

EEN TOEPASSING VAN MODELLEN BIJ HET VOORSPELLEN VAN DE TOTALE DIESELOLIEAFZET IN NEDERLAND

P.C.A. ABEN¹

1. Samenvatting

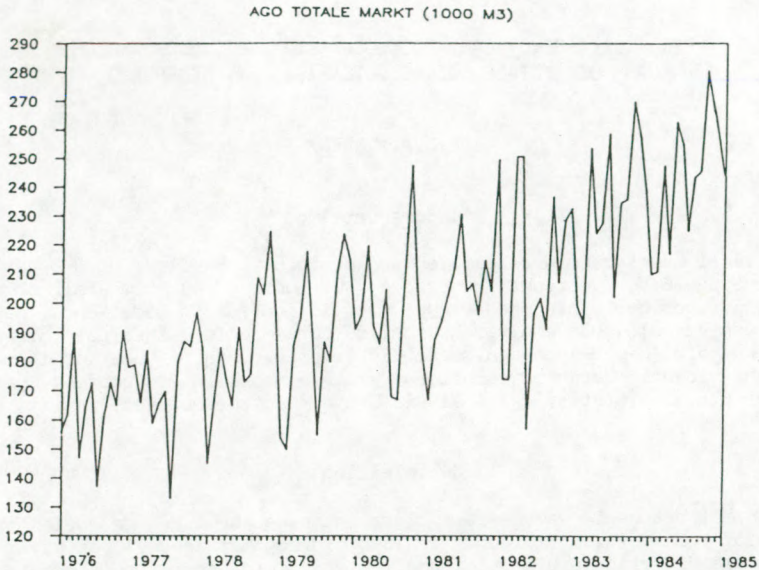
Dit artikel geeft enkele resultaten van een t.b.v. Shell Nederland Verkoopmaatschappij B.V. verricht onderzoek. Er worden een tweetal modellen opgesteld voor de totale maandelijkse dieselolieafzet in Nederland, gevolgd door een vergelijkende empirische analyse tussen beide. Het gaat hierbij om een eenvoudig AR-model en een gecompliceerder overdrachtsfunktieruis-model. Voor het identificeren en schatten van deze modellen is gebruik gemaakt van het S.A.S.- (Statistical Analysis System) softwarepakket.

2. Inleiding

Tot en met het begin van de zeventiger jaren was benzine de belangrijkste brandstof voor motorvoertuigen. In 1975 bedroeg het aandeel van benzine in de totale Nederlandse brandstoffenafzet voor motorvoertuigen, gemeten in metrische tonnen, ca. 2/3. Daarna is door substitutie op dieselolie en autogas bij personenauto's dit aandeel gestaag gedaald (in 1985 bedraagt haar aandeel nog slechts de helft). Deze substitutie hangt nauw samen met het financiële voordeel dat het rijden op dieselolie of autogas biedt bij een hoog jaarkilometrage. De lagere accijnzen op dieselolie en autogas t.o.v. benzine leiden er immers toe dat een hogere motorrijtuigenbelasting en een hogere aanschafprijs voor het motorvoertuig bij een groot jaarkilometrage meer dan gecompenseerd worden. Automobilisten rijdend op dieselolie of autogas rijden dan ook gemiddeld per jaar meer kilometers dan benzine-rijders.

Met name de totale afzet van dieselolie is in Nederland de laatste tijd flink geëxpandeerd. Ter illustratie zijn de maandelijkse totale dieselolieafzetcijfers weergegeven in figuur 1. Naast de reeds genoemde substitutie van benzine op dieselolie bij personenauto's, is deze groei te verklaren door een opleving van de economie in Nederland. Een hogere economische bedrijvigheid betekent immers onmiddellijk een groter aanbod van te vervoeren goederen voor de vervoersector. Aangezien het transport over de weg vrijwel geheel met motorvoertuigen rijdend op dieselolie geschiedt, verhoogt dit de vraag naar dieselolie.

¹Correspondentieadres: Van Meelstraat 110, 5703 ET HELMOND, tel: 04920 - 25567. De auteur bedankt de heren R.M.J. Heuts (Katholieke Hogeschool Tilburg) en G. Wolf (Shell Nederland Verkoopmaatschappij B.V.) voor hun hulp bij het tot stand komen van dit artikel.



Figuur 1 De binnenlandse totale dieselolieafzet in 1000 m³.

In dit artikel worden twee modellen voor de totale maandelijkse dieselolieafzet in Nederland besproken.² Het gaat hierbij om een naïef AR-model en een realistischer overdrachtsfunctieruis-model (OFR-model) (zie bv. Box en Jenkins (1976)). Het eerstgenoemde model wordt gepresenteerd in paragraaf drie. Zij behoort tot de klasse van ARIMA-modellen. Deze hebben als kenmerk dat zij de afhankelijke variabele volledig verklaren uit de historie van deze grootheid zelf. Dit in tegenstelling tot het OFR-model, dat in paragraaf vier besproken wordt. Bij dit soort modellen spelen tevens verklarende variabelen een rol. De geschiktheid van de volgende grootheden als verklarende variabele wordt bekeken:

²Het gaat hierbij om cijfers verstrekt door het C.B.S., hoofdafdeling statistieken van industrie en bouw nijverheid. De cijfers tot en met 1981 zijn gecorrigeerd voor industriële gasolie leveringen aan de landbouw, die ten onrechte in het dieseloliecijfer waren opgenomen.

1. De economische conjuncturele situatie in Nederland, gemeten door de Economisch Dagblad Leading Indicator (E.D.L.I.).³
2. Het substitutieeffect, gemeten door het aandeel van dieselpersonenauto's in het totale Nederlandse actieve personenautopark.⁴
3. Het werkdageffect, gemeten door het aantal werkdagen (ma. tot en met vr.) in een maand. Deze variabele is relevant omdat men in een maand met een geringer aantal werkdagen ook minder dieselolie kan afzetten (op niet-werkdagen wordt er geen dieselolie geleverd aan tankstations en bedrijfspompen). Er wordt geen rekening gehouden met feestdagen.⁵
4. Een dummyvariabele voor een bijzondere gebeurtenis in 1982, namelijk een speculatie door olie-handelaren. Dit is duidelijk te maken aan de hand van het prijsverloop van gasolie (het basisprodukt voor dieselolie) in die periode. De prijs van gasolie stijgt sterk in april 1982 ($\pm 13\%$), waarna precies op de maandwisseling het omslagpunt volgt en de prijs zich in mei weer stabiliseert. Dit duidt erop dat handelaren in april 1982 bij sterk stijgende prijzen grote speculatieve voorraden dieselolie aannamen. Omdat de opslag van dieselolie vrij kostbaar is, kan veilig verondersteld worden dat alle in april verworven speculatieve volumina in mei weer zijn afgestoten. Zodoende was het dieselolieafzetcijfer voor april 1982 hoog en het meicijfer abnormaal laag. Aangezien deze exceptionele gebeurtenis een nogal grote invloed heeft op de voorspellingen van de modellen is gekozen voor de invoer van de volgende dummyvariabele in het nog te bespreken OFR-model:

$$(2.1) \text{ DUMMY}_t = \begin{cases} 1, & t = \text{april } 1982; \\ -1, & t = \text{mei } 1982; \\ 0, & \text{elders.} \end{cases}$$

³ Deze conjunctuurindicator wordt regelmatig door het Economisch Dagblad gepubliceerd.

⁴ Het actieve personenautopark wordt berekend uit het administratieve personenautopark, zoals dit volgt uit gegevens van de Rijksdienst voor het Wegverkeer (RDW), verminderd met schattingen voor handelsvoorraden en niet bij de RDW aangemelde sloop. Bron: C.B.S., hoofdafdeling statistieken van verkeer en vervoer.

⁵ Bedrijfspompen zijn niet voor het publiek opengestelde eigen tankinstallaties bij vervoerders.

Vervolgens wordt in paragraaf vijf de voorspelkracht van het AR-model met die van het OFR-model vergeleken. Deze voorspelkracht is een belangrijk element in de onderlinge empirische vergelijking. Op voorhand kan immers al gezegd worden dat een OFR-model zal leiden tot een betere verklaring van de afhankelijke variabele. Het is echter nog maar de vraag of dit ook geldt voor de voorspelkracht. Slecht voorspelbare verklarende variabelen kunnen immers de voorspelkracht van een OFR-model nadelig beïnvloeden. Tot slot worden in paragraaf zes enkele conclusies vermeld.

3. Het AR-model

Voor de bouw van een AR-model voor de maandelijkse dieselolieafzet in m^3 in Nederland zijn C.B.S.-maandcijfers gebruikt over de periode januari 1976 tot en met december 1983 (96 waarnemingen). De geschatte autocorrelatiefunctie (acf) en partiële autocorrelatiefunctie (pacf) van deze reeks zijn weergegeven in figuur 2. De grafiek laat een licht stijgende trendmatige ontwikkeling zien, een beeld dat bevestigd wordt door een niet zo snel uitdempende acf. Verder vertoont de reeks een seizoensmatig karakter, zoals onder andere blijkt uit de significante piek in de pacf op lag zes.⁶ Deze verschijnselen leiden echter niet tot niet-stationariteit van de reeks, zoals volgt uit de hieronder vermelde geschatte modellen 3.1 en 3.2. De relevante AR-parameters in deze modellen verschillen op significante wijze van een. De getallen tussen haakjes zijn de geschatte standaardafwijkingen van de geschatte parameters.

$$(3.1) \quad (1 - 0,54 L) (D_t - 194917) = \hat{\varepsilon}_{1t} \\ (0,09) \quad (5613)$$

$$(3.2) \quad (1 - 0,64 L^6) (D_t - 196257) = \hat{\varepsilon}_{2t} \\ (0,08) \quad (6600)$$

Hierbij is:

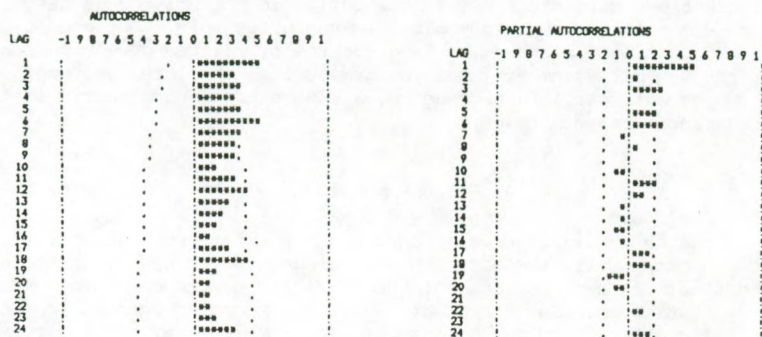
D_t : De totale dieselolieafzet in Nederland in maand t in m^3 ;

$\hat{\varepsilon}_{1t}$: De (geschatte) ruisterm in maand t .

L : De vertragingoperator, gedefinieerd als $L(D_t) = D_{t-1}$. Ook geldt:

$$L^p(D_t) = L^{p-1}(L(D_t)) = D_{t-p}.$$

Alle uitspraken zijn gebaseerd op tweezijdige toetsen bij een α van 0,05.



Figuur 2 De acf en pacf van de reeks van de maandelijkse dieselolieafzet in Nederland (waarnemingsperiode januari 1976 tot en met december 1983).

Aangezien het nemen van differenties van de reeks overbodig is, is de analyse voor de bouw van het naïeve model beperkt tot de klasse van de ARMA-modellen. Op grond van de vertoonde acf en pacf ligt een model met een Ar-parameter op lag een en zes, eventueel aangevuld met een Ar- of MA-parameter op lag drie, voor de hand. Bij het schatten blijkt dan echter dat de geschatte residuen van deze modellen geen proces van witte ruis vormen maar significante (partiële) autocorrelaties op lags 18 en 24 vertonen. Toevoeging van AR- of MA-parameters op deze lags leidt echter tot insignificantie van de AR-parameter op lag zes, zodat deze uit het model verwijderd is. Het volgende AR-model blijkt uiteindelijk in termen van percentage verklaarde variantie het beste te voldoen?

$$(3.3) \begin{pmatrix} 1 & -0,26 L & -0,18 L^3 & -0,41 L^{18} & -0,45 L^{24} \end{pmatrix} D'_t = \hat{\varepsilon}_{3t} \\ \begin{pmatrix} (0,07) & (0,07) & (0,09) & (0,10) \end{pmatrix}$$

$$D'_t = D_t - 167856 \\ (3183)$$

⁷ Onder percentage verklaarde variantie verstaan we dat deel (uitgedrukt in procenten) van de variantie van de tijdreeks van de afhankelijke variabele dat door het model verklaard wordt.

Het model geeft een percentage verklaarde variantie van 56%, en de geschatte residuen van het model vormen een proces van witte ruis. Met behulp van dit model zijn de twaalf maanden van 1984 voorspeld, uitgaande van de waarnemingen tot en met december 1983. De gemiddelde absolute voorspelfout (M.A.P.E.) over het jaar 1984 bedraagt 8,4%. De exacte voorspelresultaten zijn weergegeven in paragraaf vijf.

4. Het OFR-model

In deze paragraaf analyseren we of opname van verklarende variabelen de kwaliteit van het model verbetert. In dit verband worden enkele OFR-modellen geschat met telkens één van de in de inleiding genoemde grootheden als verklarende variabele. De structuur van de ruismodellen van de in deze paragraaf geschatte OFR-modellen is zodanig gekozen dat de geschatte residuen van deze ruismodellen een proces van witte ruis volgen. Voor opname van de bewuste verklarende variabelen in het uiteindelijke OFR-model worden de volgende criteria gebruikt:

1. Zijn de parameterschattingen voor de verklarende variabele in het OFR-model significant?
2. Geeft het geschatte OFR-model met de bewuste verklarende variabele een hoger percentage verklaarde variantie (over dezelfde waarnemingsperiode als bij het AR-model) en een betere voorspelkracht (gemeten door de M.A.P.E. over 1984), dan het AR-model?⁸

De resultaten van deze analyse zijn hieronder weergegeven:

$$(4.1) \quad D_t = 164858 + 21341 \text{ EDLI}_t + \hat{\eta}_{4t}$$

(6223) (7128)

$$(1 - 0,22 L - 0,19 L^3 - 0,42 L^{18} - 0,45 L^{24}) \hat{\eta}_{4t} = \hat{\varepsilon}_{4t}$$

(0,08) (0,08) (0,11) (0,12)

Percentage verklaarde variantie: 59%, M.A.P.E.: 8,7%

⁸De voor deze voorspellingen benodigde predikties voor de verklarende variabelen zijn als volgt verkregen. De voorspellingen voor de E.D.L.I. over 1984 zijn middels een voor deze variabele opgesteld AR-model gegenereerd. Verder produceert Shell intern reeds voorspellingen voor het substitutieeffect. De hier gehanteerde predikties voor deze variabele zijn afkomstig uit een vóór december 1983 door Shell opgesteld scenario. Verder is bij de in deze paragraaf vermelde modellen uitgegaan van gecentreerde verklarende variabelen (uitgezonderd de dummyvariabele), om een te hoge correlatie met de constante term te vermijden.

$$(4.2) D_t = 203451 + 18383 \text{ SUBST}_t + \hat{\eta}_{st}$$

(4521) (2180)

$$(1 - 0,38 L^{18} - 0,39 L^{24}) \hat{\eta}_{st} = \hat{\epsilon}_{st}$$

(0,10) (0,12)

Percentage verklaarde variantie: 59%, M.A.P.E.: 7,5%

$$(4.3) D_t = 164129 + 5381 \text{ DAG}_t + \hat{\eta}_{6t}$$

(6030) (1550)

$$(1 - 0,25 L - 0,15 L^3 - 0,46 L^{18} - 0,45 L^{24}) \hat{\eta}_{6t} = \hat{\epsilon}_{6t}$$

(0,08) (0,07) (0,10) (0,11)

Percentage verklaarde variantie: 60%, M.A.P.E.: 6,3%

$$(4.4) D_t = 162499 + 30965 \text{ DUMMY}_t + \hat{\eta}_{7t}$$

(6043) (11024)

$$(1 - 0,23 L - 0,17 L^3 - 0,45 L^{18} - 0,46 L^{24}) \hat{\eta}_{7t} = \hat{\epsilon}_{7t}$$

(0,08) (0,08) (0,11) (0,11)

Percentage verklaarde variantie: 59%, M.A.P.E.: 6,7%

Hierbij is:

D_t : De totale dieselolieafzet in Nederland in maand t in m^3 ;

$EDLI_t$: De waarde van de E.D.L.I. in maand t ;

$SUBST_t$: Het procentuele aandeel van dieselpersonenauto's in het totale actieve Nederlandse personenautopark in maand t ;

DAG_t : Het aantal werkdagen (ma. tot en met vr.) in maand t ;

$DUMMY_t$: Een dummyvariabele gedefinieerd volgens (3.2);

$\hat{\eta}_{it}$: De geschatte ruisterm van het overdrachtsfunctiegedeelte van het model;

$\hat{\epsilon}_{it}$: De geschatte ruisterm van het ruisgedeelte van het model. Deze term volgt een proces van witte ruis.

De hier geschatte OFR-modellen blijken, zoals te verwachten was, een hoger percentage verklaarde variantie op te leveren dan het AR-model. Verder zijn in alle gevallen de parameters van de toegevoegde verklarende variabelen significant. Echter waar andere verklarende variabelen een hogere voorspelkracht opleveren verzaakt de E.D.L.I. Dit kan verklaard worden uit het feit dat de E.D.L.I. zich slecht een jaar vooruit laat voorspellen. Dit in acht nemende wordt in het uiteindelijke OFR-model alleen het substitutieeffect, het dageffect en de dummy opgenomen. Het finale OFR-model heeft de volgende vorm:

$$(4.5) D_{t.} = 205909 + 19398 \text{ SUBST}_t + 5583 \text{ DAG}_t + 38613 \text{ DUMMY}_t + \hat{\eta}_{8t}$$

(3907) (2229) (1913) (11521)

$$(1 - 0,23 L^{12} - 0,40 L^{18}) \hat{\eta}_{8t} = \hat{\epsilon}_{8t}$$

(0,10) (0,11)

Percentage verklaarde variantie: 64%, M.A.P.E.: 3,6%

5. Voorspelresultaten

De maandelijkse voorspellingen van zowel het AR-model als het OFR-model voor 1984, uitgaande van waarnemingen tot en met december 1983, zijn vermeld in tabel 3. Bij de maandvoorspellingen is tussen haakjes de gemiddelde voorspelfout (A.P.E.) vermeld en bij de jaartotalen de M.A.P.E. Voor de wijze van voorspellen met behulp van AR-modellen en OFR-modellen wordt verwezen naar v.d. Genugten en Heuts (1982, p. 57 e.v. resp. p. 139 e.v.). Het AR-model voorspelt opvallend slecht voor de maanden april en mei. Dit wordt veroorzaakt door het reeds eerder vermelde speculatiegebeuren in 1982, dat wegens de faktor D_{t-24} doorwerkt in de voorspellingen.

De voorspelkracht over geheel 1984 van het OFR-model overtreft die van het AR-model in ruime mate zoals uit hun M.A.P.E.'s blijkt. Uit de tabel volgt echter wel dat het AR-model op de zeer korte termijn beter voldoet dan het OFR-model: de maanden januari, februari (en juli) worden beter voorspeld.

6. Conclusies

Uit het voorgaande is gebleken dat OFR-modellen voor de totale afzet van dieselolie in Nederland, zoals verwacht, leiden tot een hoger percentage verklaarde variantie. Het verschil is echter niet spectaculair te noemen (56% verklaarde variantie bij het AR-model en 64% bij het finale OFR-model).

OFR-modellen blijken echter niet in alle gevallen te leiden tot een betere voorspelkwaliteit. Lastig te voorspellen verklarende variabelen (zoals de

Maand '84	Realisatie	Voorspelling AR-model	Voorspelling OFR-model
Jan	210.096	221.568 (5,5%)	226.138 (7,6%)
Feb	211.381	210.333 (0,5%)	215.834 (2,1%)
Mrt	248.418	257.979 (3,8%)	250.819 (1,0%)
Apr	217.059	254.414 (17,2%)	232.141 (6,9%)
Mei	262.995	217.850 (17,2%)	249.830 (5,0%)
Jun	254.883	236.844 (7,1%)	245.825 (3,6%)
Jul	224.832	229.385 (2,0%)	230.441 (2,5%)
Aug	243.552	213.744 (12,2%)	242.065 (0,6%)
Sep	245.789	259.137 (5,4%)	245.948 (0,1%)
Okt	279.695	243.332 (13,0%)	265.657 (5,0%)
Nov	267.246	248.121 (7,2%)	256.536 (4,0%)
Dec	247.393	272.177 (10,0%)	260.493 (5,3%)
Totaal	2.913.339	2.864.884 (8,4%)	2.921.727 (3,6%)

Tabel 3 Voorspelresultaten van het AR-model en het OFR-model voor de totale Nederlandse dieselolieafzet (in m³).

E.D.L.I.) leiden, ondanks zeer significante verbanden met de afhankelijke variabele (t-waarde parameter 3,00), tot een verslechtering van de voorspelkwaliteit.

Opname van eenvoudiger en beter te voorspellen verklarende variabelen zoals het dageffect, de dummy en het substitutieffect, geven daarentegen wel een grote verbetering van de voorspelkwaliteit op de wat langere termijn.

Op de zeer korte termijn echter overtreft het AR-model het OFR-model qua voorspelkracht. Beide modellen zijn dus bruikbaar voor voorspellingen, het AR-model voor voorspellingen op de zeer korte termijn (een á twee maanden vooruit) en het OFR-model voor de voorspellingen op de langere termijn (tot

een jaar vooruit).

7. Literatuur

Box,G.,Jenkins,G.,1976. Time series analyses, forecasting and control. San Francisco: Holden-Day.

Genugten,B.v.d.,Heuts,R.,1982. Tijdreeksanalyse. Tilburg: Katholieke Hogeschool Tiburg.