

Herman H. M. Visser^{*}

Samenvatting. Indien er veel (economische) informatie omtrent toekomstige ontwikkelingen van bepaalde bedrijfsprocessen gevraagd wordt, dient men te beschikken over de nodige voorspelmodellen. In dit artikel beperken we ons tot korte termijn voorspelmodellen gebaseerd op de ARIMA structuur en de Box-Jenkins methodologie. Deze voorspelmodellen worden regelmatig geactualiseerd en vervolgens getoetst op hun validiteit.

Indien het gaat om grote aantallen voorspelmodellen is het noodzakelijk gebleken te beschikken over een monitor systeem. Het monitor systeem voert een aantal toetsen uit en geeft een aantal signalen af waaruit een deskundige kan afleiden dat bepaalde voorspelmodellen mogelijk verstoord kunnen zijn. De deskundige beslist uiteindelijk of het voorspelmodel verstoord is en of er een actie volgt om het voorspelmodel te verbeteren.

* Centrale Afdeling Statistiek en Bedrijfseconometrie, Centrale Directie PTT, Postbus 570, 9700 AN Groningen, telefoon: 050 - 60 22 76.
De auteur is dank verschuldigd aan T.A.B. Snijders, J.J.H. Verwijst en K.A. Bron voor hun waardevolle opmerkingen bij een eerdere versie van dit artikel.

Inleiding

Er bestaat binnen organisaties behoefte aan (economische) informatie over toekomstige ontwikkelingen. Deze informatie dient in veel gevallen als basis voor beleidsbeslissingen, zowel voor de lange als de korte termijn. In dit kader kan men denken aan plannings- en beheersingsprocessen. De lange termijn planning is op de realisatie van de basisdoelstellingen gericht en betreft de doelgerichtheid, de effectiviteit van de ondernemingsactiviteiten. Dit duidt men wel aan met de term: strategische planning. De korte termijn planning is op de functionering van de lopende uitvoering van de bedrijfsprocessen gericht en betreft de doelmatigheid, de efficiëncy van de functionering. Dit duidt men wel aan met de term: tactische planning.

We kunnen dit met de volgende voorbeelden illustreren. We genereren een voorspelling van het verwachte aantal telefoonaansluitingen, die van belang zijn voor de planning van het aantal nieuwe telefooncentrales (lange termijn planning). De hoeveelheid bestelde post, die bijvoorbeeld in de decembermaand verwacht wordt (waaronder de kerst- en nieuwjaarspost) is mede bepalend voor de personeelsinzet, en is daardoor van invloed op de mogelijkheid van het personeel om vrije dagen op te nemen in december (korte termijn planning).

Om aan deze toekomstgerichte informatie te komen worden voorspelmodellen geconstrueerd voor de variabelen die van invloed zijn op diverse bedrijfsprocessen. Wij beperken ons in dit kader tot korte termijn voorspelmodellen.

In een reeks artikelen en een boek, hebben Box en Jenkins(1976) hun methodologie voor de constructie van voorspelmodellen uiteengezet en geïllustreerd. Deze methodologie betreft het aanpassen aan een gegeven tijdreeks van een model uit de klasse van de 'Autoregressive Integrated Moving Average' (ARIMA) modellen. Het werk van Box en Jenkins gaf een enorme impuls aan de beoefening van de tijdreeksanalyse. De details van deze methodologie zijn thans wel bekend, zodat het niet noodzakelijk is om ze hier uitgebreid uiteen te zetten. Zie voor een discussie, bijvoorbeeld de boeken van Nelson(1973), Granger en Newbold(1977), Jenkins(1979) en McLeod(1982). Recente ontwikkelingen worden nog eens beschreven in Newbold(1983).

Indien men beschikt over een groot aantal voorspelmodellen, die bijvoorbeeld maandelijks geactualiseerd moeten worden, dan dienen de voorspelmodellen telkens getoetst te worden op hun betrouwbaarheid. Dat wil zeggen dat de voorspelprestatie, de modelparameters alsmede de overige diagnostieken getoetst moeten worden op hun significantie. Voor de toetsing van een zeer groot aantal voorspelmodellen in een relatief kort tijdsbestek is het noodzakelijk een monitor systeem te construeren. Dit artikel geeft een aanzet tot een mogelijke in-

vulling van zo'n monitor systeem in het geval van korte termijn voorspelmodellen gebaseerd op de ARIMA-structuur, en de Box-Jenkins methodologie.

De opbouw van het artikel is als volgt. Allereerst wordt ingegaan op wat de auteur verstaat onder een monitor systeem. Vervolgens besteden we aandacht aan de werking van een monitor systeem en de toetsen die in het monitor systeem zijn geprogrammeerd. Tenslotte beschrijven we enkele resultaten van de toepassing van het monitor systeem op een tiental praktijkreeksen uit een PTT-omgeving.

Wat is een Monitor systeem

We definiëren een monitor systeem als een verzameling toetsen welke afzonderlijke signalen afgeven, waaruit de deskundige tijdreeksanalist kan afleiden dat het onderhavige voorspelmodel mogelijk verstoringen vertoont. Door middel van toetsen die in het monitor systeem zijn geprogrammeerd beschouwen we de tijdreeksdata, de voorspelprestatie, de geschatte parameters van het voorspelmodel, alsmede de structuur van het voorspelmodel. Wanneer een ontwikkeld voorspelmodel operationeel is, is het verstandig om de betrouwbaarheid c.q. geschiktheid van het model te beoordelen, teneinde goede voorspellingen te kunnen waarborgen. Ons monitor systeem is niet opgezet als beslissingsinstrument maar als signaleringsinstrument. Ten alle tijde neemt de deskundige de beslissing of het onderhavige voorspelmodel betrouwbaar genoeg is, of dat het de nodige aanpassingen behoeft. Er bestaat binnen dit kader geen 'automatisme'.

McLeod(1982) geeft twee belangrijke redenen waarom een voorspelmodel als representatie van een onderliggende tijdreeks niet langer geschikt kan zijn.

- De tijdreeks kan lokaal verstoord zijn, maar dit hoeft niet te betekenen dat er op de lange termijn een verandering in het onderliggende gedrag zichtbaar wordt.
- Het onderliggende gedrag van de tijdreeks kan in de loop van de tijd veranderen.

We nemen aan dat de tijdreeks stationair is, of gemaakt kan worden door een eindig aantal differenties te nemen (of eventueel andere transformaties toe te passen). De manier waarop een bepaalde modelspecificatie in de identificatiefase¹⁾ reageert op bovenstaande veranderingen van de onderliggende tijdreeks hangt nauw samen met de mate van robuustheid van de modelspecificatie. Binnen

1) In tegenstelling tot het gebruikelijke drie-stapsproces, identificatie, schatting en verificatie, bedoelen we hier met identificatie ook de schattingsfase.

de klasse van ARIMA-modellen zijn de modellen in het algemeen vrij robuust, zie McLeod(1982)²⁾

In de praktijk wordt vaak gebruik gemaakt van tijdreeksdata, die regelmatig (bijvoorbeeld maandelijks) beschikbaar komen. Aan de beschikbare data liggen meestal administratieve processen ten grondslag. Correcties op eerder beschikbare data vormen meer regel dan uitzondering. Eén en ander kan tot gevolg hebben dat bepaalde karakteristieken van een tijdreeks zich wijzigen. Te denken valt aan een verandering van bijvoorbeeld het seizoenspatroon.

Met behulp van een monitor systeem wordt gesignaleerd of de uitbreiding en/of de correctie van de tijdreeksdata van invloed is op de ARIMA-specificatie. Een signaal wordt afgegeven indien de invloed significant is, dat wil zeggen dat de nieuwe gegevens in strijd lijken te zijn met het van te voren geschatte voorspelmodel zodat dit aanpassing behoeft.

Hoe werkt een monitor systeem

Indien men voorspelssystemen gebruikt verdient het aanbeveling om te beschikken over indicaties die inzicht geven in het functioneren van het systeem. Als indicator wordt gewoonlijk de voorspelfout genomen. Door (automatisch) te toetsen op de voorspelfout krijgt men inzicht in de betrouwbaarheid van de in het voorspelsysteem gebruikte voorspelmodellen. De meeste voorspelmodellen met tijdsonafhankelijke parameters volgen stapsgewijze veranderingen in het gemiddelde, de trend en/of het seizoenspatroon enigszins vertraagd. Indien veranderingen optreden moeten ze zo snel mogelijk worden gesignaleerd. Deze signalering kan aanleiding geven tot een interventie door een deskundige om zo nodig de specificatie van het voorspelmodel aan te passen aan de gewijzigde omstandigheden.

Abraham en Ledolter(1983) en Gardner(1983) beschrijven een aantal indicatoren die toegepast kunnen worden om te signaleren wanneer een voorspelmodel niet meer geschikt is om goede voorspellingen te genereren. Als het waarschuwingssignaal een bepaalde controlegrens overschrijdt dan concluderen ze dat de voorspelmodellen niet meer beheerst ('in control') zijn. Dit betekent dat er een grote kans bestaat dat er slechte voorspellingen gegenereerd worden. De waarschuwingssignalen van genoemde auteurs zijn afhankelijk van al dan niet

2) Hier moeten we rekening houden met de fase schatten of voorspellen. Het is in de praktijk zeer wel mogelijk dat een model met een foute specificatie goede voorspellingen genereert.

gesmoothde (gekwadrateerde) cumulatieve sommen van de één-staps-vooruit voorspelfout. Aan deze indicatoren zijn zowel voordelen als nadelen verbonden. Stel bijvoorbeeld eens dat er op een bepaald moment een grote voorspelfout optreedt, een incident waarvoor geen redelijke verklaring is te vinden. Deze voorspelfout heeft dan een grote invloed op de indicator en het kan dan voorkomen dat er een signaal afgegeven wordt, - 'let op, er is iets aan de hand met het voorspel-model' -, terwijl er in feite niets mankeert aan de modelspecificatie. Het systeem is nog steeds beheerst. Als algemeen nadeel van de ons bekende monitor systemen kunnen we opmerken dat, hoewel er signalen afgegeven worden met betrekking tot mogelijke invaliditeit van het voorspelmodel, deze signalen soms 'loos alarm' zijn. Dit is eigen aan het toetsen van hypothesen: indien de controlegrenzen voor het signaal een toets op niveau γ betekenen, zal de kans op loos alarm per signaal steeds gelijk zijn aan γ .

McLeod (1982)³⁾ heeft een monitor systeem geconstrueerd dat een ARIMA-voorspelmodel in beschouwing neemt. Het gespecificeerde ARIMA-model is gebaseerd op alle beschikbare data. In het monitor systeem worden vijf basis toetsen⁴⁾ uitgevoerd, die betrekking hebben op het opsporen van:

- afzonderlijke, abnormale residuen;
- verandering in de variantie van de residuen;
- verandering in het gemiddelde van de reeks;
- verandering van de geschatte modelparameters;
- verandering van de modelstructuur.

Wij hebben, evenals McLeod een monitor systeem geconstrueerd dat het gehele ARIMA-model in beschouwing neemt.

Op tijdstip $t=n$ hebben we de beschikking over een tijdreeks $\{y_t, t=1, \dots, n\}$ welke gerepresenteerd kan worden door een ARIMA $(p,d,q) \cdot (P,D,Q)_s$ proces.

$$(1 - \sum_{j=1}^p \phi_j B^j) (1 - \sum_{j=1}^P \Phi_j B^{js}) \nabla^d \nabla_s^D y_t = \theta_0 + (1 - \sum_{j=1}^q \theta_j B^j) (1 - \sum_{j=1}^Q \Theta_j B^{js}) a_t \quad (1)$$

waarin: ϕ_i, Φ_j (seizoens) autoregressieve parameter $i=1, \dots, p, j=1, \dots, P$
 θ_i, Θ_j (seizoens) moving average parameter $i=1, \dots, q, j=1, \dots, Q$
 B vertragsingsoperator zodanig dat $By_t = y_{t-1}$
 d, D orde (seizoens) differentie

3) Het werk van McLeod is voor een zeer groot deel afhankelijk van de onderzoeksvrchttingen van Jenkins. Zie bijvoorbeeld ook Jenkins (1983).

4) De toetsen van McLeod zijn meer gebaseerd op visuele inspectie van diverse karakteristieken, zoals in de Box Jenkins filosofie gebruikelijk is, dan op statistische toetsen.

s	seizoensperiode
λ	transformatie parameter ($\ln y_t$ ($\lambda=0$), y_t ($\lambda=1$))
θ_0	constante term
a_t	witte ruis reeks, onderling onafhankelijk identiek $N(0, \sigma_a^2)$

We veronderstellen dat een versie van model (1) op tijdstip $t=n$ is geschat en valide is bevonden. Dit levert ons de verzameling modelparameters en toetsgrootheden, $\{\beta_n, SE(\beta_n), s_a^2(n), Q^*(n)\}$, op. Hierin is β_n een geschatte parameter uit de hierboven gedefinieerde verzameling $\{\varphi, \theta, \Phi, \Theta\}$, $SE(\beta_n)$ de standaardfout van β_n , $s_a^2(n)$ de geschatte residuele variantie van de a_t en $Q^*(n)$ de waarde van de Portmanteau-toetsgrootheid. Tevens worden er voorspellingen gegenereerd, $y_n(k)$, $k=1, \dots, 24$, alsmede een 90% voorspelinterval.

Op tijdstip $t=n+1$ komt de waarneming y_{n+1} beschikbaar. Vervolgens herschatten we de gespecificeerde versie van model (1) met de beschikbare tijdreeks $\{y_t, t=1, \dots, n+1\}$. In dit stadium van het voorspelproces wordt het monitor systeem ter hand genomen. De in het monitor systeem gehanteerde toetsen worden in het navolgende toegelicht.

Monitor Toetsen

De in ons monitor systeem geprogrammeerde toetsen vallen uiteen in de volgende serie toetsen:

- toets op de voorspelprestatie
- toets op de validiteit van het voorspelmodel
- toets op een verandering van de geschatte modelparameters
- toets op de verandering van de modelstructuur.

Voordat we gaan toetsen moeten we onbetrouwbaarheidsdrempel γ vastleggen. Men moet zich dan eigenlijk afvragen in welke mate de deskundige geneigd is om een model te wijzigen als daar signalen voor worden afgegeven. Wil men een snel signaal, dan stelt men γ hoog in. Is men niet geneigd om snel actie te ondernemen, dan stelt men de onbetrouwbaarheidsdrempel laag in. We gaan achtereenvolgens op elk van de toetsen in.

Toets op de voorspelprestatie

Om te toetsen of een voorspelmodel goede voorspellingen heeft gegenereerd, gebruiken we een signaleringsmethode zoals voorgesteld door Trigg(1964). Dit

signaal vergelijkt de cumulatieve som van de één-staps-vooruit gesmoothde voorspelfouten aan het eind van elke periode met de gesmoothde gemiddelde absolute afwijking (MAD). Hiertoe bestaan een aantal actualiseringsvergelijkingen, die bij elke nieuwe waargenomen voorspelfout de volgende gedaante hebben:

$$\begin{aligned}e(t) &= Y(t) - F(t) \\ \text{SUM}(t) &= \alpha e(t) + (1 - \alpha) \text{SUM}(t-1) \\ \text{MAD}(t) &= \alpha |e(t)| + (1 - \alpha) \text{MAD}(t-1) \\ T(t) &= |\text{SUM}(t) / \text{MAD}(t)|\end{aligned}$$

De voorspelfout $e(t)$, is de gerealiseerde tijdreekswaarneming $Y(t)$, minus de voorspelling $F(t)$. Zoals gebruikelijk neemt de smoothingparameter α waarden aan tussen 0 en 1. Indien nu $T(t)$ een bepaalde van te voren vastgestelde controle-grens overschrijdt dan volgt er een signaal. Dit kan betekenen dat het voorspelmodel niet meer geschikt is om goede voorspellingen te leveren. Voor deze $T(t)$ wijst een signaal op systematisch te hoge of systematisch te lage voorspellingen.

De vraag of de voorspelfouten onafhankelijk zijn, verlangt nadere uitleg. Sommige tijdreeksmethoden, zoals de Box Jenkins benadering, zijn opgezet met de bedoeling, te komen tot een model met ongecorreleerde residuen. Dan zullen de voorspelfouten ook ongecorreleerd zijn. In andere methoden, in het bijzonder exponentiële smoothing, kan men verwachten dat de voorspelfouten gecorreleerd zijn. De aanwezigheid van deze autocorrelatie leidt tot een andere keuze van de controlegrenzen, zie Gardner(1983).

Het signaleringsmechanisme wordt opgestart met de startwaarden $\text{SUM}(0) = 0$ en $\text{MAD}(0)$ gelijk aan een schatting van de MAD. Deze schatting wordt gebaseerd op de residuen van het schattingsproces, en is gelijk te stellen aan zijn geschatte verwachting onder normaliteit, $s_a \sqrt{\frac{2}{\pi}}$. Als waarde voor de smoothingparameter wordt $\alpha = 0.1$ voorgesteld. De waarde van de controlegrens is: $T(t) = 0.47$ bij een onbetrouwbaarheidsdrempel van 10% en $T(t) = 0.54$ bij een onbetrouwbaarheidsdrempel van 5%, zie Gardner(1983).

Het gebruik van deze gestandaardiseerde CUSUM, $T(t)$, heeft voordelen ten opzichte van de eenvoudige CUSUM, $C(t)$. Als belangrijk nadeel van de eenvoudige CUSUM geldt dat grote voorspelfouten nooit worden vergeten. Dit laatste effect treedt hier niet op. We bespreken dit geval bij de toets op een verandering van de geschatte modelparameters.

Als er een signaal wordt afgegeven met betrekking tot de voorspelprestatie dan kan men met behulp van 'externe informatie' vaststellen of de opgetreden voorspelfout het mogelijke gevolg is van bijvoorbeeld een lokale verstoring

(zoals een staking). Is er voor het waarschuwingssignaal geen verklaring te vinden aan de hand van de beschikbare externe informatie, dan zullen we de resultaten van de overige monitor toetsen moeten afwachten. Levert dit geen verklaring dan moeten we besluiten dat de grote voorspelfout volledig van het toeval afhangt, of dat er 'loos alarm' is geslagen.

Tenslotte dient te worden opgemerkt dat als er een waarschuwingssignaal wordt afgegeven, de waarde van $SUM(t)$ op nul gesteld moet worden, alvorens men verder gaat. Doet men dit niet dan zal $T(t)$ de controlegrenzen blijven overschrijden, zie Wheelwright en Makridakis(1985).

Toets op de validiteit van het voorspelmodel

Is het model geschat met $n+1$ waarnemingen valide? De schattingsmethode is gebaseerd op het principe van de maximale aannemelijkheid (ML schatters). Derhalve zijn de geschatte parameters bij benadering normaal verdeeld. We kunnen dus op de gebruikelijke manier hypothesen toetsen, zie o.a. Nerlove, Grether en Carvalho(1979) of Harvey(1981). De validiteitstoetsen kunnen we uitvoeren met de verzameling modelparameters en toetsgrootheden $\{\beta_{n+1}, SE(\beta_{n+1}), s_a^2(n+1), Q^*(n+1)\}$, vergelijk onder andere Box en Jenkins(1976). Om de significantie van de parameters β_{n+1} te bepalen gebruiken we een t-toets. Als een parameter niet significant van nul verschilt dan wordt er een signaal afgegeven. Een niet significante parameter kan eventueel op basis van het 'parsimony'-principe uit het model worden geëlimineerd.

Vervolgens toetsen we de residuen op hun 'witte ruis' karakter. Dit betekent ten eerste dat hun verwachting nul is. Wordt deze nulhypothese verworpen dan volgt er een signaal. Ten tweede worden de residuen geacht geen onderlinge autocorrelatie te hebben. Dit kan worden getoetst met de Portmanteau-toetsgrootheid Q^* . Deze toetsgrootheid is onder de nulhypothese chi-kwadraat verdeeld, zie Box en Jenkins(1976), Box en Pierce(1970). Als Q^* de kritieke waarde van de bijbehorende chi-kwadraat verdeling overschrijdt, dan wordt er een signaal afgegeven. Er is dan twijfel aan de validiteit van het voorspelmodel, en dient eventueel de modelbouwfase bestaande uit het drie-staps proces, identificatie, schatting en verificatie, uitgevoerd te worden.

Toets op een verandering van de geschatte modelparameters

Het is mogelijk dat modelparameters in de loop van de tijd veranderen zonder dat er een verandering in de modelstructuur is opgetreden. Het is noodzakelijk deze veranderingen in de waarden van de modelparameters te toetsen.

De modelparameters β worden geschat met behulp van een maximale aannemelijkheidsprocedure; de schatters zijn bij benadering normaal verdeeld. We gaan uit

van de geschatte parameters op tijdstip $t=n$, en we nemen deze schattingen β_n als referentiepunt β . Indien de modelparameters zich niet of nauwelijks wijzigen zullen de schattingen β_{n+1} zich bewegen rondom de waarde β , dat wil zeggen de stochast $b(t) = (\beta_{n+1} - \beta)$ schommelt rondom nul. Vervolgens kunnen we de verandering van de modelparameters toetsen door toepassing van de eenvoudige CUSUM techniek.

$$\begin{aligned}\text{SUM}(t) &= b(t) + \text{SUM}(t-1) \\ \text{MAD}(t) &= \alpha |b(t)| + (1 - \alpha) \text{MAD}(t-1) \\ C(t) &= |\text{SUM}(t) / \text{MAD}(t)|\end{aligned}$$

Zoals gebruikelijk neemt de smoothingparameter α waarden aan tussen 0 en 1. Indien nu $C(t)$ een bepaalde van te voren vastgestelde controlegrens overschrijdt dan volgt er een signaal. Dit kan betekenen dat de modelparameters zijn veranderd. Met betrekking tot de controlegrens heeft Brown(1963) aangetoond dat de eenvoudige CUSUM $C(t)$ in het geval van exponentiële smoothing bij benadering normaal verdeeld is met gemiddelde nul en standaardfout $0.884/\sqrt{\alpha}$. De validiteit van deze normale benadering voor $C(t)$ wordt door Gardner(1983) bevestigd door middel van simulatie. Het signaleringsmechanisme wordt opgestart met de startwaarden $\text{SUM}(0) = 0$ en door $\text{MAD}(0)$ gelijk te stellen aan zijn geschatte verwachting onder normaliteit, $SE(\beta)\sqrt{\frac{2}{\pi}}$. Als waarde voor de smoothingparameter wordt $\alpha = 0.1$ voorgesteld.

Het gebruik van $C(t)$ heeft zowel voordelen als nadelen. Een van de voordelen is dat $C(t)$ onafhankelijk is van de gebruikte smoothingparameter. Als belangrijk nadeel van $C(t)$ geldt dat grote afwijkingen van de parameterschattingen ten opzichte van de referentieschattingen nooit worden vergeten. Dit laatste effect kunnen we als volgt illustreren. Stel dat één geïsoleerde grote afwijking er voor verantwoordelijk is dat $C(t)$ van een positie rondom de waarde nul verschuift naar een positie rondom de controlegrens. Vanaf deze gebeurtenis worden de afwijkingen weer klein verondersteld. We zien dat $C(t)$ rondom de controlegrens schommelt. We kunnen signalen verwachten, waaruit we mogelijk kunnen concluderen dat er significante wijzigingen in de parameterschattingen van het voorspelmodel opgetreden zijn, terwijl er feitelijk niets aan de hand is.

Toets op de verandering van de modelstructuur

Het kan voorkomen dat de parameters van een ARIMA-model veranderen ten gevolge van het optreden van een trendbreuk in de tijdreeks. Het optreden van een trendbreuk kan zelfs leiden tot een volkomen andere modelspecificatie. Uit de drie voorgaande toetsen, herhaaldelijk toegepast na toevoeging van een nieuwe

waarneming aan de tijdreeks, kunnen we signalen destilleren, die het optreden van een trendbreuk kunnen markeren. Een trendbreuk openbaart zich in de loop van de tijd en is zeker niet zichtbaar na één periode. Vandaar dat we als toevoeging op de eerdere monitortoetsen hebben gezocht naar een meer algemeen toepasbare toetsgrootheid. We hebben, naar analogie van de klassieke econometrie, de Chow-toets gekozen. Zie bijvoorbeeld Chow(1960), Maddala(1970) en Harvey (1981). Deze toets is ons insziens toepasbaar omdat de residuen a_t uit model (1) bij benadering normaal verdeelde stochasten zijn, en de parameters worden geschat met behulp van de maximale aannemelijkheids schattingsprocedure.

We schatten het voorspelmodel, bestaande uit k modelparameters, met behulp van n waarnemingen. We stellen de modelparameters voor als β_n . We definiëren RSS1 als de residuele kwadratensom gebaseerd op n waarnemingen. Het aantal graden van vrijheid is: $n-k-1$. Nadat er m 'nieuwe' waarnemingen beschikbaar zijn gekomen, $1 \leq m \leq k$, gaan we de hypothese toetsen dat de modelparameters niet zijn veranderd, d.w.z. er is geen trendbreuk opgetreden.

$$\begin{aligned} H_0 &: \beta_n = \beta_{n+m} \\ H_1 &: \beta_n \neq \beta_{n+m} \end{aligned}$$

We geven de geschatte modelparameters van het met $(n+m)$ waarnemingen geschatte voorspelmodel aan met β_{n+m} en we definiëren RSS als de residuele kwadratensom gebaseerd op $n+m$ waarnemingen. In het laatste geval hebben we $(n+m-k-1)$ graden van vrijheid.

De toetsgrootheid F ziet er als volgt uit:

$$F = \frac{(RSS - RSS1)/m}{RSS1/(n-k-1)}$$

Deze toetsgrootheid is F -verdeeld met m en $(n-k-1)$ graden van vrijheid.

De toetsing vindt bij onze toepassingen plaats voor $m = 3$, dus feitelijk per kwartaal.

Ten aanzien van de toetsgrootheid kunnen we nog opmerken dat als $m = 1$ het gebruik van de Chow toets te vergelijken is met het gebruik van het voorspel-interval.

Enkele resultaten van de Toepassing van het Monitor Systeem

De werking van het monitor systeem is getest met behulp van tien willekeurig gekozen tijdreeksen uit een PTT-omgeving. Per tijdreeks is er vooraf een ARIMA-model gespecificeerd. Deze tijdreeksen zijn, gedurende een periode van twaalf maanden, maandelijks geactualiseerd en de ARIMA-modellen zijn derhalve maandelijks herschat.

We passen de monitor toetsen toe en nemen als onbetrouwbaarheidsdrempel $\gamma = 5\%$. Het voordeel hiervan is dat in dit geval zowel de controlegrenzen als de waarden van de F-verdeling getabelleerd zijn.

Toepassing van de monitortoetsen leidt tot het resultaat dat twee van de tien tijdreeksmodellen, gedurende de gehele exercitie, geen waarschuwingssignalen afgeven. Zij beschrijven de tijdreeksen goed, en de voorspelfouten zijn beheersbaar.

De toets op de voorspelprestatie geeft voor vijf van de tien tijdreeksen geen enkel waarschuwingssignaal. Slechts in 8 van de 120 toetsingen komt er een waarschuwingssignaal (6,7%). Drie van de vijf tijdreeksen hadden een éénmalig optredend waarschuwingssignaal. De resterende twee tijdreeksen waren verantwoordelijk voor de overige waarschuwingssignalen. Voor een van de twee tijdreeksen heeft het optreden van een grote voorspelfout geleid tot herspecificatie van het ARIMA-model. We komen hier in het vervolg nog op terug. Voor de overblijvende zeven waarschuwingssignalen was er geen verklaring te geven voor het optreden van een grote voorspelfout. Deze grote voorspelfouten zijn derhalve geïnterpreteerd als toevallige verstoringen.

De toets op de validiteit van het voorspelmodel leidt in 1 van de 120 gevallen (0,8%) tot verwerping van het tot dan toe gebruikte voorspelmodel. Dit signaal is ook bij de toets op de voorspelprestatie al waargenomen. De verwerping van het voorspelmodel werd veroorzaakt door instabiliteit van het model. De parameterschatting in een eerste orde moving average polynoom werd groter dan 1. Herspecificatie van het model heeft plaatsgevonden door uitbreiding van de orde van het moving average polynoom tot 2.

De toets op een verandering van de geschatte modelparameters geeft voor zes van de tien tijdreeksen geen enkel waarschuwingssignaal. Gedurende de toetsingsfase zijn er 348 modelparameters geschat. Slechts in vijf gevallen is er een waarschuwingssignaal afgegeven (1,4%). Deze waarschuwingssignalen kwamen alleen voor bij de seizoensparameters in de seizoens-moving average polynomen.

De toets op de verandering van de modelstructuur kon in 16 van de 120 mogelijkheden (13,3%) niet worden toegepast. De teller van de toetsgrootte kreeg een negatieve waarde. Dit verschijnsel wordt veroorzaakt door een combinatie

van 'backforecasting' en het niet lineaire iteratieve schattingsalgorithme, zoals dat wordt toegepast in de programmatuur van Jenkins & Partners. Voor zeven van de tien tijdreeksen verkrijgen we geen waarschuwingssignalen. Herspecificatie van de modellen leidde niet tot een verbetering van de gebruikte voorspelmodellen.

Referenties

Abraham, B. en J. Ledolter, (1983), Statistical Methods for Forecasting, John Wiley & Sons, New York

Box, G.E.P. en G.M. Jenkins, (1976), Time Series Analysis: Forecasting and Control, Holden-Day, San Francisco, revised edition

Box, G.E.P. en D.A. Pierce, (1970), 'Distribution of residual autocorrelation in autoregressive integrated moving average time series models', Journal of the American Statistical Association, vol. 65, pp. 1509-1526

Brown, R.G., Smoothing, Forecasting and Prediction of Discrete Time Series, Englewood Cliffs, N.J.: Prentice Hall

Chow, G.C., (1960), 'Tests of Equality between Subsets of Coefficients in Two Linear Regressions', Econometrica, vol. 28, pp. 591-605

Harvey, A.C., (1981), The Econometric Analysis of Time Series, Philip Allan, Oxford

Gardner, E.S., (1983), 'Automatic Monitoring of Forecast Errors', Journal of Forecasting, vol. 2, pp. 1-21

Granger, C.W.P. en P. Newbold, (1977), Forecasting Economic Time Series, Academic Press, New York

Jenkins, G.M., (1979), Practical Experiences with Modelling and Forecasting Time Series, GJP publications, Lancaster

Maddala, G.S., (1977), Econometrics, McGraw-Hill

McLeod, G., (1982), Box Jenkins in Practice, GJP publications, Lancaster

Nelson, C.R., (1973), Applied Time Series Analysis for Managerial Forecasting, Holden-Day, San Francisco

Nerlove, M., D.M. Grether en J.L. Carvalho, (1979), Analysis of Economic Time Series: A Synthesis, Academic Press, New York

Newbold, P., (1983), 'ARIMA Model Building and the Time Series Analysis Approach to Forecasting', Journal of Forecasting, vol. 2, pp. 23-35

Trigg, D.W., (1964), 'Monitoring a Forecast System' Operational Research Quarterly, vol. 15, pp. 271-274

Wheelwright, S.C. en S. Makridakis, (1985), Forecasting Methods for Management, John Wiley & Sons, New York, 4th edition

Ontvangen: 85
Geaccepteerd: 24-09-86