt naar minimalisatie van ρ_{ME} of minimalisatie van ρ_P. Bijgevolg hoeft in principe slechts één van beide doelfuncties geoperationaliseerd te worden.

3. Afleiding van selectie criteria

In deze paragraaf wordt de taak opgevat de doelstellingsfunctie ρ_P (en impliciet ook ρ_{ME}) om te zetten in een toepasbaar criterium. Hierbij treedt helaas een moeilijkheid op, indien dezelfde data gebruikt worden voor zowel het selecteren van X_1 als het schatten van β d.m.v. $\hat{\beta}_{(1)}$. In dit geval zou men bij het evalueren van de verwachting in (2) de kansverdeling van $\hat{\beta}_{(1)}$ eigenlijk moeten conditioneren op het feit dat X_1 geselecteerd is uit $\{X_1\}$ (men spreekt hier van "selection bias", zie Miller 1990). Een dergelijke conditionering is echter pas mogelijk, wanneer de keuze verzameling $\{X_1\}$, alsmede het selectie criterium en de zoekprocedure bekend zijn. Aangezien deze drie zaken onbekend zijn (het selectie criterium moet juist afgeleid worden!), is conditionering onmogelijk en derhalve zal gewoon de onconditionele verdeling van $\hat{\beta}_{(1)}$ gehanteerd worden, zoals neergelegd in de assumpties van paragraaf 2.

Gegeven (6) en (2) kan ρ_P als volgt worden uitgewerkt:

$$\begin{aligned}
\rho_P &= \sigma^2 + E[(X\hat{\beta}_{(1)} - E[X\hat{\beta}_{(1)}])'(X\hat{\beta}_{(1)} - E[X\hat{\beta}_{(1)}])]/n + (X\beta - E[X\hat{\beta}_{(1)}])'(X\beta - E[X\hat{\beta}_{(1)}])/n \\
&= \sigma^2 + E[\{X_1(X_1'X_1)^{-1}X_1'\varepsilon\}'\{X_1(X_1'X_1)^{-1}X_1'\varepsilon\}]/n + (M_1X_2\beta_2)'(M_1X_2\beta_2)/n \\
&= \sigma^2 + E[\text{spoor}\{X_1(X_1'X_1)^{-1}X_1'\varepsilon\varepsilon'X_1(X_1'X_1)^{-1}X_1'\}]/n + \beta_2'X_2'M_1X_2\beta_2/n \\
&= \sigma^2 + \sigma^2 k_1/n + \gamma_1/n,
\end{aligned} \tag{7}$$

waarin X_2 is gedefinieerd als de kolommen van X die niet in X_1 zijn vervat, β_2 als de vector van coëfficiënten behorende bij X_2, $M_1 = I - X_1(X_1'X_1)^{-1}X_1'$ en $\gamma_1 = \beta_2'X_2'M_1X_2\beta_2$. De drie componenten waaruit ρ_P is opgebouwd zijn respectievelijk de variantie van y_p, de gemiddelde variantie van $x_i'\hat{\beta}_{(1)}$ en de gemiddelde kwadratische vertekening van $x_i'\hat{\beta}_{(1)}$. Blijkens (7) komt het minimaliseren van ρ_P of ρ_{ME} neer op het maken van de juiste afweging tussen de model omvang k_1 en de vertekeningsfactor γ_1. Het komt er dus op aan deze afweging zo goed mogelijk te operationaliseren.

Een voor de hand liggende operationalisering ontstaat door voor ieder model $X_1 \in \{X_1\}$ de waarde van ρ_P zo goed mogelijk te schatten. Echter, bij het vergelijken van

modellen zijn de waarden van ρ_P op zich niet van belang, maar wel de *verschillen* tussen de waarden van ρ_P voor de alternatieve modellen en daarom is het effectiever om deze verschillen zo goed mogelijk te schatten. Stel dat men een voorkeur wil bepalen tussen twee modellen, zeg X_a, $X_b \in \{X_1\}$, dan is het verschil in doelwaarde

$$\Delta\rho_P = \sigma^2(k_a - k_b)/n + (\gamma_a - \gamma_b)/n. \tag{8}$$

Goede schatting van dit verschil kan nu worden vertaald in goede schatting van $\gamma_a - \gamma_b$ en σ^2. Het is handig dit probleem d.m.v. conditionering op te splitsen in twee deelproblemen:

(1) schatting van $\delta = \gamma_a - \gamma_b$ conditioneel op σ^2,
(2) schatting van σ^2.

Deze deelproblemen worden achtereenvolgens besproken in subparagrafen 3.1 en 3.2.

3.1. Toepassing van krimp-schatting

In deze subparagraaf wordt conditioneel op σ^2 een goede schatter gezocht van $\delta = \gamma_a - \gamma_b$. Uit

$$E[y'M_1y] = E[(M_1X_2\beta_2 + M_1\varepsilon)'(M_1X_2\beta_2 + M_1\varepsilon)] = \gamma_1 + \sigma^2(n - k_1) \tag{9}$$

volgt dat $\hat{\gamma}_1 = y'M_1y - \sigma^2(n - k_1)$ conditioneel op σ^2 een zuivere schatter is van γ_1. Bijgevolg is

$$\hat{\delta} = \hat{\gamma}_a - \hat{\gamma}_b = (y'M_ay - y'M_by) + \sigma^2(k_a - k_b) \tag{10}$$

conditioneel op σ^2 een zuivere schatter van δ. Nu blijkt uit Thompson (1968) en Blight (1971) dat de schatter $\hat{\delta}$ verbeterd kan worden, indien het gerechtvaardigde vermoeden bestaat dat de afstand tussen δ en een zekere waarde δ_0 klein is in verhouding tot de variantie van $\hat{\delta}$. Zoals hieronder zal blijken is deze situatie hier zeer relevant, omdat men in de praktijk een keuze wenst te maken tussen modellen met een kleine vertekeningsfactor γ_1, waardoor δ meestal dichtbij $\delta_0 = 0$ zal liggen. In het uiterste geval zijn X_a en X_b beide "ware" modellen, waarvoor exact geldt $\delta = 0$. De verbetering van de schatter bestaat hieruit dat $\hat{\delta}$ getrokken wordt in de richting van $\delta_0 = 0$ door de "krimp-schatter" $\omega\hat{\delta}$ te hanteren, waarin ω een nader te bepalen "krimp-parameter" is, die voldoet aan $0 < \omega \leq 1$ (door ook $\omega = 1$ toe te laten wordt $\omega\hat{\delta}$ een generalisatie van $\hat{\delta}$). Om de verbetering duidelijk te maken wordt verondersteld dat gestreefd wordt naar

minimalisering van de verwachte kwadratische schattingsfout

$$E[(\omega\hat{\delta} - \delta)^2] = \omega^2 \, Var[\hat{\delta}] + (1-\omega)^2 \, \delta^2. \tag{11}$$

De optimale waarde van ω kan dan worden verkregen door de afgeleide naar ω gelijk aan nul te stellen:

$$\omega^* = \frac{\delta^2}{\delta^2 + Var[\hat{\delta}]}. \tag{12}$$

De *optimale krimp-schatter* van δ is dus $\omega^*\hat{\delta}$. Merk op dat $0 \leq \omega^* < 1$ en dat ω^* kleiner is naarmate $\delta^2/Var[\hat{\delta}]$ kleiner is (aangenomen dat $Var[\hat{\delta}] > 0$). Indien X_a en X_b beide "ware" modellen zijn, dan geldt $\delta = 0$ en dus $\omega^* = 0$. Het relatieve voordeel van $\omega^*\hat{\delta}$ t.o.v. $\hat{\delta}$ is af te lezen aan

$$\frac{E[(\omega^*\hat{\delta}-\delta)^2]}{E[(\hat{\delta}-\delta)^2]} = \frac{\delta^4 Var[\hat{\delta}] + (Var[\hat{\delta}])^2 \, \delta^2}{(\delta^2+Var[\hat{\delta}])^2 \, Var[\hat{\delta}])} = \omega^*. \tag{13}$$

Bij kleinere ω^* is het relatieve voordeel dus groter.

Helaas is $\omega^*\hat{\delta}$ slechts een utopische schatter, aangezien ω^* onbekend is en bovendien afhangt van het paar modellen X_a, $X_b \in \{X_1\}$. Om die reden wordt nu gepoogd iets meer te zeggen over de grootte van ω^*. Veronderstel dat de kolommen van X_a een deelverzameling vormen van de kolommen X_b (d.w.z. model X_a is genest in model X_b) en dat $k_a < k_b$, dan geldt $\delta \geq 0$ (dit is intuïtief onmiddellijk duidelijk; voor een bewijs zie Van Casteren 1994 blz. 154). Veronderstel bovendien dat ε_i normaal verdeeld is, dan heeft $\hat{\delta}/\sigma^2 + k_b - k_a = (y'M_a y - y'M_b y)/\sigma^2$ een niet-centrale χ^2-verdeling met $k_b - k_a$ vrijheidsgraden en niet-centraliteitsparameter $\delta/\sigma^2 \geq 0$ (zie Van Casteren 1994 blz. 155). Dit betekent dat

$$Var[\hat{\delta}] = \sigma^4 \, Var[\hat{\delta}/\sigma^2 + k_b - k_a] = 4\sigma^2 \, \delta + 2\sigma^4 \, (k_b - k_a). \tag{14}$$

Substitutie hiervan in (13) geeft de optimale waarde van ω:

$$\omega^*_{NN} = \frac{\delta^2/\sigma^4}{\delta^2/\sigma^4 + 4\delta/\sigma^2 + 2(k_b - k_a)}, \tag{15}$$

waarin het subscript NN de beperking tot geneste modellen en normaliteit uitdrukt.

Vergelijking (15) biedt nog steeds geen bruikbare inperking van de optimale waarde van ω, want ω^*_{NN} is nog steeds afhankelijk van δ. Deze afhankelijkheid kan worden opgeheven door uit te gaan van de situatie waarin beide modellen X_a en X_b even goed zijn, d.w.z. $\Delta\rho_P = 0$. Dit is een redelijk uitgangspunt, want indien men zou uitgaan

van $\Delta\rho_P > 0$ of $\Delta\rho_P < 0$, dan zou men bevooroordeeld zijn ten gunste van één van beide modellen. Nu leidt $\Delta\rho_P = 0$ via (8) tot $\delta = \sigma^2(k_b - k_a)$ en substitutie hiervan in (15) geeft

$$\omega^*_{NNE} = \frac{1}{1 + 6/(k_b - k_a)}, \qquad (16)$$

waarin de toevoeging van het subscript E het evenwichtige, onbevooroordeelde uitgangspunt uitdrukt. Met (16) is de optimale waarde van ω teruggebracht tot een uitdrukking, die berekend kan worden voor gegeven modellen X_a en X_b. Helaas is ω^*_{NNE} nog wel afhankelijk van het verschil in aantal parameters. Tabel 2 geeft een kwantificering van deze afhankelijkheid (voor later gebruik is tevens $1+1/\omega^*_{NNE}$ berekend).

Tabel 2. De optimale waarde ω^*_{NNE} (en $1+1/\omega^*_{NNE}$) als functie van $k_b - k_a$.

$k_b - k_a$	1	2	3	4	5	6	7	8	9	10	∞
ω^*_{NNE}	0.14	0.25	0.33	0.40	0.45	0.50	0.54	0.57	0.60	0.60	1
$1+1/\omega^*_{NNE}$	8	5	4	3.5	3.2	3.0	2.68	2.75	2.67	2.60	2

Als de normaliteitsassumptie acceptabel is en $\{X_1\}$ bevat slechts twee modellen, die ook nog genest zijn, dan kan men via (16) de keuze van ω bepalen. In de praktijk echter hoeven niet alle mogelijke paren modellen genest te zijn zodat (16) niet altijd toepasbaar is en $\{X_1\}$ kan meer dan één paar geneste modellen bevatten, zodat er meerdere waarden van ω^*_{NNE} kunnen zijn. Toch verschaft (16) ook in dergelijke gevallen bruikbare informatie. Vergelijkingen van niet geneste paren modellen kunnen vaak in stappen uitgevoerd worden via twee of meer "geneste" model vergelijkingen. Wanneer er meerdere geneste paren modellen zijn, dan kan men zich afvragen welke modellen *a priori* als het meest concurrerend (met een lage waarde van ρ_P resp. ρ_{ME}) aangemerkt kunnen worden, waaruit een indicatie verkregen wordt welke waarden van $k_b - k_a$ het meest relevant zijn. Immers, het elimineren van apert slechte modellen is gemakkelijker dan het selecteren van het beste model onder de goede modellen, en dit laatste stelt daarom hogere eisen aan de keuze van ω.

Een redelijke aanpak is de mogelijke waarden van $k_b - k_a$ a priori te wegen en ω te bepalen als gewogen gemiddelde van de bijbehorende waarden van ω^*_{NNE}. Men zou kunnen stellen dat in de praktijk de mogelijke waarden van $k_b - k_a$ vaak niet boven 6 uitkomen, zodat $k_b - k_a$ varieert van 1 tot en met 6. Binnen dit domein varieert ω^*_{NNE} van 0.14 tot en met 0.5 en is het ongewogen gemiddelde 0.35. Dit duidt erop dat het in de praktijk vaak voordelig is om ω beduidend kleiner dan 1 te kiezen, waarbij volgens (13) het relatieve voordeel t.o.v. $\omega=1$ flink kan oplopen. Deze conclusie, die nog nader

ondersteund zal worden d.m.v. asymptotische analyse (§6) en een Monte Carlo experiment (§7), vormt een overtuigende rechtvaardiging voor het introduceren van de krimp-parameter ω. Daar het moeilijk is om in de praktijk een exacte waarde voor ω te kiezen, zal de keuze van ω nog verder worden besproken in §6 en §7.

Het gebruik van de krimp-schatter $\omega\hat{\delta}$ van $\delta = \gamma_a - \gamma_b$ kan geëffectueerd worden door γ_1 in de doelfunctie (7) te vervangen door $\omega\hat{\gamma}_1$, zodat ρ_p verandert in:

$$\begin{aligned}\rho_P' &= \sigma^2 + \sigma^2 k_1/n + \omega\hat{\gamma}_1/n \\ &= \sigma^2 + \sigma^2 k_1/n + \omega[y'M_1y - \sigma^2(n-k_1)]/n \\ &= \sigma^2[1-\omega + (1+\omega)k_1/n] + \omega y'M_1y/n. \end{aligned} \tag{17}$$

Immers, er geldt dan

$$\Delta\rho_P' = \sigma^2(k_a-k_b)/n + \omega(\hat{\gamma}_a - \hat{\gamma}_b)/n, \tag{18}$$

zodat $\delta = \gamma_a - \gamma_b$ in (8) *de facto* wordt geschat door $\omega\hat{\delta} = \omega(\hat{\gamma}_a - \hat{\gamma}_b)$.

3.2. Schatting van de variantie

Gegeven de uitdrukking voor ρ_P' in (17) resteert de taak om de minimalisatie van ρ_P' verder te operationaliseren via schatting van σ^2. Onder de veronderstelling dat rang(X) = $k < n$, kan m.b.v. het meest ruime model X de zuivere schatter $\hat{\sigma}^2 = y'My/(n-k)$ met $M = I - X(X'X)^{-1}X'$ bepaald worden. Echter, onder normaliteit van ε_i heeft $\tilde{\sigma}^2 = y'My/(n-k+2)$ een lagere verwachte kwadratische schattingsfout dan $\hat{\sigma}^2$, zie b.v. Theil (1971 blz. 128). Daarom kan men bij normaliteit van ε_i de voorkeur geven aan $\tilde{\sigma}^2$. Merk op dat $\tilde{\sigma}^2$ in feite een krimp-schatter is, afgeleid van $\hat{\sigma}^2$, met krimp-parameter $(n-k)/(n-k+2)$. Kiest men $\tilde{\sigma}^2$, dan wordt $\Delta\rho_P'$ geschat door

$$\tilde{\sigma}^2(k_a-k_b)/n + \omega(\hat{\gamma}_a - \hat{\gamma}_b)/n$$

en dit wordt geëffectueerd door

$$\tilde{\sigma}^2[1-\omega + (1+\omega)k_1/n] + \omega y'M_1y/n$$

als schatter van ρ_P' te nemen. Omdat $\omega > 0$, is minimalisering hiervan equivalent met minimalisering van (monotoon stijgende transformatie):

$$\text{SDMEP}_\omega = [y'M_1y + (1+1/\omega)\tilde{\sigma}^2 k_1]/n \qquad (0 < \omega \leq 1), \tag{19}$$

waarin SDMEP staat voor "Set-Dependent Mean Estimation and Prediction criterion". Wanneer $\hat{\sigma}^2$ wordt gebruikt i.p.v. $\tilde{\sigma}^2$ dan verkrijgt men een generalisatie van Mallows' criterium (C_P, zie Mallows 1973; MC, zie Amemiya 1980), die voor $\omega = 1$ equivalent is met Mallows' criterium. Met "set-dependent" wordt bedoeld dat SDMEP via $\tilde{\sigma}^2$ afhangt van de complete verzameling regressoren X. Dit heeft het nadeel dat eerst duidelijk moet zijn welke regressoren deel van X uitmaken, voordat SDMEP berekend kan worden, en bovendien is vereist dat rang(X) = k < n. Daarom wordt hieronder een zogenaamd "set-independent" criterium ontwikkeld.

Een alternatieve aanpak is de schatter van σ^2 te baseren op X_1, d.w.z. het model dat men juist overweegt te selecteren. Analoog aan $\hat{\sigma}^2$ en $\tilde{\sigma}^2$ kunnen gedefinieerd worden: $\hat{\sigma}_1^2 = y'M_1y/(n-k_1)$ en $\tilde{\sigma}_1^2 = y'M_1y/(n-k_1+2)$. Merk op dat $\hat{\sigma}_1^2$ een niet negatieve vertekening heeft, d.w.z. $E[\hat{\sigma}_1^2]-\sigma^2 \geq 0$, zoals blijkt uit (9) en $\gamma_1 \geq 0$. De vertekening van $\tilde{\sigma}_1^2$ kan positief of negatief zijn. Onder normaliteit van ε_i kan men $\tilde{\sigma}_1^2$ prefereren boven $\hat{\sigma}_1^2$, vanwege een lagere verwachte kwadratische schattingsfout (zie Van Casteren 1994 blz. 79).

Indien men in (17) de onbekende σ^2 zou vervangen door de schatter $\hat{\sigma}_1^2$ of $\tilde{\sigma}_1^2$, dan zou dit de modellen met een relatief hoge schatting (meestal de kleinere modellen) direct benadelen en bovendien zou σ^2 in (18) dan gelijktijdig op twee verschillende manieren geschat worden. Derhalve is dit geen geschikte mogelijkheid. Een betere aanpak is de volgende. Volgens (17) moet er bij het vergelijken van modellen een afweging plaatsvinden tussen k_1 en $y'M_1y$ en de manier waarop dit moet gebeuren wordt vastgelegd door de richting van de gradiënt van ρ_P' (hierbij wordt k_1 beschouwd als een continue variabele), d.w.z. door de marginale substitutie verhouding

$$\frac{\partial \rho_p'/\partial k_1}{\partial \rho_p'/\partial (y'M_1y)} = (1+\frac{1}{\omega})\sigma^2 \qquad (20)$$

en de tekens van de partiële afgeleiden (beide zijn positief). De gevraagde afweging tussen k_1 en $y'M_1y$ kan dus geoperationaliseerd worden door σ^2 in (20) te vervangen door een schatter en vervolgens te bekijken welke criteriumfunctie daarmee overeenstemt. Men kan dit het "marginale substitutie verhouding principe" noemen. Hanteert men bijvoorbeeld de schatter $\tilde{\sigma}^2$, net als hierboven, dan voldoet minimalisering van $SDMEP_\omega$ aan dit principe, omdat

$$\frac{\partial SDMEP_\omega /\partial k_1}{\partial SDMEP_\omega /\partial (y'M_1y)} = (1+\frac{1}{\omega})\tilde{\sigma}^2, \qquad (21)$$

terwijl beide partiële afgeleiden positief zijn. De bedoeling was echter de schatter $\tilde{\sigma}_1^2$ te hanteren en in dat geval voldoet minimalisering van het criterium

$$\text{SIMEP}_\omega = \frac{y'M_1y}{(n-k_1+2)^{1+1/\omega}} \qquad (0 < \omega \leq 1), \tag{22}$$

genaamd "Set-Independent Mean Estimation and Prediction criterion". Immers, er geldt

$$\frac{\partial \text{SIMEP}_\omega / \partial k_1}{\partial \text{SIMEP}_\omega / \partial (y'M_1y)} = (1+\frac{1}{\omega})\tilde{\sigma}_1^2 \tag{23}$$

en beide partiële afgeleiden zijn positief. Op soortgelijk wijze voldoet minimalisering van $y'M_1y/(n-k_1)^{1+1/\omega}$, indien $\hat{\sigma}_1^2$ als schatter van σ^2 gehanteerd wordt. Hier moet worden opgemerkt dat SIMEP_ω niet rechtstreeks van ρ_P' is afgeleid, maar door uitproberen totdat een criterium was gevonden, dat voldoet aan (23). Overigens voldoet ook iedere monotoon stijgende transformatie van SIMEP_ω. Verder kan SIMEP_ω niet zomaar opgevat worden als schatter van ρ_P' of ρ_P. Indien men ρ_P wil schatten op basis van X_1 na toepassing van SIMEP_ω, dan kan dit door achtereenvolgens γ_1 en σ^2 in (7) te vervangen door $\hat{\gamma}_1$ en $\tilde{\sigma}_1^2$, hetgeen leidt tot $\tilde{\rho}_{P1} = \tilde{\sigma}_1^2(n+k_1+2)/n = y'M_1y/n$. Soortgelijk kan men $\rho_{ME} = \rho_p - \sigma^2$ schatter door $\tilde{\rho}_{ME1} = \tilde{\sigma}_1^2 (k_1+2)/n$.

In vergelijking met SDMEP_ω heeft SIMEP_ω het praktische voordeel dat niet eerst X bepaald hoeft te worden en bovendien is rang(X) = k < n niet vereist. Welk van beide schatters $\tilde{\sigma}_1^2$ (in SIMEP_ω) en $\tilde{\sigma}^2$ (in SDMEP_ω) de beste is, is niet direct te zeggen (zie Van Casteren 1994, hoofdstuk 4).

Om een vergelijking van SDMEP_ω en SIMEP_ω met informatiecriteria mogelijk te maken wordt ook nog een derde criterium geïntroduceerd. Stel dat σ^2 in (20) geschat wordt door de maximum likelihood (ML) schatter $y'M_1y/n$ (onder normaliteit van ε_i), dan voldoet

$$\text{SIIC}_\omega = n \ln(y'M_1y/n) + (1+1/\omega)(k_1+1), \tag{24}$$

genaamd "Set Independent Information Criterion", aan het marginale substitutie verhouding principe. Dit criterium is een bijzonder geval van het informatiecriterium, dat wordt besproken in §8, en komt overeen met AIC als $1+1/\omega = 2$ (d.w.z. $\omega=1$) en met SBIC als $1+1/\omega = \ln n$. Omdat SIIC_ω hier is afgeleid op een soortgelijke manier als SDMEP_ω en SIMEP_ω, kunnen alle drie ook op dezelfde manier beoordeeld worden. In dit artikel wordt overigens niet veel aandacht besteed aan SIIC_ω, aangezien dit criterium voortkomt uit een schatter van σ^2, die minder aantrekkelijk is dan $\tilde{\sigma}_1^2$ of $\hat{\sigma}_1^2$.

4. Mate van zuinigheid

Elk van de in §3 afgeleide criteria brengt een afweging tot stand tussen de omvang van het model k_1 en de omvang van de fouten $y'M_1y$. Aan de marginale substitutie verhouding kan afgelezen worden welke marginale afname van $y'M_1y$ vereist is om een marginale toename van k_1 te compenseren (hierbij wordt k_1 beschouwd als continue variabele). Hoe hoger de vereiste afname van $y'M_1y$ des te zuiniger is het criterium met het toevoegen van parameters en daarom kan de marginale substitutie verhouding geïnterpreteerd worden als de *mate van zuinigheid* van het criterium.

Aan (20), (21) en (23) is te zien dat binnen de aanpak van §3 de zuinigheid wordt bepaald door ω en de schatter van σ^2. Hoe kleiner ω hoe zuiniger het criterium, dus de generalisatie van $\omega = 1$ tot $0 < \omega \leq 1$ opent de weg naar zuiniger criteria. Een hogere schatting van σ^2 leidt ook tot meer zuinigheid. De hoogte van de alternatieve schatters van σ^2 wordt verduidelijkt door de onderstaande relaties, welke mede-gebaseerd zijn op (9) en $\gamma_1 \geq 0$:

$$E[\tilde{\sigma}^2] < E[\hat{\sigma}^2] = \sigma^2 \leq E[\hat{\sigma}_1^2], \tag{25}$$

$$E[\tilde{\sigma}^2] \leq E[\tilde{\sigma}_1^2] < E[\hat{\sigma}_1^2], \tag{26}$$

waarbij m.b.t. $\tilde{\sigma}^2$ en $\hat{\sigma}^2$ verondersteld wordt dat $\text{rang}(X) = k < n$. Uit (25) blijkt dat gemiddeld genomen $\tilde{\sigma}^2$ een *lage* en $\hat{\sigma}_1^2$ een *hoge* schatting van σ^2 geeft, terwijl $\hat{\sigma}^2$ gemiddeld goed zit. Daarnaast laat (26) zien dat $\tilde{\sigma}_1^2$ gemiddeld tussen de lage en de hoge schatting uitkomt. De eerste relatie in (26) impliceert dat SIMEP_ω gemiddeld minstens zo zuinig is als SDMEP_ω bij eenzelfde keuze van ω.

In de praktijk zijn de verschillen tussen de alternatieve schatters van σ^2 meestal niet zo groot in verhouding tot het belang van de keuze van ω (als men b.v. ω verlaagt van 1 naar 0.35 dan ondergaat $1+1/\omega$ een verhoging van 93%, n.l. van 2 naar 3.86). Het feit dat $\tilde{\sigma}_1^2$ gemiddeld tussen de lage en de hoge schatter in zit en de doorgaans relatief geringe verschillen tussen de alternatieve schatters van σ^2 maakt dat men de praktische aspecten zou kunnen laten prevaleren en dus SIMEP_ω prefereren boven SDMEP_ω. De kwaliteit van beide beslissingsregels wordt nog verder bestudeerd aan de hand van een Monte Carlo experiment in §7.

5. Geïmpliceerde F-toets

In deze paragraaf worden SDMEP_ω en SIMEP_ω omgezet in kritieke waarden bij F-toetsen van geneste modellen. Veronderstel dat X_a, $X_b \in \{X_1\}$, X_a is genest in X_b (d.w.z. de kolommen van X_a vormen een deelverzameling van de kolommen X_b) en $k_a < k_b$. Dan is de F-waarde gedefinieerd als

$$F = \frac{(y'M_a y - y'M_b y)/(k_b - k_a)}{y'M_b y/(n-k_b)}. \qquad (27)$$

Past men $SDMEP_\omega$ toe met $X = X_b$, dan wordt X_b geprefereerd boven X_a, indien

$$y'M_a y + (1+1/\omega)\tilde{\sigma}_b^2 k_a > y'M_b y + (1+1/\omega)\tilde{\sigma}_b^2 k_b$$

en dit is equivalent met het *kritieke gebied*

$$F > (1+\frac{1}{\omega})\frac{n-k_b}{n-k_b+2} = FMEP_\omega^1, \qquad (28)$$

waarin $FMEP_\omega^1$ impliciet is gedefinieerd (bij gebruik van $\hat{\sigma}^2$ i.p.v. $\tilde{\sigma}^2$ is de kritieke F-waarde $1+1/\omega$). Past men $SIMEP_\omega$ toe, dan wordt X_b geprefereerd boven X_a, indien

$$\frac{y'M_a y}{(n-k_a+2)^{1+1/\omega}} > \frac{y'M_b y}{(n-k_b+2)^{1+1/\omega}}$$

en dit is equivalent met

$$F > \left[\left(\frac{n-k_a+2}{n-k_b+2}\right)^{1+1/\omega} - 1\right]\frac{n-k_b}{k_b-k_a} = FMEP_\omega^2. \qquad (29)$$

Uit (28) en (29) valt via Taylor-ontwikkelingen het volgende af te leiden (zie Van Casteren 1994 blz. 49-50):

$$FMEP_\omega^1 = 1+1/\omega + O(n^{-1}),$$
$$FMEP_\omega^2 = 1+1/\omega + O(n^{-1}).$$

Voor voldoende grote n kunnen beide kritieke waarden dus benaderd kunnen worden door $1+1/\omega$. Ter illustratie geeft tabel 3 voor enkele gevallen $FMEP_\omega^1$ en $FMEP_\omega^2$. In deze tabel worden ook de bijbehorende onbetrouwbaarheidsdrempels gegeven, die onder normaliteit van ε_i berekend kunnen worden (via de F-verdeling met k_b-k_a en $n-k_b$ vrijheidsgraden):

$$\alpha MEP_\omega^1 = \Pr[F > FMEP_\omega^1 \mid X_a \text{ is waar}],$$
$$\alpha MEP_\omega^2 = \Pr[F > FMEP_\omega^2 \mid X_a \text{ is waar}].$$

Tabel 3. Kritieke F-waarden $FMEP_\omega^1$ en $FMEP_\omega^2$ en onbetrouwbaarheidsdrempels αMEP_ω^1 en αMEP_ω^2 als functie van ω, $k_b - k_a$ en $n - k_b$.

ω		1		0.5		0.33		0.25	
$k_b - k_a$		1	5	1	5	1	5	1	5
	$n - k_b$								
	10	1.67	1.67	2.50	2.50	3.33	3.33	4.17	4.17
$FMEP_\omega^1$	50	1.92	1.92	2.88	2.88	3.85	3.85	4.81	4.81
	∞	2.00	2.00	3.00	3.00	4.00	4.00	5.00	5.00
	10	1.74	2.01	2.71	3.69	3.77	6.06	4.92	9.41
$FMEP_\omega^2$	50	1.94	2.02	2.94	3.17	3.96	4.44	5.00	5.83
	∞	2.00	2.00	3.00	3.00	4.00	4.00	5.00	5.00
	10	.23	.23	.14	.10	.10	.050	.069	.026
αMEP_ω^1	50	.17	.11	.10	.023	.055	.0050	.033	.0011
	∞	.16	.075	.083	.010	.046	.0012	.025	.00014
	10	.22	.16	.13	.038	.081	.0078	.051	.0015
αMEP_ω^2	50	.17	.092	.093	.015	.052	.0020	.030	.00026
	∞	.16	.075	.083	.010	.046	.0012	.025	.00014

Bij het toepassen van de bovenstaande kritieke F-waarden dient men goed te beseffen dat het gebruik van $FMEP_\omega^2$ exact overeenkomt met minimalisering van $SIMEP_\omega$ en dus eenduidig tot een optimaal model leidt, terwijl het gebruik van $FMEP_\omega^1$ niet overeenkomt met minimalisering van $SDMEP_\omega$. Dit laatste komt doordat $FMEP_\omega^1$ is afgeleid van $SDMEP_\omega$ onder de keuze $X = X_b$, zodat de waarde van $\tilde{\sigma}^2$ bij iedere toets weer anders kan zijn. Het gevolg is dat een reeks achtereenvolgende model-vergelijkingen m.b.v. $FMEP_\omega^1$ niet tot een eenduidige model-voorkeur hoeft te leiden. Om die reden wordt gebruik van $FMEP_\omega^1$ niet aangeraden.

6. Asymptotiek

In deze paragraaf wordt ingegaan op de optimale keuze van ω voor $n \to \infty$. In de literatuur vindt men twee soorten asymptotische analyse, die allebei besproken zullen worden. Daarna zal de relevantie van beide soorten resultaten aan de orde worden gesteld. Om het

verschil duidelijk te maken wordt X_0 gedefinieerd als de kolommen van X waarvan de regressiecoëfficiënten ongelijk aan nul zijn (relevante regressoren), en k_0 als het aantal kolommen in X_0. Terwijl X een waar model voorstelt, is X_0 dus het kleinst mogelijke ware model (verondersteld dat er verder geen parameter-restricties zijn) en men kan X_0 daarom het *minimaal ware model* noemen. Een waar model kan dus irrelevante regressoren (met $\beta_j=0$) bevatten, maar het minimaal ware model X_0 kan dit niet. Er ontstaan nu twee verschillende soorten asymptotische analyse door één van de volgende assumpties aan te nemen:

(1) X_0 is eindig en vast (dus k_0 is eindig en onafhankelijk van n);

(2) k_0 is oneindig of nadert naar oneindig voor $n\to\infty$.

Eerst wordt verondersteld dat k_0 eindig en vast is, en bovendien dat $X_0 \in \{X_1\}$. Dit laatste betekent dat $\text{rang}(X_0) = k_0 < n$. Volgens (7) geldt

$$\rho_P = \sigma^2 + \sigma^2 k_1/n \quad \text{als } \beta_2 = 0 \text{ (waar model)}, \tag{30}$$

dus van alle ware modellen is de beste die met het kleinste aantal regressoren (dit geldt voor alle n). Dit impliceert dat X_0 het beste model is onder alle ware modellen in $\{X_1\}$ (dezelfde conclusie volgt met ρ_{ME}). Nu rest nog te bekijken of X_0 ook beter is (asymptotisch) dan alle niet-ware modellen in $\{X_1\}$. Onder de veronderstelling dat $n^{-1}X'X$ voor $n\to\infty$ convergeert naar een symmetrische positief definiete matrix, kan worden aangetoond dat tevens $n^{-1}X_2'M_1X_2$ voor iedere $X_1 \in \{X_1\}$ convergeert naar een symmetrische positief definiete matrix (zie Van Casteren 1994 blz. 99). Bijgevolg convergeert ook γ_1/n, zeg $c_1 = \lim_{n\to\infty} \gamma_1/n$. Nu geldt

$$c_1 = 0 \quad \text{als } \beta_2 = 0 \text{ (waar model)}, \tag{31a}$$

$$c_1 > 0 \quad \text{als } \beta_2 \neq 0 \text{ (niet-waar model)}, \tag{31b}$$

en met (7) volgt

$$\lim_{n\to\infty} \rho_P = \sigma^2 + c_1, \tag{32}$$

hetgeen groter is dan σ^2 als $\beta_2 \neq 0$ en gelijk aan σ^2 als $\beta_2 = 0$. Dit betekent dat X_0 asymptotisch een lagere waarde van ρ_P (en ρ_{ME}) heeft dan ieder niet-waar model in $\{X_1\}$. Uiteindelijk is de conclusie dus dat het minimaal ware model X_0 asymptotisch de voorkeur verdient.

Gegeven de asymptotische suprematie van X_0, kan een selectie criterium *asymptotisch optimaal* genoemd worden, indien dit criterium voor $n\to\infty$ zorgt voor de selectie van X_0. Men kan dit op verschillende manieren opvatten en men spreekt in dit

verband van zwakke of sterke dimensie consistentie. Een criterium heet *zwak dimensie consistent*, indien de kans om daarmee X_0 te selecteren nadert naar één voor $n\to\infty$, aangenomen dat $X_0 \in \{X_1\}$. Onder normaliteit van ε_i kan worden aangetoond dat de kans op het selecteren van een niet-minimaal waar model ("overfitting") naar nul nadert bij gebruik van $SDMEP_\omega$ of $SIMEP_\omega$, mits

$$\lim_{n\to\infty} \omega_n = 0 \tag{33}$$

(hierbij wordt aangenomen dat ω afhankelijk van n gekozen wordt) en tevens kan worden aangetoond dat er situaties zijn waarin deze kans niet naar nul gaat, indien $0 < \lim_{n\to\infty} \omega_n \leq 1$ (zie Van Casteren 1994 blz. 137-138, 147-148). Voorts kan onder de aanname dat ε_i normaal verdeeld is en dat $n^{-1}X'X$ convergeert naar een symmetrische positief definiete matrix worden aangetoond dat de kans op het selecteren van een niet-waar model ("underfitting") nadert naar nul bij gebruik van $SDMEP_\omega$ of $SIMEP_\omega$, mits

$$\lim_{n\to\infty} \frac{1}{n\omega_n} = 0 \tag{34}$$

(zie Van Casteren 1994 blz. 137-138, 147-148). Hieruit blijkt dat $SDMEP_\omega$ en $SIMEP_\omega$ zwak dimensie consistent zijn en dus asymptotisch efficiënt, indien ω afhankelijk van n gekozen wordt zodanig dat voldaan wordt aan (33) en (34). Over het begrip *sterke dimensie consistentie* zij hier slechts vermeld dat dit de extra eis oplegt, dat ω_n met een bepaalde snelheid naar nul moet gaan: i.p.v. (33) komt dan $\lim_{n\to\infty} \omega_n \ln n = 0$ of $\lim_{n\to\infty} \omega_n \ln \ln n = 0$ (zie Pötscher 1989, Van Casteren 1994 blz. 162).

De tweede soort asymptotische analyse gaat uit van de veronderstelling dat k_0 oneindig is of nadert naar oneindig voor $n\to\infty$. Enkele relevante asymptotische resultaten zijn gegeven door Shibata (1981). Op basis daarvan kan geconcludeerd worden dat $SDMEP_\omega$ en $SIMEP_\omega$ met $\omega = 1$ onder bepaalde assumpties asymptotisch optimale selectie criteria zijn in de zin dat ρ_P asymptotisch geminimaliseerd wordt (en ρ_{ME} dus ook). De gebruikte assumpties zijn o.a. $k_0 = \infty$, ε_i is normaal verdeeld, $\{X_1\}$ breidt zich alsmaar uit bij toenemende n,

$$\lim_{n\to\infty} \max_{X_i\in\{X_i\}} k_i/n = 0, \tag{35}$$

$$\lim_{n\to\infty} \sum_{X_i\in\{X_i\}} \lambda^{\sigma^2 k_i+\gamma_i} = 0 \quad \text{voor elke } 0 < \lambda < 1. \tag{36}$$

Volgens (35) moet de omvang van het grootste model in de keuze-verzameling minder snel toenemen dan n en volgens (36) moet γ_i voor iedere $X_i \in \{X_i\}$ voldoende snel naar oneindig gaan in verhouding tot de snelheid waarmee $\{X_i\}$ zich uitbreidt.

De twee soorten asymptotische analyse leiden klaarblijkelijk tot twee

onverzoenbare condities, n.l. $\omega_n \to 0$ tegenover $\omega_n \to 1$. Deze schijnbare tegenstelling kan worden verklaard m.b.v. de optimale krimp-parameter, afgeleid in §3.1. Bovendien kan daarmee de relevantie van beide asymptotische analyses worden verduidelijkt. In §3.1 is voor een willekeurig paar modellen X_a, $X_b \in \{X_i\}$ de optimale waarde ω^* afgeleid, zie (12). Ieder paar modellen in $\{X_i\}$ kent een afzonderlijke waarde van ω^* en de overall-optimale waarde van ω, over *alle* mogelijke paren in $\{X_i\}$, wordt voornamelijk bepaald door de paren modellen met relatief lage waarden van ρ_P resp. ρ_{ME} (zie ook §3.1). Wanneer n toeneemt, dan verandert de overall-optimale waarde van ω dus ten eerste doordat voor ieder vast modellen-paar ω^* verandert en ten tweede doordat bepaalde paarsgewijze modelvergelijkingen belangrijker worden en andere juist onbelangrijker worden voor de overall-optimale keuze van ω.

Om het resultaat van de eerste soort asymptotische analyse te verklaren wordt nu aangenomen dat k_0 eindig en vast is en dat $n^{-1}X'X$ voor $n\to\infty$ convergeert naar een symmetrische positief definiete matrix. Dan is er volgens (31) en (32) asymptotisch een duidelijker verschil tussen X_0 en niet-ware modellen dan tussen ware modellen onderling. Het is daarom relatief gemakkelijk om te voorkomen dat asymptotisch een niet-waar model geselecteerd wordt (en dit kan gerealiseerd worden door conditie (34) in acht te nemen). Belangrijker voor de keuze van ω zijn de vergelijkingen tussen ware modellen onderling, waarbij steeds geldt $\gamma_a = \gamma_b = 0$, dus $\delta = 0$ en volgens (12) $\omega^* = 0$. Hiermee is verklaard waarom bij dit type analyse de asymptotisch optimale waarde $\omega = 0$ verkregen wordt.

Neem vervolgens aan dat $k_0 = \infty$ of $k_0 \to \infty$ voor $n\to\infty$, dat $\{X_i\}$ zich alsmaar uitbreidt bij toenemende n en dat $\{X_i\}$ louter onware modellen bevat. Dan is juist uitsluitend het vergelijken van onware modellen van belang voor de (asymptotisch) optimale keuze van ω. Nu zou binnen de analyse, waarbij k_0 eindig en vast is en $n^{-1}X'X$ convergeert naar een symmetrische positief definiete matrix, (31b) van toepassing zijn, zodat bij het vergelijken van twee vaste geneste onware modellen X_a en X_b met $c_a > c_b$ zou gelden $\delta/n \to c_a - c_b > 0$ en dus $\delta \to \infty$, hetgeen via (15) zou resulteren in $\omega^*_{NN} \to 1$ (onder normaliteit van ε_i). Echter, wat de belangrijkste paren modellen zijn is afhankelijk van n, zodat niet direct duidelijk is wat er precies gebeurt met δ (≥ 0), $k_b - k_a$ (≥ 1) en ω^*_{NN} voor de belangrijkste paren geneste modellen. Niettemin kan men uit (15) afleiden dat voor de belangrijkste geneste paren $\omega^*_{NN} \to 1$ voor $n\to\infty$, indien voor deze paren geldt $\delta^2/[\sigma^4(k_b - k_a)] \to \infty$. De laatste eis is verwant aan (36), zodat hiermee de door Shibata (1981) verkregen asymptotisch optimale waarde $\omega = 1$ verduidelijkt is. Het dient echter te worden opgemerkt dat $\omega^*_{NN} \to 0$, indien $\delta^2/[\sigma^4(k_b - k_a)] \to 0$, hetgeen suggereert dat binnen dit type asymptotische analyse ook $\omega = 0$ als asymptotisch optimale waarde verkregen kan worden door uit te gaan van andere veronderstellingen. In feite toont de cruciale rol van $\delta^2/[\sigma^4(k_b - k_a)]$ voor het gedrag van ω^*_{NN} op overtuigende wijze dat de asymptotisch optimale ω bij dit type analyse afhangt van de wijze waarop de belangrijkste modellen en hun waarden van γ_i en k_i zich ontwikkelen bij toenemende n.

In de praktijk zijn de waarden van γ_1 voor $X_1 \in \{X_1\}$ onbekend en is tevens onbekend welke paren modellen het belangrijkste zijn, waardoor men in het ongewisse blijft over de meest relevante waarden van $\delta^2/[\sigma^4(k_b-k_a)]$ en ω^*_{NN}. Een redelijk, onbevooroordeeld uitgangspunt is dan, zoals reeds aangegeven in §3.1, dat bij de belangrijkste geneste paren modellen de beide waarden van ρ_P (resp. ρ_{ME}) ongeveer even groot zijn, hetgeen betekent dat $\delta = \sigma^2(k_b-k_a)$, oftewel $\delta^2/[\sigma^4(k_b-k_a)] = k_b-k_a$, waardoor ω^*_{NN} reduceert tot ω^*_{NNE} in (16). Het moge duidelijk zijn, dat dit uitgangspunt een *tussenpositie* inneemt tussen de bovengenoemde mogelijkheden waarin ω^*_{NN} nadert naar 1 of 0 voor $n\to\infty$.

Er kan nog worden opgemerkt dat met de ingevoerde eis $\delta = \sigma^2(k_b-k_a) > 0$ impliciet verondersteld wordt dat $\gamma_a > 0$. Niet-ware modellen blijven dus relevant voor alle mogelijke n. Hiermee is ook impliciet een keuze gemaakt voor de tweede soort asymptotische analyse, waarbij $k_0 = \infty$ of $k_0 \to \infty$ voor $n\to\infty$. Dit is een overtuigde keuze, omdat in de praktijk geen enkel model een perfecte representatie van de werkelijkheid is. Dit wil echter niet zeggen dat de andere soort asymptotische analyse (met vaste k_0) irrelevant zou zijn. Een voorbeeld moge dit verduidelijken. Stel dat $k_0 = \infty$ en laat $\beta_j \to 0$ voor $n\to\infty$ voor alle $j = 11,12,\ldots$, dan is het model met uitsluitend de eerste 10 regressoren steeds een onwaar model, totdat de limiet is bereikt. Dit illustreert dat er een continue overgang is van de assumptie $k_0 = \infty$ naar de assumptie k_0 is eindig en vast. Voor eindige n zal de optimale waarde van ω dan ook weinig verschillen tussen het geval waarin $\beta_{11},\beta_{12},\ldots$ bijna hun limiet hebben bereikt en het geval waarin zij die geheel hebben bereikt, zodat ook de ontwikkeling van de optimale ω bij toenemende, eindige n weinig zal verschillen. Dit verklaart waarom via beide typen asymptotische analyses de optimale waarde $\omega = 0$ verkregen kan worden. Concluderend, zijn beide typen asymptotische analyses relevant, ook al is een eindig waar model onbestaanbaar. Gezien het bovenstaande zou men eigenlijk de *effectieve* k_0 moeten definiëren als het aantal regressoren met niet-verwaarloosbare coëfficiënten (gegeven de bereikbare precisie bij de actuele waarde van n). Het *effectieve* minimaal ware model (*effectieve* X_0) behoeft dan slechts een goede en geen exacte representatie van de werkelijkheid te zijn. Een illustratie hiervan wordt gegeven in §7.

7. Monte Carlo experiment

De werking van $SDMEP_\omega$ en $SIMEP_\omega$ zal nu geïllustreerd worden middels een Monte Carlo experiment. Daarbij zullen de resultaten voor wat betreft de optimale waarde van ω worden vergeleken met de analyses van §3.1 (ω^*, ω^*_{NN}, ω^*_{NNE}) en §6 (asymptotiek). In het experiment worden artificiële data gegenereerd door het volgende proces:

$$y_i = \sum_{j=1}^{20} \beta_j \, z_{i-j+1} + \varepsilon_i, \tag{37}$$

waarin

$$z_i = 1.6 \, z_{i-1} - 0.64 \, z_{i-2} + \eta_i \tag{38}$$

met ε_i en η_i voor i = 1,2,3,... onafhankelijke trekkingen uit de standaard normale verdeling. Merk op dat de regressoren in (37) als constanten beschouwd kunnen worden door op de uitkomsten van z_i te conditioneren. Wanneer waarden voor $\beta_1,\ldots,\beta_{20}$ gekozen worden, is het data genererende proces volledig gespecificeerd. Tabel 4 geeft zes mogelijkheden (A,...,F) die gebruikt zullen worden. Om aansluiting te vinden bij de bestaande literatuur zijn de processen A, B en C, inclusief (37) en (38), gekozen in navolging van Geweke en Meese (1981) en Teräsvirta en Mellin (1986). De processen D, E en F zijn toegevoegd om *beide* soorten asymptotische analyse te illustreren (daarbij zijn de coëfficiënten in vaste verhoudingen genomen, n.l. D:E = 1:5 en D:F = 1:10).

Tabel 4. Data genererende processen A,...,F: waarden van $\beta_1,\ldots,\beta_{20}$ in (37).

j	A	B	C	D	E	F
1	1	0.5	1	1	5	10
2		0.8	1	0.6	3	6
3		0.8	1	0.3	1.5	3
4		0.8	1	0.1	0.5	1
5		0.8	0.9	0.06	0.3	0.6
6		0.8	0.8	0.03	0.15	0.3
7		0.8	0.7	0.01	0.05	0.1
8		0.5	0.6	0.006	0.03	0.06
9			0.5	0.003	0.015	0.03
10			0.4	0.001	0.005	0.01
11			0.3	0.001	0.005	0.01
12			0.2	0.001	0.005	0.01
13			0.1	0.001	0.005	0.01
14,...,20				0.001	0.005	0.01
k_0	1	8	13	20	20	20

In alle gevallen wordt genomen: k = 20, $x_i = (z_i,\ldots,z_{i-k+1})'$ en $\{X_1\}$ = {de eerste

k_1 kolommen van X}. Het selectie probleem reduceert dus tot het probleem de orde van de verdeelde vertraging in (37) te bepalen. Gegeven elk van de zes processen wordt achtereenvolgens n = 50, 100, 200 genomen, zodat in totaal 18 cases resulteren. In elke case worden 1000 replicaties van een complete data verzameling gegenereerd, bestaande uit n waargenomen perioden (i = 1,...,n) en daaropvolgend 100 voorspelperioden (p = 1,...,100), welke worden gebruikt om de selectie criteria te evalueren.

Het evalueren van selectie criteria geschiedt als volgt. In elke replicatie wordt met ieder criterium een model geselecteerd, waarmee twee verlieswaarden worden berekend:

$$\ell_{ME} = \frac{1}{n} \sum_{i=1}^{n} (x_i'\beta - x_i'\hat{\beta}_{(1)})^2, \tag{39}$$

$$\ell_p = 1 + \frac{1}{100} \sum_{p=1}^{100} (x_p'\beta - x_p'\hat{\beta}_{(1)})^2. \tag{40}$$

Vervolgens worden beide verlieswaarden gemiddeld over de 1000 replicaties, hetgeen leidt tot r_{ME} en r_P, die als meetbare versies van ρ_{ME} resp. ρ_P beschouwd kunnen worden. Een eventueel verschil tussen r_{ME} en $r_p - 1$ kan zijn oorzaak vinden in het tweemaal gebruiken van dezelfde data (zie §3) of in onjuistheid van (5). In het huidige experiment is (5) bij benadering correct.

Om de optimale waarde van ω, waarvoor r_{ME} resp. r_P geminimaliseerd wordt, te kunnen bepalen, worden $SDMEP_\omega$ en $SIMEP_\omega$ toegepast met 10 mogelijke waarden: ω = 0.1, 0.2,...,1. Tabel 5 geeft de optimale waarde van ω in elke case bij evaluatie door r_{ME} alsmede de bijbehorende gemiddelde omvang van het geselecteerde model en tabel 6 is soortgelijk voor r_P.

Tabel 5. De optimale ω bij evaluatie door r_{ME} en de bijbehorende gemiddelde geselecteerde k_1.

proces		optimale ω						gemiddelde geselecteerde k_1					
		A	B	C	D	E	F	A	B	C	D	E	F
$SDMEP_\omega$	n = 50	.1	.3	.5	.4	.4	.6	1.0	8.1	12.1	4.1	6.6	9.8
	n = 100	.1	.2	.4	.5	.5	1	1.0	8.0	12.2	4.8	7.9	13.4
	n = 200	.1	.1	.2	.5	.9	1	1.0	8.0	12.1	5.2	11.1	15.2
$SIMEP_\omega$	n = 50	.1	.3	.5	.5	.4	.7	1.0	8.0	12.0	4.1	6.4	9.7
	n = 100	.1	.2	.4	.5	.6	1	1.0	8.0	12.2	4.7	8.1	13.3
	n = 200	.1	.1	.3	.5	.8	1	1.0	8.0	12.3	5.2	10.7	15.2

Tabel 6. De optimale ω bij evaluatie door r_P en de bijbehorende gemiddelde geselecteerde k_1.

proces		optimale ω A	B	C	D	E	F	gemiddelde geselecteerde k_1 A	B	C	D	E	F
$SDMEP_\omega$	n = 50	.1	.3	.4	.4	.4	.6	1.0	8.1	11.9	4.1	6.6	9.8
	n = 100	.1	.2	.4	.5	.5	1	1.0	8.0	12.2	4.8	7.9	13.4
	n = 200	.1	.1	.2	.5	.9	1	1.0	8.0	12.1	5.2	11.1	15.2
$SIMEP_\omega$	n = 50	.1	.3	.4	.5	.4	.7	1.0	8.0	11.9	4.1	6.4	9.7
	n = 100	.1	.2	.4	.5	.6	1	1.0	8.0	12.2	4.7	8.1	13.3
	n = 200	.1	.1	.3	.5	1	1	1.0	8.0	12.3	5.2	11.4	15.2

Vergelijkt men de optimale waarden van ω in tabel 5 met die in tabel 6, dan ziet men nauwelijks verschil, en dit betekent dat men kan volstaan met één waarde voor ω te kiezen voor zowel het minimaliseren van ρ_{ME} als voor het minimaliseren van ρ_P, zoals reeds geopperd in §2.

Vergelijkt men vervolgens de optimale waarden van ω voor $SDMEP_\omega$ met die voor $SIMEP_\omega$, dan ziet men ook weinig verschil, zij het dat de optimale ω voor $SDMEP_\omega$ op slechts één uitzondering na altijd gelijk of kleiner is dan die voor $SIMEP_\omega$. Dit strookt met de observatie in §4 dat $SIMEP_\omega$ bij gelijke ω gemiddeld minstens zo zuinig is als $SDMEP_\omega$, zodat $SDMEP_\omega$ voor het bereiken van een even grote mate van zuinigheid een gelijke of kleinere waarde van ω nodig heeft.

Vergelijkt men tenslotte de zes processen, dan ziet men totaal verschillende patronen in de optimale waarden van ω. Bij processen A en B ziet men lage optimale waarden van ω (≤ 0.3) en een dalende tendens (tot de laagst beschouwde waarde 0.1) en hetzelfde patroon is te zien bij proces C, zij het iets minder uitgesproken. Bij processen E en F, en in wat mindere mate bij proces D, ziet men juist relatief hoge optimale waarden van ω (≥ 0.4) en een stijgende tendens. Hier zijn duidelijk de twee typen asymptotiek te herkennen met de twee asymptotisch optimale waarden $\omega = 0$ en $\omega = 1$ (zie §6). De verklaring hiervoor is dat bij processen A en B de gemiddelde geselecteerde k_1 (bij optimale keuze van ω) reeds gelijk is aan k_0 (bij proces C bijna gelijk aan k_0), zodat de situatie goed gekarakteriseerd wordt door de assumptie dat k_0 eindig en vast is, terwijl bij processen D, E en F de gemiddelde geselecteerde k_1 nog ver beneden k_0 ligt, waardoor de situatie beter gekarakteriseerd wordt door de assumptie dat k_0 oneindig is of naar oneindig gaat voor $n \to \infty$. Het naast elkaar bestaan van twee tegengestelde patronen bevestigt het nut van beide typen asymptotische analyse. De resultaten tonen tevens de toepasselijkheid van de optimale waarden ω^* en ω^*_{NN} (zie §3.1): als de gemiddelde geselecteerde k_1 dicht in de buurt van k_0 komt (processen A, B en C), dan is het geval $\delta = 0$ belangrijk (kleine

optimale ω) en wordt steeds belangrijker naarmate n toeneemt (dalende optimale ω), maar als de gemiddelde geselecteerde k_1 nog ver onder k_0 blijft (processen D, E en F), dan kunnen juist relatief grote waarden van δ^2 belangrijk zijn (grote optimale ω bij E en F) en kunnen deze bovendien snel toenemen bij stijgende n (stijgende optimale ω bij E en F).

Om het onderscheid tussen beide typen patronen nog verder te verduidelijken is het bovengenoemde experiment herhaald met vervanging van processen A, B en C door processen A', B' en C', die uit A, B en C verkregen worden door op de open plekken in tabel 4 steeds de waarde $\beta_j = 10^{-6}$ in te vullen. Nu resulteren voor processen A', B' en C' dezelfde optimale waarden van ω als voor processen A, B en C, zoals gerapporteerd in tabellen 5 en 6. Dit illustreert dat de optimale ω een continue functie is van β (zoals ook in de uitdrukkingen voor ω^* en ω^*_{NN}), waardoor de overgang tussen het wel of niet waar zijn van een bepaald model een continue is met slechts een geleidelijk effect op de optimale waarde van ω. Bovendien toont het feit dat processen A', B' en C' in de analyse van de optimale ω goed gerepresenteerd kunnen worden door processen A, B en C dat een minimaal waar model in de analyse gerepresenteerd kan worden door een effectief minimaal waar model, bestaande uit het minimaal aantal benodigde regressoren om een bijna waar model te vormen.

De resultaten tonen dat er een enorm gevaar schuilt in het gebruik van Monte Carlo experimenten voor het vergelijken van model selectie criteria, zoals er zo vele in de literatuur te vinden zijn, b.v. Akaike (1979), Atkinson (1980), Bhansali en Downham (1977), Geweke en Meese (1981), Hannan en Quinn (1979), Teräsvirta en Mellin (1986). Men kan de uitkomst immers volstrekt manipuleren door de keuze voor een bepaald type data genererend proces. Door uit te gaan van een proces als A of B (effectieve k_0 dichtbij de gemiddeld geselecteerde k_1; kleine waarden van δ^2 zijn dominant) kan men een "bewijs" creëren dat zuiniger criteria het beter doen dan minder zuinige, terwijl men met een proces als E of F (effectieve k_0 veel groter dan de gemiddelde geselecteerde k_1; grote waarden van δ^2 zijn dominant) een "bewijs" kan leveren dat minder zuinige criteria beter zijn. Derhalve dient men aan een goed Monte Carlo experiment de eis te stellen dat beide typen processen, corresponderend met beide typen asymptotiek, vertegenwoordigd zijn. In het hier besproken experiment is aan deze eis voldaan (processen A en B tegenover E en F). Additioneel zou men kunnen eisen dat het tussenliggende geval (processen als C en D) vertegenwoordigd is.

In het bovenstaande experiment stuit men klaarblijkelijk op de zelfde problemen als bij de theoretische analyses van §3.1 en §6, n.l. de optimale waarde van ω is onbekend en hangt af van het data genererende proces en van het aantal waarnemingen, waardoor er niet één waarde van ω te noemen is, die in alle gevallen superieur is. Om op basis van dit experiment toch één geschikte waarde van ω te kunnen vinden wordt nu overgegaan tot het wegen van de 18 cases. Er was reeds geconstateerd dat in dit experiment elk type proces en het tussenliggende geval ieder met twee processen vertegenwoordigd is, en daarom wordt ervoor gekozen elk van de processen A,...,F hetzelfde gewicht toe te kennen, daarbij toegevend dat dit een arbitraire keuze is. Evenzo wordt de arbitraire

keuze gemaakt de drie waarden n = 50, 100, 200 even zwaar te laten wegen. Gezien de arbitraire weging en de arbitraire keuze van de zes data genererende processen wordt uitdrukkelijk niet gepretendeerd het ultieme antwoord te kunnen geven op de vraag welke waarde van ω in de praktijk over het algemeen het beste is.

Om de voorspelprestaties op een geschikte manier te kunnen middelen over de 18 cases worden eerst de evaluatie-waarden r_{ME} en r_P genormaliseerd. Stel dat men perfecte kennis heeft over ℓ_{ME} en ℓ_P voor ieder alternatief model, zoals bij dit gecontroleerde experiment het geval is, dan kan men steeds perfecte model selectie toepassen door ℓ_{ME} resp. ℓ_P te minimaliseren in plaats van $SDMEP_\omega$ of $SIMEP_\omega$. Schrijft men r_{ME}^{min} resp. r_P^{min} voor het gemiddelde van deze minima over alle 1000 replicaties in een bepaalde case, dan worden bij toepassing van een bepaald criterium in deze case de genormeerde waarden van r_{ME} en $r_P - 1$ gegeven door de indices

$$I_{ME} = 100\, r_{ME} / r_{ME}^{min}, \tag{41}$$

$$I_P = 100(r_P - 1) / (r_P^{min} - 1), \tag{42}$$

die beide minstens 100 moeten zijn. Hierbij is $r_P - 1$ genormaliseerd i.p.v. r_P (d.w.z. $\sigma^2 = 1$ wordt er eerst uit gehaald) omdat het niet zinvol is te bekijken hoe goed de foutencomponent ε_p voorspeld wordt. Een bijkomend voordeel is dat I_{ME} en I_P dan beter vergelijkbaar zijn. Het verschil tussen I_{ME} of I_P t.o.v. de waarde 100 kan geïnterpreteerd worden als het relatieve verlies dat optreedt door niet-perfecte model selectie in de betreffende case. Het samennemen van de 18 cases geschiedt nu op twee manieren. Ten eerste wordt het ongewogen gemiddelde van I_{ME} resp. I_P bepaald over de 18 cases en ten tweede wordt de maximum waarde van I_{ME} resp. I_P bepaald over de 18 cases. Gebruik van de maximum waarde veronderstelt dat vooral extreem hoge uitkomsten van I_{ME} resp. I_P vermeden moeten worden (risico mijdend gedrag).

De eindresultaten staan in tabel 7. Daarin zijn de uitkomsten vermeld voor $SDMEP_\omega$,$SIMEP_\omega$ en $SIIC_\omega$ (met $\omega = 0.1,\ldots,1$) en bovendien voor SBIC. Vergelijkt men de uitkomsten voor de gemiddelde indices met die voor de maximum indices, dan is er nauwelijks verschil in de optimale waarden van ω te constateren, dus het maakt voor de keuze van ω weinig uit of men gemiddeld zo goed mogelijk uit wil komen dan wel risico mijdend wil zijn.

Volgens tabel 7 ligt de beste waarde van ω nabij 0.3 voor $SDMEP_\omega$ en in de buurt van 0.4 voor $SIMEP_\omega$. Hiermee is een drie-dubbele motivering verkregen om in het algemeen ω kleiner dan 1 te kiezen, ergens in het midden tussen 0 en 1, n.l. de analyse m.b.v. ω^*_{NNE} (§3.1), de twee tegengestelde asymptotische analyses (§6) en de uitkomst van het Monte Carlo experiment (§7).

Tabel 7. Samengevatte prestaties van selectie criteria over 18 cases. De minima zijn gemarkeerd.

criterium	ω	gemiddelde I_{ME}	maximum I_{ME}	gemiddelde I_P	maximum I_P
$SDMEP_\omega$	1.0	186.41	488.64	197.12	475.56
	0.9	175.81	427.20	185.28	406.77
	0.8	165.45	376.18	173.44	352.63
	0.7	156.64	326.35	164.14	307.70
	0.6	148.45	282.42	156.20	265.23
	0.5	140.25	230.46	148.23	220.69
	0.4	135.05	189.09	143.63	183.90
	0.3	132.96	156.94	142.47	177.79
	0.2	136.61	181.00	147.33	202.37
	0.1	161.14	256.51	175.77	285.37
$SIMEP_\omega$	1.0	176.89	396.09	185.05	368.83
	0.9	166.54	347.80	174.24	325.62
	0.8	157.57	298.63	164.83	277.46
	0.7	149.96	262.27	156.73	238.96
	0.6	142.48	222.48	149.84	207.37
	0.5	136.98	194.71	144.85	186.59
	0.4	133.10	164.92	141.35	168.88
	0.3	133.15	157.60	142.55	185.43
	0.2	139.53	190.22	150.87	213.65
	0.1	172.10	277.76	188.13	306.70
$SIIC_\omega$	1.0	196.28	548.27	209.40	536.73
	0.9	183.32	466.23	195.31	455.73
	0.8	170.10	388.74	179.98	367.37
	0.7	158.36	310.63	167.35	290.98
	0.6	147.50	249.76	155.92	235.63
	0.5	139.02	203.44	147.45	192.93
	0.4	133.73	172.20	142.44	167.87
	0.3	132.19	153.66	141.53	180.19
	0.2	137.38	185.56	148.66	207.60
	0.1	168.21	274.09	183.58	303.38
SBIC		133.46	170.26	143.02	190.97

Bij vergelijking van $SDMEP_\omega$ met $SIMEP_\omega$ valt op dat de prestaties bij de optimale keuze van ω nauwelijks afwijken, m.u.v. maximum I_p, die lichtelijk in het voordeel spreekt van $SIMEP_\omega$. Het effect van niet-optimale keuze van ω is ook iets gunstiger voor $SIMEP_\omega$. Immers, als men $SDMEP_\omega$ vergelijkt met $SIMEP_{\omega+0.1}$ voor $0.1 \leq \omega \leq 0.9$, dan blijkt dat de laatste het in veel gevallen een stuk beter doet en in het ergste geval maar een beetje slechter (en daarnaast is $SIMEP_\omega$ met $\omega = 0.1$ duidelijk beter dan $SDMEP_\omega$ met $\omega = 1$). Deze resultaten suggereren dat $SIMEP_\omega$ niet alleen handiger is dan $SDMEP_\omega$, maar ook nog effectiever en veiliger.

Vergelijkt men $SIIC_\omega$ met $SDMEP_\omega$ dan valt te constateren dat de prestaties voor optimale ω niet erg verschillen, maar dat voor niet-optimale ω de prestaties van $SIIC_\omega$ zichtbaar slechter zijn. Dit bevestigt dat $SIIC_\omega$ niet de voorkeur verdient.

Tenslotte is SBIC het meest vergelijkbaar met $SIIC_\omega$ omdat SBIC neerkomt op $SIIC_\omega$ met $1+1/\omega = \ln n$. Het enige verschil is dat ω bij $SIIC_\omega$ constant genomen is, terwijl bij SBIC de keuze van ω varieert met n volgens $\omega = 1/(\ln n - 1)$. Om het effect van dit verschil te zien vergelijke men SBIC met $SIIC_\omega$ waarbij $\omega = 0.28$, n.l. de gemiddelde waarde van $\omega = 1/(\ln n - 1)$ voor $n = 50, 100, 200$. De prestaties van $SIIC_\omega$ met $\omega = 0.28$ zijn niet direct beschikbaar, maar kunnen benaderd worden door interpolatie van de resultaten bij $\omega = 0.3$ en 0.2. Dit geeft 133.23, 160.24, 142.96 en 185.67 in dezelfde volgorde als in tabel 7 (deze benaderingen zijn aan de hoge kant vanwege convexiteit van de uitkomsten als functie van ω). De vaste keuze $\omega = 0.28$ geeft in vergelijking met SBIC dus iets betere resultaten voor maximum I_{ME} en maximum I_p en nauwelijks een verschil voor de gemiddelde I_{ME} en de gemiddelde I_p. In dit experiment pakt de vaste keuze van ω dus iets gunstiger uit dan de keuze volgens SBIC.

8. Informatie criteria afgeleid via krimp-schatting

In dit artikel is langs de weg van krimp-schatting een generalisatie gegeven van de aanpak van Amemiya (1980) voor het afleiden van selectie criteria. Een soortgelijke generalisatie kan toegepast worden op de aanpak van Akaike (1974) bij het afleiden van AIC. Akaike's aanpak is toepasbaar wanneer de modellen geschat worden via "maximum likelihood" en het gegeneraliseerde criterium is:

$$SIIC_\omega = -2 \ln(\text{maximum likelihood}) + (1+1/\omega)(\text{aantal vrije parameters}) \qquad (43)$$

("Set-Independent Information Criterion"). Dit criterium dient geminimaliseerd te worden over alle alternatieve modellen. $SIIC_\omega$ is gelijk aan AIC voor $1+1/\omega = 2$ en het is gelijk aan SBIC voor $1+1/\omega = \ln n$, waarbij n het aantal waarnemingen voorstelt. Zoals blijkt uit (24) kan $SIIC_\omega$ in het geval van univariate lineaire modellen met normaal verdeelde witte ruis storingen tevens worden afgeleid via de aanpak van §3. In het algemene geval zijn soortgelijke formules voor ω^*, ω^*_{NN} en ω^*_{NNE} af te leiden, waaruit blijkt dat ook hier

een waarde van ω ergens tussen 0 en 1 gepast is. Dit blijkt tevens uit het feit dat volgens de ene soort asymptotische analyse de waarde $\omega = 0$ asymptotisch optimaal is (zie Kohn 1983 en Pötscher 1989), terwijl volgens de andere soort $\omega = 1$ asymptotisch optimaal is (zie Akaike 1978; er is hier sprake van een minimax type optimaliteit). Ongetwijfeld zullen deze beweringen ook gestaafd kunnen worden middels een Monte Carlo experiment (daarbij dient men de alternatieve criteria te evalueren aan de hand van de Kullback-Leibler informatie maatstaf).

9. Samenvatting en conclusies

Gegeven een verzameling alternatieve univariate lineaire regressie modellen met witte ruis storingen is de vraag met behulp van welke beslissingsregel men het beste een model kan selecteren. Dit artikel geeft een analyse van toepasselijke model selectie regels, met als doel zo klein mogelijke voorspelfouten te verkrijgen en de verwachte waarden van de waarnemingen aan de afhankelijke variabele zo goed mogelijk te schatten. Daarbij wordt een kwadratische verliesfunctie gehanteerd, zie ρ_P in (3) en ρ_{ME} in (2).

Uit diverse soorten analyses blijkt dat minimalisering van $SIMEP_\omega$ met $0 < \omega \leq 1$, zie (22), een handige en relatief goede methode is om de model voorkeur te bepalen. Deze methode is equivalent met toepassing van de kritieke F-waarde $FMEP_\omega^2$, zie (29). Een moeilijkheid is evenwel de keuze van de krimp-parameter ω.

De keuze van ω is zeer bepalend voor de mate van zuinigheid van het criterium (de afweging tussen de omvang van het model en de omvang van de residuen, zie §4). De keuze $\omega = 1$ (dit heeft ongeveer dezelfde uitwerking als C_p, PC, GCV, S_p, AIC en BIC) is nogal extreem in de zin dat de mate van zuinigheid uiterst gering is. Een keuze van ω zeer dicht bij 0 is juist extreem bevorderlijk voor de zuinigheid. Daartussen ligt een heel gebied van mogelijkheden met een ruime variatie in de neiging tot zuinigheid. De keuze $\omega = 1/(\ln n - 1)$ heeft ongeveer dezelfde uitwerking als SBIC en BEC.

De vraag welke waarde van ω het beste gekozen kan worden is een lastige, maar niettemin is hierop langs drie wegen hetzelfde duidelijke antwoord verkregen. De theoretisch optimale waarden ω^*, ω^*_{NN} en ω^*_{NNE} (§3.1), de twee soorten asymptotische analyse (§6) en het uitgevoerde Monte Carlo experiment (§7) wijzen het volgende uit: indien de concurrerende modellen bijna waar zijn, dan kan men ω het beste dicht bij nul kiezen; indien deze modellen nogal strijdig zijn met de werkelijkheid, dan kan men ω het beste dicht bij één kiezen; in het meest voorkomende geval, waarin men dit onderscheid moeilijk kan maken, is het aanbevelenswaardig ω ergens in het midden tussen nul en één te kiezen. Wordt b.v. $\omega = 0.5$ gekozen, dan impliceert dit bij grote steekproeven dat bij F-toetsen de kritieke waarde ca. $1+1/\omega = 3$ moet bedragen, hetgeen bij een t-toets correspondeert met een kritieke $|t|$-waarde van ca. $\sqrt{3} = 1.732$ en een significantie niveau van ca. 8.3%. Kiest men $\omega = 0.33$, dan correspondeert dit met een kritieke $|t|$-waarde van ca. $\sqrt{4} = 2$, dus een significantieniveau van 4.6%, waarmee toch enigszins

een rechtvaardiging voor het arbitraire 5% significantieniveau bij t-toetsen verkregen is.

De hier gegeven uitbouw van Amemiya's beslissingstheoretische aanpak, met invoering van de krimp-parameter ω, is ook toepasbaar op de aanpak van Akaike en dit mondt uit in een generalisatie van AIC en SBIC (zie §8).

Literatuur

Akaike, H. (1974) "A New Look at the Statistical Model Identification", *IEEE Transactions on Automatic Control*, AC−19, 716−723.

Akaike, H. (1978) "A Bayesian Analysis of the Minimum AIC Procedure", *Annals of the Institute of Statistical Mathematics*, 30, Part A, 9−14.

Akaike, H. (1979) "A Bayesian Extension of the Minimum AIC Procedure of Autoregressive Model Fitting", *Biometrika*, 66, 237−242.

Amemiya, T. (1980) "Selection of Regressors", *International Economic Review*, 21, 331−354.

Atkinson, A.C. (1980) "A Note on the Generalized Information Criterion for Choice of a Model", *Biometrika*, 67, 413−418.

Bhansali, R.J., and D.Y. Downham (1977) "Some Properties of the Order of an Autoregressive Model Selected by a Generalization of Akaike's FPE Criterion", *Biometrika*, 64, 547−551.

Blight, B.J.N. (1971) "Some General Results on Reduced Mean Square Error Estimation", *The American Statistician*, 25, 24−25.

Breiman, L. and D. Freedman (1983) "How Many Variables Should Be Selected in a Regression Equation?", *Journal of the American Statistical Association*, 78, 131−136.

Casteren, P.H.F.M. van (1994) *Statistical Model Selection Rules*. Proefschrift, Vrije Universiteit, Amsterdam.

Chow, G.C. (1981) "A Comparison of the Information and Posterior Probability Criteria for Model Selection", *Journal of Econometrics*, 16, 21−33.

Geweke, J. and R. Meese (1981) "Estimating Regression Models of Finite But Unknown Order", *International Economic Review*, 22, 55−70.

Golub, G.H., M. Heath and G. Wahba (1979) "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter", *Technometrics*, 21, 215−223.

Hannan, E.J. and B.G. Quinn (1979) "The Determination of the Order of an Autoregression", *Journal of the Royal Statistical Society*, Series B, 41, 190−195.

Hocking, R.R. (1976) "The Analysis and Selection of Variables in Linear Regression", *Biometrics*, 32, 1−49.

Kohn, R. (1983) "Consistent Estimation of Minimal Subset Dimension", *Econometrica*, 51, 367−376.

Mallows, C.L. (1973) "Some Comments on C_P", *Technometrics*, 15, 661–675.

Miller, A.J. (1990) *Subset Selection in Regression*. London: Chapman and Hall.

Pötscher, B.M. (1989) "Model Selection under Nonstationarity: Autoregressive Models and Stochastic Linear Regression Models", *Annals of Statistics*, 17, 1257–1274.

Sawa, T. (1978) "Information Criteria for Discriminating among Alternative Regression Models, *Econometrica*, 46, 1273–1291.

Schwarz, G. (1978) "Estimating The Dimension of a Model", *Annals of Statistics*, 6, 461–464.

Shibata, R. (1981) "An Optimal Selection of Regression Variables", *Biometrika*, 68, 45–54.

Teräsvirta, T. and I. Mellin (1986) "Model Selection Criteria and Model Selection Tests in Regression Models", *Scandinavian Journal of Statistics*, 13, 159–171.

Theil, H. (1961) *Economic Forecasts and Policy*, 2nd ed. Amsterdam: North Holland.

Theil, H. (1971) *Principles of Econometrics*. New York: Wiley.

Thompson, J.R. (1968) "Some Shrinkage Techniques for Estimating the Mean", *Journal of the American Statistical Association*, 63, 113–122.

Ontvangen: 6 juni 1995
Geaccepteerd: 25 maart 1996