

# Vuistregels voor voorspellingen in de marketing

Ph.H. Franses

Econometrisch Instituut  
Erasmus Universiteit Rotterdam

**Marketingbeslissingen op korte termijn worden met regelmaat gebaseerd op lange-termijnvoorspellingen. Hierbij worden modellen gebruikt waarbij uitspraken worden gedaan over lange-termijntrends en over eventuele verzadigingsniveaus. Aan de hand van de voorspelling over de penetratie van CD-spelers worden twee vuistregels gegeven die de kans op foutieve beslissingen bij zulke voorspellingen verkleinen.**

In veel praktische situaties kan de marketeer gebruikmaken van voorspellingen voor bijvoorbeeld de verkopen of cumulatieve verkopen van een produkt. Deze voorspellingen zijn van belang bij het formuleren van beleidsbeslissingen met betrekking tot bijvoorbeeld het inzetten van marketinginstrumenten als reclame en distributie. Vaak worden voorspellingen aangeleverd door een stafmedewerker of door een commercieel marktonderzoekbureau dat bij de rapportage toekomstprojecties levert. Hierbij worden veelal scenario's gehanteerd voor resultaten bij ongewijzigd beleid of bij een beleid waar een toename van marketinginspanningen wordt verondersteld.

Al naar gelang de uitkomsten van de verschillende scenario's overeenstemmen met de doelstellingen van de betreffende marketeer, zal men het beleid trachten om te buigen of bij te stellen. De effecten van deze ombuigingen kunnen eventueel weer worden betrokken bij nieuwe projecties, zodat de marketeer in staat is de effecten van verschillende beleidsmaatregelen vooraf te evalueren.

In de praktijk van het voorspellen maakt men onderscheid in korte-termijn- en lange-termijnvoorspellingen. Korte-termijnvoorspellingen beslaan meestal periodes korter dan een jaar, terwijl lange-termijnvoorspellingen over 5 tot 10 jaar of nog langer kunnen gaan. Het ligt voor de hand dat hoe verder men in de toekomst voorspelt, hoe onzekerder de uitkomst wordt. Een marktsituatie zal binnen nu en een half jaar niet zo snel veranderen als eenzelfde situatie binnen nu en tien jaar. Concreet betekent dit dat korte-termijnvoorspellingen kunnen worden gebaseerd op modellen die de huidige marktsituatie zo nauwkeurig mogelijk beschrijven. Hierbij wordt er van uit gegaan dat er in het consumentengedrag en de marktsituatie op korte termijn geen rigoureuze veranderingen optreden. Dit soort modellen is dan ook betrekkelijk eenvoudig te hanteren.

Soms echter heeft men behoefte aan projecties over de lange termijn, bijvoorbeeld over een tiental jaren. Een concreet voorbeeld hiervan is een voorspelling van het autopark in 2010 die nodig is om plannen voor de aanleg van wegen te maken. Een ander voorbeeld is het voorspellen van de penetratie van CD-spelers in Nederland. Men kan zich indenken dat het huidige of het korte-termijn inkoopbeleid of produktontwikkelingsbeleid af kunnen hangen van de lange-termijnprojecties.

Bij het maken van lange-termijnvoorspellingen moet men vaak een beslissing nemen over het al dan niet bestaan van een trend en

over het al dan niet bestaan van een verzadigingsniveau. Hoe een lange-termijntrend beschreven kan worden en hoe het verzadigingsniveau bepaald wordt, wordt hierna uitgelegd.

Het zal blijken dat de schijnbaar eenvoudige keuze voor een bepaald lange-termijnmodel geen onschuldige keuze is. Sterker nog, deze keuze zal de toekomstprojectie volledig domineren, en daarmee mogelijk misleidende korte-termijnbeleidsbeslissingen impliceren.

## Voorspellingen zonder verzadigingsniveau

Veel economische en marketingvariabelen hebben een trendend patroon zonder dat ze worden begrensd door een verzadigingsniveau. Voorbeelden zijn de niet-duurzame consumptieve bestedingen, de prijs van een nieuwe auto en de hoogte van de reclamebestedingen. Stel dat we zo'n variabele weergeven met het symbool  $y(t)$ , waarbij de index  $t$  het tijdstip van waarneming aanduidt. Het waarnemingsinterval kan dagen, weken, maanden of jaren zijn. Er zijn twee eenvoudige manieren om dit trendende karakter van  $y(t)$  te beschrijven in modelvorm. De eerste is het zogenaamde deterministische trend-model (DT-model), bijvoorbeeld:

$$y(t) = \alpha + \beta t.$$

De  $t$  staat voor de reeks  $0,1,2,\dots$ . Deze  $t$  wordt de deterministische trend genoemd omdat alle volgende en voorgaande waarnemingen van  $t$  perfect voorspelbaar zijn. Nadat  $\alpha$  en  $\beta$  zijn bepaald, is een voorspelling voor  $y(t)$  voor  $h$ -periodes vooruit te berekenen als:

$$y(t+h) = \alpha + \beta(t+h).$$

Dit betekent dat de verandering van  $y(t+h)$  ten opzichte van  $y(t+h-1)$  gelijk is aan  $\beta$ .

Een alternatief model om een trend weer te geven is het zogenaamde geïntegreerde trend-model (IT-model). Een eenvoudig voorbeeld hiervan is:

$$y(t) = y(t-1) + \mu.$$

Als deze  $\mu$  bekend is, dan kan een voorspelling voor  $h$ -periodes vooruit gemaakt worden als:

$$y(t+h) = y(t) + h\mu \text{ en geldt dat } y(t+h) - y(t+h-1) \text{ gelijk is aan } \mu.$$

In kader 1 worden de eigenschappen van deze twee trend-modellen verder uitgewerkt. Op basis van deze eigenschappen kan men de beide modellen, althans theoretisch, van elkaar onderscheiden.

In de praktijk moet men vaak op basis van waargenomen variabelen in het verleden, dus  $y(0), y(1), \dots, y(t-1), y(t)$ , een beslissing maken of het DT-model of het IT-model het beste is. In de praktijk liggen de waarden van  $\mu$  en  $\beta$  vaak heel dicht bij elkaar en domineert de lange-termijntrend vaak over korte-termijnfluctuaties in een variabele; het is dan ook begrijpelijk dat het moeilijk is een





Foto: Benelux Press

Soms heeft men behoefte aan projecties over de lange termijn, bijvoorbeeld over een tiental jaren. Een concreet voorbeeld hiervan is een voorspelling van het autopark in 2010 die nodig is om plannen voor de aanleg van wegen te maken.

### Kader 1

## Onderscheidende eigenschappen van DT-model en IT-model

Om precies aan te geven welke de essentiële verschillen zijn tussen het DT- en het IT-model beschouwen we de opbouw van de variabele  $y(t)$  in ieder model vanaf tijdstip 0. We introduceren ook een storings-term  $u(t)$  met een gemiddelde 0.

#### Voor het DT-model:

$$y(t) = \alpha + \beta t + u(t)$$

geldt simpelweg dat

$$y(t) = y(0) + \beta t + u(t) - u(0)$$

waarbij  $y(0)$  staat voor de waarneming op tijdstip 0.

#### Voor het IT-model:

$$y(t) = y(t-1) + \mu + u(t)$$

geldt na herhaaldelijk substitueren (integreren):

$$y(t) = y(t-2) + \mu + u(t-1) + \mu + u(t) =$$

$$= y(t-3) + \mu + u(t-2) + \mu + u(t-1) + \mu + u(t) \text{ enz.}$$

dat

$$y(t) = y(0) + \mu t + u(t) + u(t-1) + u(t-2) + \dots + u(1).$$

Het essentiële verschil tussen het DT-model en het IT-model is gegeven door de accumulatie van de variabele  $u(t)$ . Voor het DT-model geldt dat alleen  $u(t)$  van belang is op tijdstip  $t$ , terwijl voor het IT-model geldt dat alle  $u(t)$  tot op heden de waarde van  $y(t)$  bepalen. Dus alle oude storingen of afwijkingen worden verwerkt in de huidige waarde. Met ander woorden, schokken als gevolg van bijvoorbeeld beleidswijzigingen hebben een permanent effect op  $y(t)$  in het IT-model, terwijl ze een verdwijnend effect hebben bij het DT-model.

keuze te maken tussen het DT- en het IT-model. De keuze voor een van beide modellen geeft echter een andere kijk op een economische variabele als die men zou hebben wanneer men het andere model had verkozen. Dit blijkt uit *kader 1*.

Van belang voor de marketeer die zich baseert op onderzoek van derden is de vraag welke van de twee modellen het beste buiten de steekproef voorspelt.

### Geïntegreerd model de beste voorspeller

Recentelijk is een uitvoerige studie uitgevoerd naar de relatieve voorspelkwaliteit van de beide modellen, zonder dat van tevoren vaststond welk model binnen de steekproef het beste was. De conclusie van die studie is dat het IT-model in het algemeen het beste voorspelt. Deze conclusie geldt voor verschillende voorspelhorizonten en voor verschillende tijdstippen, waarop de voorspellingen worden gemaakt. De conclusie blijkt dus robuust voor de steekproefgrootte. Een belangrijke oorzaak van de goede voorspelkwaliteit van het IT-model is dat de voorspellingen in zekere mate bestand zijn tegen structuurbreuken in de voorspelperiode. In plaats van veel inspanning te steken in het selectieproces van het DT-model versus het IT-model kan men dus beter de vuistregel hantieren dat in geval van de veel voorkomende twijfel tussen DT en IT, men beter het IT-model kan kiezen.

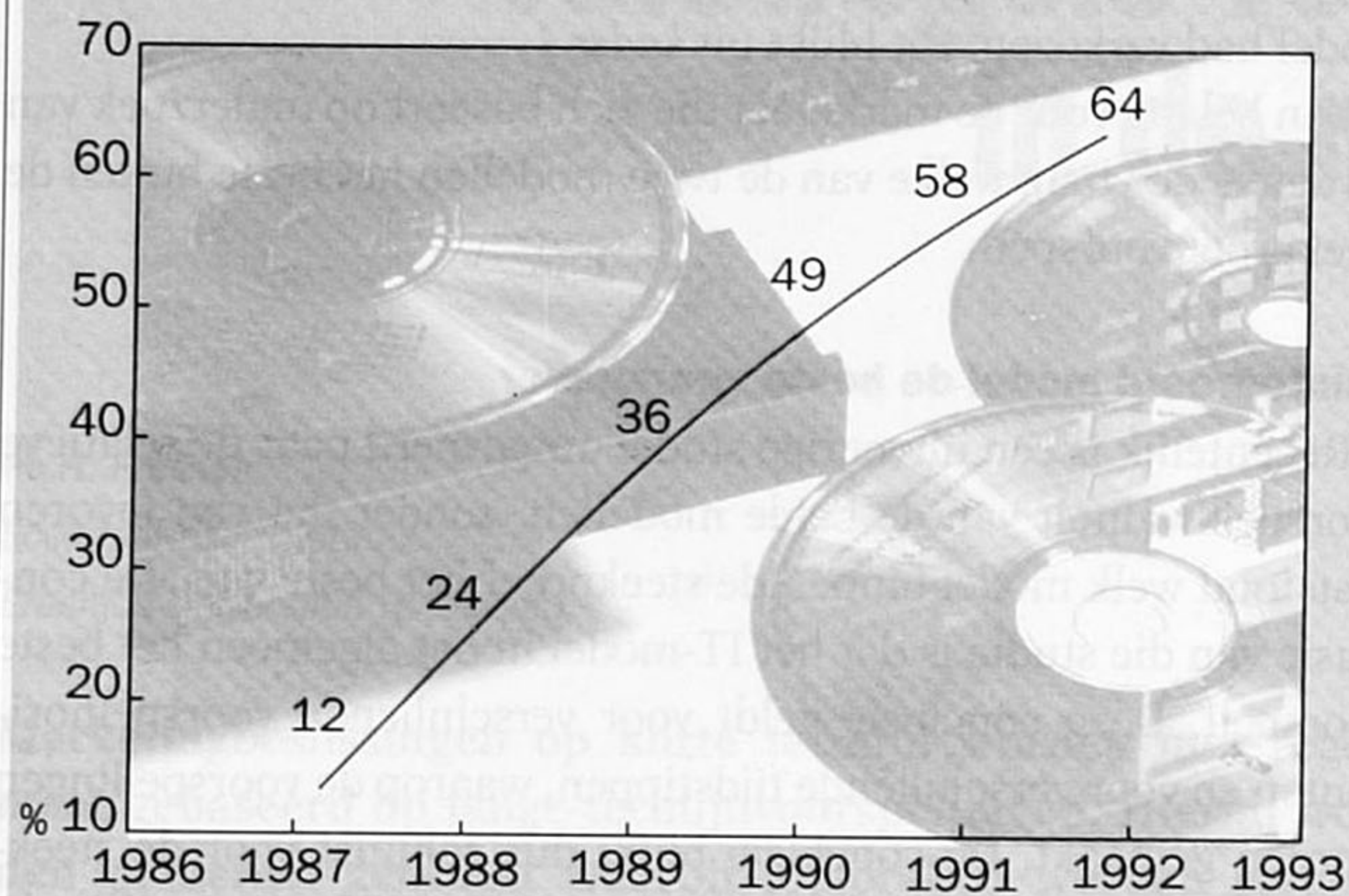
### Voorspellingen met verzadigingsniveau

Lange-termijnvoorspellingen zijn dikwijls nodig bij het voorspellen van verkopen of cumulatieve verkopen, waarbij men een zeker verzadigingsniveau kan verwachten. Zulke lange-termijnvoorspellingen worden vaak gebaseerd op zogenoemde groeicurves. Deze groeicurves nemen aan dat de verkopen of de cumulatieve verkopen alleen afhangen van de tijd. Dit reduceert de afhankelijkheid van allerlei grootheden, die op de korte termijn wellicht van belang zijn, maar die op de lange termijn alleen maar grote onzekerheid rondom voorspellingen genereren. De eenvoudigste groeicurve is de lineaire trendcurve, ofwel het DT-model. Deze lineaire curve geeft voorspellingen waarin een almaar stijgende lijn zit. In veel marketingsituaties is dit wellicht niet het patroon dat correspondeert met eerdere ervaring met andere produkten of met de produktiecapaciteiten. Het ligt vaak meer in de lijn der verwachting dat de groei in de verkopen zal afzwakken naarmate er meer verkopen zijn. Met ander woorden, de verkopen en de cumulatieve verkopen convergeren naar alle waarschijnlijkheid naar een zeker verzadigingsniveau. Een voorbeeld van een groeicurve is de Gompertz-curve (zie *kader 2*).

Bij het gebruik van de Gompertz-groeicurve stelt men het verzadigingsniveau gelijk aan 1 of aan 100%. Dit betekent bijvoorbeeld dat op de lange termijn iedereen een CD-speler heeft. Deze veronderstelling is in veel gevallen in de marketingpraktijk niet realistisch. Door de introductie van concurrerende of nieuwe produkten zal het maximale verzadigingsniveau niet vaak worden gehaald.

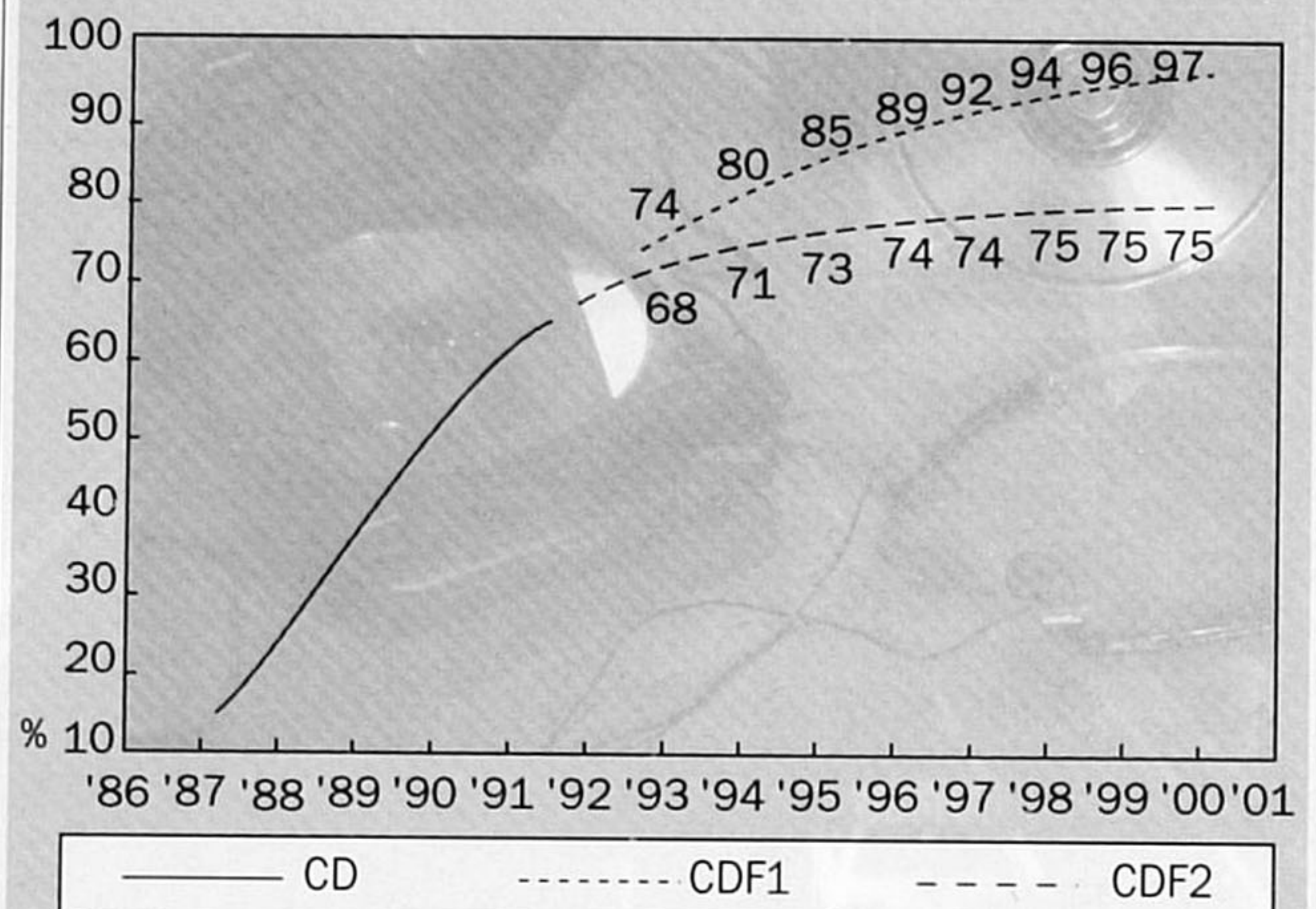
Een methode om de parameters in een Gompertz-model te bepalen, waarbij niet van tevoren wordt opgelegd dat het verzadigingsniveau gelijk is aan 1 of 100%, staat beschreven in *kader 3*. Het ligt voor de hand dat de waarde van het verzadigingsniveau in de lange-termijnvoorspellingen domineert. Het is duidelijk dat een te hoog verzadigingsniveau de toekomstvoorspelling opwaarts beïnvloedt. Dit kan te optimistische verwachtingen bij de marketeer teweegbrengen, waardoor marketinginstrumenten onjuist kunnen worden aangewend. Het is dus van groot belang om voor dit verzadigingsniveau een geschikte waarde te hanteren. Het lijkt dan ook niet zinvol om van tevoren vast te leggen dat dit niveau gelijk is aan 1 of aan 100%. Beter is het om een methode te gebruiken



**Figuur 1. Penetratie CD-speler, 1987-1992**

Het percentage van de Nederlandse huishoudens dat een CD-speler bezit, gemeten over de periode 1987-1992.

Bron: Cijferschrift 1992, NVPI, Hilversum.

**Figuur 2. Penetratie CD-speler, 1987-2000**

Voorspellingen van het percentage van de Nederlandse huishoudens dat een CD-speler bezit over de periode 1993-2000. De voorspellingen zijn gegenereerd uit een model dat een 100%-verzadigingsniveau veronderstelt (CDF1) en een model dat geen veronderstellingen maakt voor dit niveau (CDF2).

die de waarde van het verzadigingsniveau mede bepaalt. Als daaruit blijkt dat dit inderdaad 1 of 100 is, dan kan men die waarde in het vervolg hanteren.

### Penetratie van de CD-speler

De penetratie van CD-spelers is een typisch voorbeeld van een grootheid die naar alle waarschijnlijkheid niet zal convergeren naar 100%. In *figuur 1* staan zes waarnemingen van de penetratie van de CD-speler in Nederlandse huishoudens van 1987-1992 (Cijferschrift 1992, uitgegeven door het NVPI in Hilversum). Een stijgende lijn in de penetratie is duidelijk waarneembaar. Tevens lijkt deze stijgende lijn iets af te zwakken na 1990. Nieuwe produk-

ten zullen in de toekomst de CD als muziekdrager gaan vervangen. Verder ligt het in de lijn der verwachting dat de afbuiging naar een verzadigingsniveau zich reeds voordoet. Een stabilisatie van de penetratiegraad is ook wat de experts van het NVPI verwachten (Cijferschrift 1992, p. 7). Het is dan ook interessant om voorspellingen te maken met behulp van een Gompertz-groei-model en te zien of het vastleggen van het verzadigingsniveau realistische voorspellingen geeft. De methode die aanneemt dat het verzadigingsniveau 100% is, heeft als lange-termijnvoorspelmodel (voor meer details zie *kader 3*):

$$y(t) = 100 \cdot \exp(-1.98 \exp(-0.32t))$$

### Kader 2

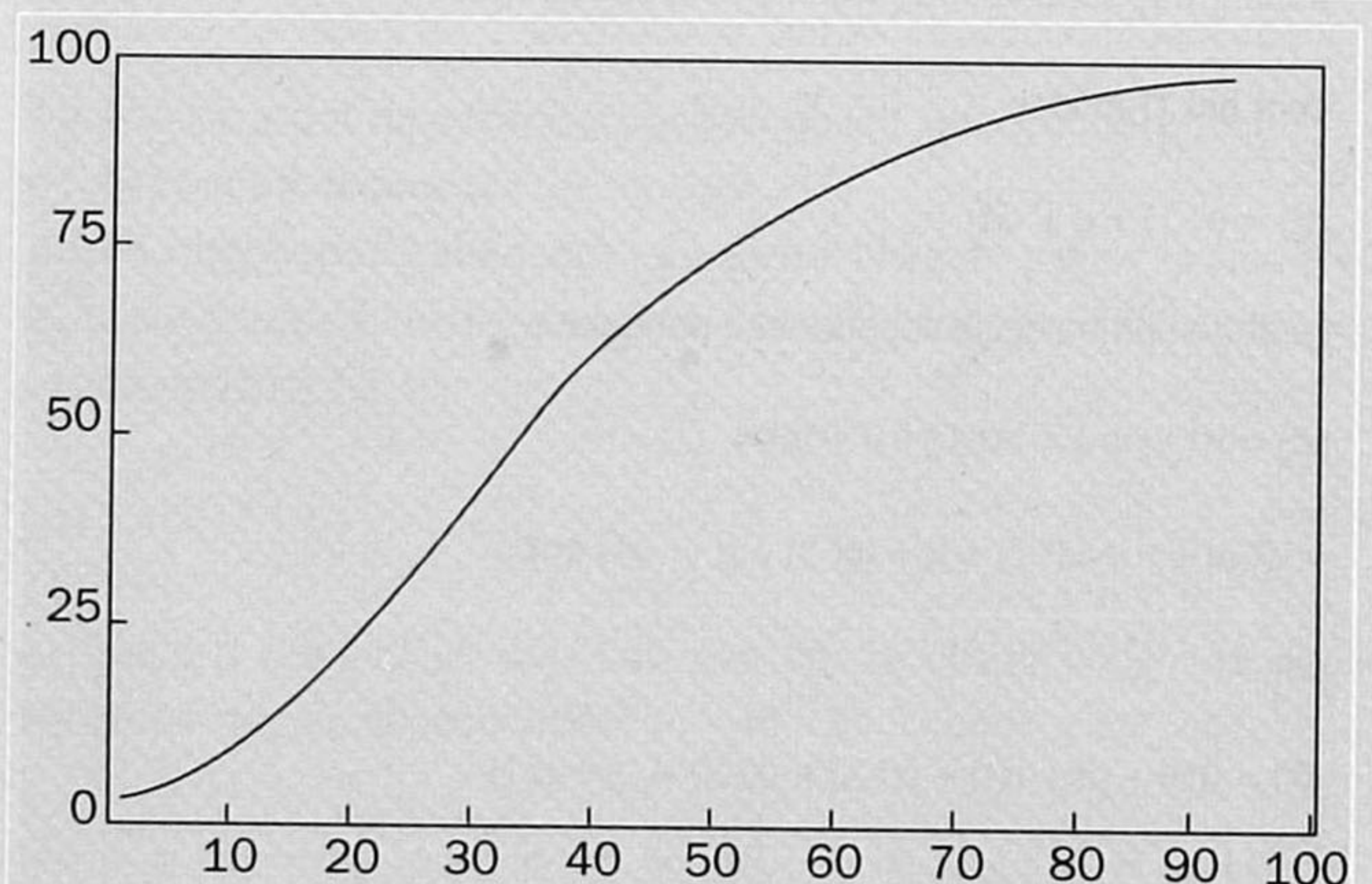
#### De Gompertz-groei-curve

De Gompertz-groei-curve wordt weergegeven door:

$$y(t) = a \cdot \exp(-b \cdot \exp(-ct))$$

waarbij  $y(t)$  de variabele van belang is,  $t$  geeft een lineaire trend weer. Er zijn nu drie parameters onbekend:  $a$ ,  $b$  en  $c$ . De waarde van  $a$  geeft het verzadigingsniveau aan. Dit niveau fungeert als een bovengrens want het zal nooit worden bereikt. Immers, als  $t$  almaar groter wordt, dan komt de waarde van de term  $\exp(-b \cdot \exp(-ct))$  steeds dichterbij 1 te liggen, maar zal nooit gelijk aan 1 worden. In de figuur hiernaast is een voorbeeld gegeven van een variabele die gegenereerd is volgens deze groei-curve. Er zijn natuurlijk nog meerdere typen groei-curves, zoals bijvoorbeeld de logistische groei-curve.

Het probleem bij het hanteren van de Gompertz-groei-curve is dat de drie parameters  $a$ ,  $b$  en  $c$  moeten worden geschat op basis van één verzameling waarnemingen. Een veel gebruikte procedure is om de waarde van  $a$ , dit is het verzadigingsniveau, vast te leggen.



Een voorbeeld van een Gompertz-groei-curve met verzadigingsniveau 100. Een typisch aspect van een Gompertz-curve is dat het proces eerst flink groeit en na het buigpunt minder sterk groeit tot vlak bij het verzadigingsniveau. Het is dus een voorbeeld van een asymmetrische groei-curve.

Bij voorspellingen van het marktaandeel zet men deze  $a$  gelijk aan 1 of 100%. In *kader 3* is aangegeven hoe men met een bekende  $a$  de overige twee parameters kan bepalen.



## Kader 3

**Methode om parameters in Gompertz-model te bepalen**

De standaardmethode om de parameters in een Gompertz-model te schatten gaat ervan uit dat de waarde van het verzadigingsniveau,  $a$  in het model

$$y(t) = a \cdot \exp(-b \cdot \exp(-ct))$$

bekend is. Als dat zo is, dan kan het Gompertz-model worden geschreven als:

$$y(t)/a = \exp(-b \cdot \exp(-ct)).$$

Het aan beide kanten nemen van natuurlijke logaritmes levert op:

$$\log(y(t)/a) = -b \cdot \exp(-ct).$$

Het minteken naar de andere kant halen en nog een keer logaritmes nemen geeft:

$$\log(-\log(y(t)/a)) = \log b - ct.$$

Deze laatste uitdrukking is een lineaire vergelijking. Dit betekent dat standaardtechnieken als kleinste kwadraten kunnen worden gebruikt om waarden voor  $b$  en  $c$  te bepalen.

Een alternatieve methode, die er vanuit gaat dat  $a$  onbekend is en dat het verzadigingsniveau zelf ook geschat moet worden, heeft als uitgangspunt dat het Gompertz-model wordt herschreven als:

$$\log y(t) = \log a - b \cdot \exp(-ct).$$

Een volgende stap is om  $\log y(t-1)$  van deze vergelijking af te trekken en daarna weer logaritmes nemen, ofwel

$$\log(\log y(t) - \log y(t-1)) = \log(b \cdot \exp(-b)) - ct.$$

De parameters in deze lineaire vergelijking,  $b$  en  $c$ , kunnen nu worden geschat met behulp van niet-lineaire kleinste kwadraten. Een schatting van  $a$  kan worden gemaakt door de gevonden waarden van  $b$  en  $c$  in te vullen in de originele vergelijking van het Gompertz-model.

De voorspellingen uit dit model voor de periode 1993-2000 zijn weergegeven in *figuur 2*. Het is duidelijk dat deze voorspellingen erg hoog liggen. Sterker nog de voorspelling voor 1993 is 74, hetgeen betekent dat de eerder gesignaleerde afbuiging niet doorzet en dat er (gezien vanuit 1992) weer een periode van flinke groei aankomt. Dit alles lijkt niet erg realistisch. De methode die de waarde van het verzadigingsniveau vrijlaat en in feite mede bepaalt, levert het Gompertz-model:

$$y(t) = 76 \cdot \exp(-1.86 \exp(-0.48t)).$$

Het verzadigingsniveau van de penetratie van de CD-speler wordt hier geschat op 76%. De voorspelling voor 1993 is 68%; deze waarde lijkt een stuk realistischer. Verder zien we dat de groei snel afneemt na 1995 en dat het verzadigingsniveau in 2000 zo goed als bereikt is.

**Vuistregels**

Twee eenvoudige vuistregels bij het maken van lange-termijnvoorspellingen in de marketing en bij het evalueren van marktonderzoeksrapportages zijn:

1. Bij onzekerheid over de keuze van een geschikt model, dat een lange-termijntrend beschrijft, is het model waarin een variabele wordt beschreven in termen van veranderingen te prefereren boven een model dat een perfect voorspelbare trend veronderstelt.
2. Het van te voren vaststellen van een verzadigingsniveau van bijvoorbeeld 100% geeft ongeschikte, vaak te optimistische voorspellingen. Tevens is een eenvoudige methode gegeven om het verzadigingsniveau mede te bepalen.

Onjuiste voorspellingen kunnen leiden tot het verkeerd aanwenden van marketinginspanningen, omdat beleid dan gebaseerd is op incorrecte projecties. Het is daarom van belang om inzicht te verkrijgen in de wijze waarop lange-termijnvoorspellingen worden ge-

maakt en bijvoorbeeld na te gaan hoe commerciële bureaus te werk zijn gegaan bij het maken van hun lange-termijnprojecties. Het kan zijn dat de marketeer een veel te optimistische toekomst wordt geschetst, terwijl dit met eenvoudige vuistregels vermeden had kunnen worden. TM



**Dr. Ph.H. Franses** is verbonden aan het Econometrisch Instituut van de Erasmus Universiteit Rotterdam als onderzoeker van de Koninklijke Nederlandse Akademie van Wetenschappen. Zijn onderzoek concentreert zich op modelkeuze in econometrische modellen met een nadruk op voorspellen.

**Noten**

1. *Unit roots in the Nelson-Plosser data: Do they matter for forecasting?* door Ph.H. Franses. Een exemplaar is op te vragen bij de auteur, Econometrisch Instituut, Erasmus Universiteit Rotterdam, Postbus 1738, 3000 DR Rotterdam.
2. Als twee variabelen beschreven kunnen worden met een IT-model, maar een lineaire combinatie beschreven kan worden met een model zonder trend, dan heten die beide variabelen gentergreerd. Een toepassing in de marketing van dit concept wordt gegeven in 'Modeling new product sales: An application of cointegration analysis' door Ph.H. Franses. Deze studie verschijnt binnenkort in de *International Journal of Research in Marketing*.
3. Details van deze methode staan beschreven in Ph.H. Franses, 'Fitting a Gompertz Curve', *Journal of the Operational Research Society*, 45 (1994), p. 109-113.