

DE WISKUNDIGE ECONOMIE IN TILBURG EN NEDERLAND EN DE WISKUNDIGE ECONOMIE VAN PIETER H.M. RUYS

Claus Weddepohl*

Universiteit van Amsterdam

1 Tilburg, het begin

Halverwege de jaren vijftig slaagde Professor Jan Dalmulder er in, binnen de economische faculteit van de toenmalige Katholieke Economische Hogeschool Tilburg, een studierichting econometrie ingevoerd te krijgen. De studierichting omvatte nog specialisaties voor "Algemene Econometrie" en voor "Bedrijfseconometrie". Soortgelijke ontwikkelingen deden zich aan de andere economische faculteiten voor. Dalmulder had, net als zijn naaste collega's Theil in Rotterdam, de Wolff in Amsterdam en Rijken van Olst in Groningen, een brede leeropdracht, die naast Econometrie in ruime zin (inclusief de Wiskundige Economie), ook de Mathematische Statistiek omvatte.

Een heel klein groepje studenten meldde zich in die begintijd voor de specialisatie Algemene Econometrie waaronder in opeenvolgende jaren Ad Pikkemaat, schrijver dezes en Pieter Ruys. Bij Dalmulder volgden we colleges Econometrie en Mathematische Statistiek. In afwijking van wat toen gebruikelijk was bij de economische opleidingen, werden die colleges gegeven uit recente Amerikaanse boeken. Samen met de professor probeerden we de gepresenteerde theorieën te begrijpen, want het was ook voor de docent nog geen gesneden koek. Dalmulder was eigenlijk het meest thuis in de Wiskundige Economie, maar voor het doceren van dat vak had hij geen tijd, want hij gaf ook nog "Bedrijfseconometrie", c.q. Lineair programmeren.

In het begin van de jaren zestig studeerden wij af. Dat was in de tijd dat de aantallen studenten in het wetenschappelijk onderwijs in korte tijd verveelvoudigden en waarin het overheidsbeleid een ongekende groei van de personeelsformaties van de universiteiten mogelijk maakte. Een gouden tijd dus. Vele van de toen aangestelden zijn aan de universiteit gebleven en die zijn nu, net als wij, allemaal aan hun pensioen toe.

* *Adres:* Afdeling Kwantitatieve Economie, Universiteit van Amsterdam, Roetersstraat 11, 1018 WB Amsterdam. *Email:* H.N.Weddepohl@uva.nl.

We werden dus uitgenodigd om als assistent bij Dalmulder te komen werken. Hij wilde het onderzoek, waaraan hij zelf nauwelijks was toegekomen na zijn proefschrift bij Tinbergen, op poten zetten. Wij mochten zelf bepalen waar we ons mee bezig zouden houden. Voor mij en ook voor Ad Pikkemaat en Pieter Ruys betekende dat de Wiskundige Economie, terwijl enkele anderen zich concentreerden op Econometrie en Statistiek. Wij stortten ons op de recente en de klassieke literatuur op ons vakgebied, met in ons achterhoofd dat er ooit een proefschrift zou moeten komen, maar dat was voorlopig nog niet aan de orde. Dalmulder vond dat we eerst internationale ervaring moesten opdoen. En zo ging Pikkemaat eerst naar Parijs en toen naar de VS, ik studeerde een jaar in Parijs en Ruys verbleef een jaar in Rochester bij Lionel McKenzie, een van de founding fathers van de algemene evenwichtstheorie.

Daarna konden we serieus gaan denken om zelf iets aan de wetenschap te gaan doen. Een extra stimulans daartoe kregen we in het najaar van 1969. Gérard Debreu bracht toen zijn sabbatical door bij het CORE in Leuven en hij had zich bereid verklaard een serie colleges te geven in Tilburg. In vijf sessies verklaarde hij ons de algemene evenwichtstheorie, zoals door hem geformuleerd in zijn meesterwerk "Theory of Value" (Debreu, 1959).

Leuven was ook verder een belangrijke bron van inspiratie. Het CORE had zich, door de inspanningen van zijn oprichter Jacques Drèze, ontwikkeld tot het Europese centrum voor economische theorie. Er kwamen daar geleerden van over de hele wereld voor een korter of langer verblijf en die gaven daar een seminar op maandag over de hot topics in de Wiskundige Economie. Eens in de veertien dagen of zo tuften we, meestal in mijn eend, naar Leuven, om het seminar bij te wonen. Zo maakten we meteen kennis met talloze collega's op ons vakgebied.

Aan het eind van de zestiger jaren was Ad Pikkemaat gepromoveerd op een proefschrift over informatie in de economie (Pikkemaat, 1969) en vertrokken naar Groningen, bevond mijn proefschrift over keuzemodellen en consumptietheorie (Weddepohl, 1971) zich in zijn eindfase en werkte Ruys aan de theorie van de publieke goederen (Ruys, 1974a).

2 Wiskundige Economie in Nederland

Aan het begin van de zeventiger jaren was in Nederland de moderne Wiskundige Economie, met toen als centrale thema de algemene evenwichtstheorie, alleen in Tilburg en in Groningen onderwerp van onderzoek en onderwijs.

In Rotterdam was men meer bezig met empirische toepassingen van de Wiskundige Economie, vooral op het terrein van het consumentengedrag. Daaraan werkte eerst Theil, naast zijn zuiver econometrische werk, samen met o.a. zijn promovendus Anton Barten, die later hoogleraar werd in Leuven, en later de hoogleraar Wiskundige Economie Somermeyer met zijn medewerker Jan van Daal, en Somermeyers opvolger Bernard van Praag. Uiteindelijk werd Sanjeev Goyal de hoogleraar Wiskundige Economie in Rotterdam. Intussen had Jan van Daal zich ontwikkeld tot een internationaal erkend deskundige op het terrein van het werk van Walras, de grondlegger van de algemene evenwichtstheorie.

Ook aan de beide Amsterdamse universiteiten was men vooral empirisch bezig:

Piet de Wolff aan de UvA en later diens promovendus Nol Merkies aan de VU. Na het vervoegde emeritaat van de Wolff werd de Wiskundige Economie achtereenvolgens door enkele buitenlandse geleerden verzorgd (von Weiszcker, Levhari, Phelps) en in 1980 nam ik het over. Tegelijkertijd volgde Pieter Ruys Dalmulder op in zijn functie van hoogleraar Wiskundige Economie. Merkies had Gerard van der Laan en Dolf Talman aangetrokken als promotiemedewerkers voor de Wiskundige Economie. Zij ontwierpen algoritmen voor het berekenen van evenwichten, waarover beiden promoveerden (Talman, 1980, van der Laan, 1980) en zij werden specialisten op het terrein van evenwichts- en onevenwichtstheorie. Van der Laan werd later hoogleraar aan de VU en Talman versterkte de groep wiskundig economen in Tilburg, na enige tijd ook als hoogleraar.

En om het plaatje compleet te maken: Talman's promovendus Jean-Jacques Herings werd, kort na zijn promotie in 1995, hoogleraar Economie in Maastricht, maar hij is overduidelijk een representant van de Wiskundige Economie. Daarnaast zijn er nu veel meer wiskundig economen werkzaam aan de Nederlandse universiteiten en andere instellingen, waarvan de meesten promovendi zijn van de hiervoor genoemde hoogleraren. Helaas verdween de Wiskundige Economie met het vroege emeritaat van Pikkemaat uit Groningen: hij werd niet opgevolgd en zijn medewerkers werden verspreid over andere leerstoelen.

Daarbij moet worden opgemerkt dat het gebruik van wiskundige modellen nu algemeen gangbaar is onder economen, anders dan toen wij begonnen. Maar er wordt soms slordig met de theorie omgesprongen en men blijft vaak steken in voorbeelden.

3 De wetenschappelijke bijdragen van Pieter Ruys

Evenwicht, publieke goederen en dualiteit

Pieter Ruys en ik hadden een gezamenlijke liefhebberij ontwikkeld in de convexe analyse en de daarmee verband houdende dualiteitstheorie. Deze laatste komt, kort gezegd, hierop neer dat formele modellen in termen van goederenhoeveelheden, kunnen worden vertaald naar soortgelijke formele modellen in termen van goederenprijzen, terwijl "oplossingen" in beide modellen met elkaar corresponderen.

Dat leidt tot een bijzonder elegante manier om de, oorspronkelijk door Lindahl ontworpen theorie van de publieke goederen te modelleren, zoals in Ruys (1974a) wordt uitgelegd. Dit proefschrift handelt over allocatieproblemen en de wijze waarop optimale allocaties via gecentraliseerde en gedecentraliseerde procedures kunnen worden bereikt, met de nadruk op de allocatie van *publieke* goederen. Kenmerk van zogenaamde *private* goederen is, dat de consumptie daarvan exclusief is voor elke consument, dus wat de een heeft, heeft de ander niet. De totale beschikbare hoeveelheid moet dus worden verdeeld onder de consumenten. "Zuiver" publieke goederen daarentegen, worden gezamenlijk door alle consumenten gebruikt, zodat ze allemaal precies dezelfde hoeveelheden tot hun beschikking hebben.

In een evenwicht in een economie met alleen private goederen betaalt iedereen voor deze goederen dezelfde prijs. Maar welke prijzen zouden er idealiter betaald moeten worden voor het gebruik van de gemeenschappelijke hoeveelheden van de publieke goederen? Lindahl had laten zien dat iedereen dan zou moeten betalen

overeenkomstig zijn betalingsbereidheid: hij zou de prijs betalen, die hij er voor over heeft, dat is de prijs waarbij hij precies de gegeven hoeveelheid zou kiezen, als hij zou kunnen kiezen, zodat de bekende gelijkheid van de verhouding van marginale nutten en van prijzen blijft gelden. Die voor de verschillende consumenten verschillende prijzen moeten dan optellen tot de kostprijs. Om deze oplossing te realiseren moet ieders betalingsbereidheid bekend worden, maar consumenten zullen niet geneigd zijn deze te tonen door hun schaduwprijs correct door te geven: ze laten het betalen liever aan de anderen over en gedragen zich zo als *free rider*.

Ruys paste nu de dualiteitstheorie toe en formuleerde de publieke goederen economie in de prijzenruimte. En nu komt een evenwicht met (uitsluitend) publieke goederen in de prijzenruimte formeel op precies hetzelfde neer als een evenwicht met private goederen in de goederenruimte: in het laatste geval zijn de betaalde prijzen identiek en de hoeveelheden tellen op tot een totaal, in het eerste geval zijn de hoeveelheden identiek en de prijzen tellen op tot een totaal. De hoeveelheden private goederen vormen een optimum bij gegeven prijzen, de prijzen van publieke goederen vormen een optimum bij gegeven hoeveelheden. (Dit resultaat werd eerder gepubliceerd in Ruys (1972)). Deze methode is evenwel ook toepasbaar en zeer verhelderend in een economie met zowel publieke als private goederen, zoals in het proefschrift uiteengezet. En verder wordt daarin de organisatie van het allocatieproces behandeld, mede omvattend procedures om de "ware" preferenties voor de publieke goederen te achterhalen.

Voor zijn bijzonder fraaie proefschrift, waarop hij cum laude promoveerde, ontving Ruys de Kluwer prijs.

De theorie over de publieke goederen stond in de zeventiger jaren sterk in de belangstelling, maar is de laatste tijd min of meer vergeten, waarschijnlijk omdat geen praktisch realiseerbare methoden zijn gevonden om het free rider probleem op te lossen. Toch valt dit te betreuren, omdat deze theorie beslist zeer verhelderend is.

Na zijn promotie is Ruys eerst nog even verder gegaan met de convexe analyse en dualiteit. In Ruys (1974b) werkte hij dualiteit nader uit voor de productietheorie en in Ruys en Weddepohl (1979) vatten we nog eens samen wat we gezamenlijk over dualiteit en de economische toepassingen daarvan op consumptie- en productietheorie en evenwicht, wisten. Daarna hielden we allebei dit onderwerp verder voor gezien.

De theorie van de onevenwichtigheid, die zijn oorsprong vond in het werk van Drèze (1975) en Benassy (1975) e.a., is het onderwerp van Ruys (1982). Het omvat een uitwerking van het idee om in geval van onevenwichtigheid op markten (ongelijkheid van vraag en aanbod bij vaste prijzen), een evenwicht niet door een rantsoeneringssysteem te laten ondersteunen, zoals de hiervoor genoemde auteurs voorstelden, maar door "effort". In geval van onevenwicht op de markt, zullen de vragers of aanbieders aan de lange kant van de markt (waar dus een tekort heerst) zich moeten inspannen om een zo groot mogelijk aandeel te verwerven van de vraag of het aanbod op de korte kant van de markt, in die zin, dat men om meer te krijgen meer inspanning moet leveren. Naast de normale prijs in termen van de numéraire die men betaalt of ontvangt, betaalt men ook een "effort"-prijs in termen van inspanning, wat een nutsoffer betekent. De auteur beschouwt dit als een

zelf organiserend mechanisme om tot een evenwicht te geraken, dat geen centrale verdelende instantie eist.

Interactie

De bij het schrijven van zijn proefschrift over publieke goederen verworven inzichten, hebben in belangrijke mate richting gegeven aan Ruys' verdere werk. Dit komt tot uiting in zijn opvatting over wat de kern is van de economische theorie, zoals verwoord in zijn Tilburgse oratie (Ruys, 1981), waarin hij zegt dat het daar "niet zonder meer gaat om het bestuderen van het keuze-gedrag van mensen in een economie: het gaat veeleer om een georiënteerd en gecoördineerd gedrag, dat een sociale betekenis heeft en waar mensen niet geheel vrij zijn in het kiezen van een gedragscode". Sleutelbegrippen in zijn denken zijn "communicatie", "relaties", "netwerken", maar bovenal "interactie", want interacties tussen mensen spelen voortdurend mee bij de voortbrenging, de consumptie en de allocatie van goederen en diensten.

Interacties in algemene zin, voor zover ze zijn geïnstitutionaliseerd, worden bestudeerd in Ruys (1999). Hij gebruikt daar het zogenaamde "tripolaire model", een eenvoudige algebraïsche structuur in het platte vlak, waaraan hij zeer gehecht is en waaraan hij ook het laatste hoofdstuk in zijn proefschrift had gewijd. Aan de hand van dat model formuleert hij verschillende vormen van interactie en laat hij zien hoe die interacties binnen een spel, respectievelijk binnen een ruileconomie, in deze structuur passen.

In het proefschrift ging het over een economie met "zuiver" private goederen (slechts één consument profiteert van de hem toekomende hoeveelheid van een goed) en "zuiver" publieke goederen (alle consumenten in de economie profiteren van de totaal beschikbare hoeveelheid), waarbij hun waarderingen voor private zowel als publieke goederen verschillen. Een belangrijke conclusie van deze analyse is dat de voorziening van elk van beide typen goederen een volstrekt andere organisatie vereist. Voor zuiver private goederen voldoet een marktorganisatie, maar voor zuiver publieke goederen moet het in elk geval anders, en de optimale manier is te vinden in de theorie over Lindahl evenwichten, hoewel de implementatie daarvan problematisch blijft. Het behandelde model geeft een vollediger beeld van de echte economie dan het standaardmodel met alleen zuiver private goederen, waarvan in de economische theorie meestal wordt uitgegaan. Dat model hebben de meeste economen ook in hun hoofd, als ze over de wereld nadenken, waardoor ze nog al eens tot onverantwoorde conclusies komen.

Maar eigenlijk zijn we er met deze twee typen nog lang niet. Er zijn allerlei tussenvormen, die noch zuiver publiek, noch zuiver privaat zijn: min of meer private goederen waarvan er altijd medegebruikers zijn, of die alleen in kleinere groepen kunnen worden gebruikt, en ook min of meer publieke goederen die alleen lokaal of voor groepen van geïnteresseerden van belang zijn, enz. Voor de voorziening en de financiering van zulke goederen past noch de voorziening via de markt, noch door een algemene publieke organisatie. Een aantal bijdragen van Ruys gaan nu precies over de nadere classificatie van goederen en diensten, - naar de wijze van gebruik en de mate van interactie tussen de gebruikers, en - naar de wijze waarop hun voorziening kan worden georganiseerd, niet alleen via private bedrijven of centrale

overheden, maar ook via andere organisaties.

Daarvoor heeft Ruys, alleen en samen met anderen, waaronder enkele van zijn promovendi, modellen ontwikkeld. Die zijn gericht op het beter aankleden van de gebruikelijke economische concepten, die zeer (te) schematisch van karakter zijn en die vaak een te simpel beeld van economische verschijnselen geven. Het Walrasiaanse evenwichtsmodel blijft meestal het skelet, maar er wordt gepoogd dit skelet van meer vlees te voorzien, waardoor het meer op een "levend" organisme gaat lijken. Daarmee neemt Ruys een geheel eigen plaats in onder de Nederlandse wiskundig economen.

Direct aansluitend op het Walrasiaanse model is een studie samen met Gilles over *communicatiestructuren*. In het Walrasiaanse evenwicht in een economie met alleen private goederen, worden alle agenten geacht met elkaar te handelen, goederen onderling te ruilen of aan elkaar te verkopen. Een evenwicht is een toestand waarin ieder precies die transacties afsluit die hij de beste vindt bij de gegeven prijzen of ruilverhoudingen. Hoe dat precies gebeurt blijft meestal in het midden, maar er is een sociale structuur nodig, om de gewenste transacties te realiseren. Gilles en Ruys (1990) voorzien de economie niet alleen van preferenties en resources, maar daarnaast van een communicatiestructuur. Ze definiëren een *relationally structured population*, door middel van een reflexieve en symmetrische binaire relatie over tweetallen in de verzameling economische agenten, die betekent dat de twee elkaar "kennen", en die wordt uitgebreid met alle indirecte relaties (kennissen van kennissen, enz). Voor een goede werking van de economie wordt het nodig geacht dat ieder ieder ander kent, direct of indirect (*full communication property*). Om te garanderen dat aan deze eigenschap voldaan is, worden voorwaarden opgelegd, mede m.b.v. de individuele kenmerken van de agenten. Als b.v. alle agenten die bepaalde individuele kenmerken gemeen hebben elkaar kennen en, binnen elke twee groepen met verschillende kenmerken, bestaan er steeds twee die elkaar kennen, is aan deze eis voldaan.

Semi-publieke goederen

Een theorie over *semi-publieke goederen* vormt een verdere uitbouw van de theorie uit het proefschrift. Zuiver publieke goederen, waarvan iedereen gebruik kan maken zonder anderen te hinderen, vormen een extreem geval en niet het enige soort goed met een publiek karakter. Onder een semi-publiek goed wordt een algemeen gebruikte voorziening (infrastructuur, zoals een wegennet of een vliegveld) verstaan, die het karakter van een publiek goed heeft, en die nodig is voor de consumptie c.q. het gebruik van een of meer private goederen (voertuigen, vliegtuigen), waarbij de mogelijke omvang van dit gebruik afhangt af van de omvang van de infrastructuur.

Een model daarvoor wordt gepresenteerd in Ruys (1984). Voor de consumptie van de private goederen betaalt men in elk geval de prijs daarvan, plus een uniforme opslag voor het gebruik van het semi-publieke goed, onder de veronderstelling dat dit de betalingsbereidheid van de verschillende consumenten weergeeft. Alleen in zoverre deze toeslag niet genoeg oplevert om de infrastructuur te financieren, wordt nog een tweede prijs betaald door iedereen, overeenkomstig zijn eigen betalingsbereidheid (schaduwprijs) voor de infrastructuur als geheel. Alle goederen wor-

den door winstmaximaliserende bedrijven geproduceerd, maar de organisatie en de afhandeling van de infrastructuur wordt geregeld door een centrale planner.

Dit wordt nader uitgewerkt in Ruys en van der Laan (1987) voor een economie met semi-publieke goederen naast gewone private goederen. Een evenwicht wordt zorgvuldig gedefinieerd, met een wat ingewikkelder tariefstructuur voor de semi-publieke goederen. De gebruikers (consumenten en producenten) die de beschikbare infrastructuur als een beperking ervaren (en haar dus zouden willen zien uitgebreid) betalen een toeslag, overeenkomstig hun betalingsbereidheid, die dient om de infrastructuur geheel of gedeeltelijk te financieren. Om zo'n evenwicht te implementeren moet de betalingsbereidheid voor de toeslag dus kunnen worden waargenomen, en daarnaast moet de individuele directe bijdrage voor financiering van de infrastructuur (als voor het zuiver publieke goed) worden gevonden, waardoor tevens het niveau van de infrastructuur wordt bepaald. Die vaststelling van de betalingsbereidheid blijft natuurlijk een probleem. Het stuk wordt gecompleteerd met een aantal uitgewerkte numerieke voorbeelden.

In Ruys (1988) komt hij nog eens terug op de semi-publieke goederen, zij het dat het betoog hier is ingebed in een meer algemene beschouwing over de structuur van markten enerzijds en van bedrijfstakken anderzijds. Hij onderscheidt daar vrije en gereguleerde bedrijfstakken, met daartussen "georganiseerde" bedrijfstakken, waarvan het optimale functioneren wordt bevorderd door een al dan niet gouvernementele organisatie met regelende bevoegdheden. Op de markt onderscheidt hij dan private en publieke bedrijven, met daartussen "gespecialiseerde" bedrijven die aan een contract gebonden zijn, dat hun in hun beslissingen beperkt. Er zijn, als in Ruys en van der Laan (1987), semi-publieke en private goederen (nu semi-private goederen genoemd), die samen samengestelde (compound) goederen vormen. Een semi-privaat goed wordt door private bedrijven geleverd via de markt, een semi-publiek goed door een gespecialiseerde publieke instelling, van het hierboven beschreven intermediaire type, die haar investering zoveel mogelijk, liefst geheel, financiert uit de door de private bedrijven geheven opslagen op de prijs.

Gilles en Ruys (1994b) stellen een nog verfijndere indeling van de goederen voor. Op basis van de wijze waarop deze worden geconsumeerd zijn er drie typen goederen: alleen de gebruiker profiteert er van ("no interaction"), velen profiteren ervan, maar beïnvloeden elkaar niet ("external interaction") en velen profiteren ervan, maar de voordelen worden mede beïnvloed door het gebruik door de anderen ("social interaction"). In elk type goed kan privaat, publiek of in een netwerk worden voorzien, zodat er negen combinaties ontstaan. De eerder besproken zuiver private, zuiver publieke en gedeeltelijk ook de semi-publieke goederen (nu infra-publieke goederen genoemd), vormen de groep "no interaction", zodat er nog zes andere typen worden onderscheiden, zij het dat hier ook omgevingsfactoren onder vallen die meestal niet als goederen of diensten worden aangemerkt, zoals cultuur, ondernemingscultuur of persoonlijke hygiëne. En voor deze verschillende goederen worden verschillende organisatievormen bekeken, waarin de voorziening van goederen en diensten wordt gerealiseerd, en die passen in de hiervoor besproken classificatie, zoals de overheid, sociale netwerken, non-profit organisaties, politieke processen, etc.

Organisatiestructuren

Een aantal bijdragen betreft de interne organisatie van bedrijven en andere economische waarde scheppende instituties, die hun plaats vinden in de hiervoor besproken indelingen.

Het model van een bedrijf in Ruys en van den Brink (1999) en in Ruys, van den Brink en Semenov (2000) heeft aanzienlijk meer interne structuur dan gebruikelijk in de theorie, waar het bedrijf meestal is gereduceerd tot een productiefunctie of een productieverzameling. Hier wordt het bedrijf uitgerust met een hiërarchische bestuursstructuur, weergegeven door een *gerichte graaf*, waarvan de punten de personeelsleden voorstellen en de lijnen daartussen de geldende gezagsrelaties. Onderaan staan de "front workers", en bovenaan de eigenaar, met een variabel aantal lagen daar tussen. Elke tussenliggende laag omvat managers die ieder hetzelfde aantal agenten in de volgende laag onder zich hebben, die elk weer de top vormen van de volgende laag. Het aantal lagen groeit met de omvang van het bedrijf, c.q. het aantal front workers. De front workers realiseren de productie. De grootte daarvan wordt bepaald via een productiefunctie, met als argumenten het aantal front workers en het aantal lagen: met het aantal lagen nemen zowel de productiviteit van de front workers als de kosten toe. Voor de verschillende niveaus wordt een aan de speltheorie ontleende relatieve beloningsstructuur gehanteerd, die eerder was uitgewerkt in van den Brink (1994), en waar een functionaris op een hoger niveau een hogere beloning krijgt dan die daaronder. Gegeven deze beloningsstructuur en gegeven de productiefunctie zijn, bij exogeen gegeven minimumlonen, verschillende bedrijfsomvangs met bijbehorend aantal tussenlagen mogelijk, waarvan er één, die het voordeligst is voor de eigenaar, zal worden gerealiseerd. Dit geheel wordt dan ingebed in een competitieve markt en leidt tot een evenwicht dat een "Institutional Market Equilibrium" wordt genoemd.

In Ruys, van den Brink en Semenov (2000) wordt ook een gelijksoortige structuur voorgesteld voor *clubs*, dat zijn organisaties die hun leden van goederen of diensten voorzien. De beheersstructuur wordt door dezelfde gerichte graaf weergegeven als bij de bedrijven. Ook hier zijn er verschillende lagen, met steeds een topman ("officer") verbonden met een aantal agenten. De leden van de club, die de dienst vragen, vormen het laagste niveau. Alleen is de relatie tussen topman en agenten omgekeerd: elke laag controleert het hogere niveau (bottom up). Het laagste niveau verschaft de middelen die het hogere niveaus mogelijk maken hen de gewenste diensten te verschaffen. Elke hogere laag geeft weer instructies en middelen aan het naast hogere niveau.

In Ruys (2002) gaat het om een model voor een diensteneconomie. Een dienst ontstaat uit een interactie tussen twee of meer individuen. De organisatie waarbinnen dit wordt geregeld, wordt een waardescheppend netwerk genoemd, weergegeven door een gerichte graaf. Dat netwerk bestaat uit vele subnetwerken, onderscheiden naar netwerken van dienstverleners (performers) en van dienstenontvangers (receivers) voor elk soort dienst. Het dienstverlenersnetwerk heeft een hiërarchische structuur, als geïntroduceerd in Ruys en van den Brink (1999) voor een bedrijf: een aantal front performers, met daarboven managers en zo een aantal lagen, met een topmanager bovenaan (top down). Een ontvangersnetwerk lijkt op de club uit Ruys,

van den Brink en Semenov (2000): de ontvangers van de diensten (voters) geven instructies aan de managers, wederom georganiseerd in verschillende lagen, waarin elke manager de vertegenwoordiger is van degenen die hij coördineert (bottom up). Netwerken voor dezelfde dienst worden samengevoegd tot één netwerk, via de interactie tussen front performers and front receivers, de eigenlijke dienstverlening.

Binnen dit kader wordt een *service economy* gedefinieerd, waar de top managers van de performers netwerken hun inkomen maximaliseren. Dat leidt er toe dat voor de verschillende diensten aanbodfuncties ontstaan. De ontvangers worden in dit geval vragers naar de diensten met vraagfuncties. Omdat vragers hun eigen managers kunnen zijn, hebben ze geen netwerk meer nodig. In deze economie bestaat nu een algemeen evenwicht, waarbij alle markten ruimen.

De theorie van de core

Samen met Gilles en Haller hield Ruys zich ook bezig met de theorie van de core, waarover Gilles eerder bij hem was gepromoveerd (Gilles, 1990). De core is een speltheoretisch evenwichtsconcept dat, toegepast op een ruileconomie, allocaties *uitsluit*, waarvoor geldt dat een coalitie (deelverzameling) van consumenten een allocatie die voor hen beter is, kan realiseren, zonder medewerking van consumenten buiten de coalitie, zodat ze de voorliggende allocatie kunnen blokkeren. De core omvat daarmee die allocaties die niet kunnen worden geblokkeerd. Een bekende eigenschap is dat Walrasiaanse evenwichten tot de core horen en dat voor (volgens bepaalde criteria) "grote" economieën, de core en de verzameling Walrasiaanse evenwichten samenvallen, wat bekend staat als het "core equivalence theorem". Coalities zijn daarbij alle bestaande groepen van consumenten.

De vraag waar het in Gilles, Haller en Ruys (1994, 1998) om gaat, is in hoeverre een dergelijke eigenschap ook geldt als de verzameling toegelaten coalities slechts een meer beperkte groep deelverzamelingen omvat. Immers de mogelijke vorming van elke denkbare coalitie en het daarvoor aanwijzen van een blokkerende allocatie, eist de uitwisseling van zeer veel informatie. Meer beperkte en beter overzienbare groepen met een duidelijke interne structuur zouden daartoe wellicht ook in staat zijn. Men onderzoekt nu de allocaties in een ruileconomie, die tot stand kunnen worden gebracht of juist worden uitgesloten via coalities, die zijn gebaseerd op relaties die binnen groepen en tussen verschillende groepen bestaan, en op eigenschappen die verschillende individuen gemeenschappelijk hebben, zoals in de relationally structured population uit Gilles en Ruys (1990). Die groepen vormen dan gezamenlijk een netwerk, waarin de transacties tot stand kunnen komen. Onder bepaalde voorwaarden, m.n. wat betreft de omvang van het netwerk, blijft het core equivalence theorem geldig.

4 Slot

Hierboven heb ik aandacht besteed aan de wetenschappelijke publicaties van Ruys. Daarnaast heeft hij ook een aantal stukken geschreven die hij zelf omschrijft als "miscellaneous publications" die een meer toegepast karakter hebben en waarvan een aantal in minder toegankelijke media zijn verschenen. Zo gaat het in van der

Laan, Ruys en Stokman (1995) over de mogelijke inrichting van de drinkwatervoorziening in Nederland en over de mogelijkheden voor marktwerking. In Verbruggen, van der Laan en Ruys (1996) wordt het overheidsbeleid op het gebied van de energie, als uiteengezet in de Derde Energienota (1995) van het Ministerie van Economische zaken, bekritiseerd. Daarin wordt gesteld dat een groot aantal problemen bij de energievoorziening, samenhangend met het streven naar marktwerking, efficiëntie en bescherming van het milieu, onvoldoende worden onderkend.

Niet alleen door zijn eigen onderwijs en onderzoek, maar ook als begeleider van een groot aantal promovendi, heeft Ruys een belangrijke bijdrage geleverd aan de ontwikkeling van de Wiskundige Economie in Nederland. In de proefschriften van een aantal van hen zijn duidelijk de sporen van zijn denken terug te vinden. Anderzijds waren ze voor hem ook bronnen van inspiratie. Hij was ook de drijvende kracht achter de samenwerking tussen de wiskundig economen van de verschillende faculteiten, die gestalte kreeg in het project Competitie en Coöperatie, dat destijds was begonnen als v(oorwaardelijk) f(inancierings) project. De met het project verbonden driemaandelijks bijeenkomsten vormden en vormen nog steeds een stimulans voor jonge en oude onderzoekers.

Referenties

- Brink, J.R. van den (1994), *Relational Power in Hierarchical Organizations*, Proefschrift, Katholieke Universiteit Brabant, Tilburg.
- Benassy, J.P. (1975), “Neo-Keynesian disequilibrium in a monetary economy”, *Review of Economic Studies*, 42, 503-523.
- Debreu, G. (1959), *Theory of Value*, Wiley, New York.
- Drèze, J.H. (1975), “Existence of exchange equilibrium under price rigidities”, *International Economic Review*, 16, 301-320.
- Gilles, R.P. (1990), *Core and Equilibria of Socially Structured Economies: The Modelling of Social Constraints in Economic Behaviour*, Proefschrift, Katholieke Universiteit Brabant, Tilburg.
- Gilles, R.P., H.H. Haller and P.H.M. Ruys (1994), “Economies with relational constraints on coalition formation”, Ch. 5 in: R.P. Gilles and P.H.M. Ruys (1994a), 89-136.
- Gilles, R.P., H.H. Haller and P.H.M. Ruys (1998), “Semi-core equivalence”, *Economic Theory*, 11, 101-127.
- Gilles, R.P., and P.H.M. Ruys (1990), “Characterization of economic agents in arbitrary communication structures”, *Nieuw Archief voor Wiskunde*, 8, 325-345.
- Gilles, R.P., and P.H.M. Ruys (1994a), (eds.), *Imperfections and Behavior in Economic Organizations*, Theory and Decision Library, Series in Mathematical Economics and Game Theory, Kluwer Academic Publishers, Boston.

- Gilles, R.P., and P.H.M. Ruys (1994b), “Inherent imperfections of economic organizations”, Ch. 1 in R.P. Gilles and P.H.M. Ruys (1994a), 1–14.
- Laan, G. van der (1980), *Simplicial Fixed Point Algorithms*, Proefschrift, Vrije Universiteit, Amsterdam.
- Laan, G. van der, P.H.M. Ruys en F.N. Stokman (1995), “Water is goud waard”, *Economisch Statistische Berichten*, 80, nr. 4036, 1092–1097.
- Pikkemaat, G.F.W.M. (1969), *Informatie, Communicatie en Economische Organisatie*, Proefschrift, Tilburg.
- Ruys, P.H.M. (1972), “On the existence of an equilibrium for an economy with public goods only”, *Zeitschrift für Nationalökonomie*, 32, 189–202.
- Ruys, P.H.M. (1974a), *Public Goods and Decentralization: The Duality Approach in the Theory of Value*, Kluwer Academic Publishers, Dordrecht.
- Ruys, P.H.M. (1974b), “Production correspondences and convex algebra”, in: W. Eichhorn, R. Henn, O. Opitz and R.W. Shephard, (eds.), *Production Theory*, Springer Verlag, Berlin, 231–252.
- Ruys, P.H.M. (1981), “Rationeel gedrag”, *Maandschrift Economie*, 45, 266–280.
- Ruys, P.H.M. (1982), “Disequilibrium characterized by implicit prices in terms of effort”, in: J.E.J. Plasmans, (ed.), *Econometric Modelling in Theory and Practice*, The Hague, 111–129.
- Ruys, P.H.M. (1984), “Algemeen evenwicht met semi-publieke goederen”, in P.A. Verheyen e.a. (eds.), *De praktijk van de econometrie*, Stenfert Kroese, Leiden, 85–96.
- Ruys, P.H.M. (1988), “Industries with private and public enterprises”, *European Economic Review*, 32, 415–421.
- Ruys, P.H.M. (1999), “The role of the medium in an interaction structure”, in: H. de Swart, (ed.), *Logic, Game Theory, and Social Choice*, Tilburg University Press, Tilburg, 314–329.
- Ruys, P.H.M. (2002), “A general equilibrium of managed services”, in C. Hommes, R. Ramer and C.A. Withagen, (eds.), *Equilibrium, Markets and Dynamics, Essays in the Honour of Claus Weddepohl*, Springer, Berlin, 117–139.
- Ruys, P.H.M., and R. van den Brink (1999), “Positional abilities and rents on equilibrium wages and profits”, in: P.J.J. Herings, G. van der Laan and A.J.J. Talmán, (eds.), *The Theory of Markets*, North-Holland, Amsterdam.
- Ruys, P.H.M., R. van den Brink and R. Semenov (2000), “Values and governance systems”, in: Cl. Ménard (ed.), *Institutions, Contracts and Organizations: Perspectives from New Institutional Economics*, Edward Elgar, Cheltenham, 422–445.

- Ruys, P.H.M., and G. van der Laan (1987), “Computation of an industrial equilibrium”, in: A.J.J. Talman and G. van der Laan, (eds.), *The Computation and Modelling of Economic Equilibria*, North Holland, Amsterdam, 205–229.
- Ruys, P.H.M., and H.N. Weddepohl (1979), “Economic theory and duality”, in: J. Kriens, (ed.), *Convex Analysis and Mathematical Economics*, Springer, Berlin, 1–72.
- Talman, A.J.J. (1980), *Variable Dimension Fixed Point Algorithms and Triangulations*, Proefschrift, Vrije Universiteit, Amsterdam.
- Verbruggen, A., G. van der Laan en P.H.M. Ruys (1996), “Het wankelmoedig energiebeleid”, *Economisch Statistische Berichten*, 81, nr. 4052, 302–305.
- Weddepohl, H.N. (1971), *Axiomatic Choice Models and Duality*, Wolters Noordhoff, Groningen.

SOLVABILITEIT VAN PENSIOENFONDSEN

Jan H. Koning*

Grafische Bedrijfsfondsen (GBF)

De solvabiliteit van pensioenfondsen geniet sinds geruime tijd grote belangstelling. Voor een pensioenfonds kan solvabiliteit worden omschreven als de mate waarin de nakoming van (alle) toekomstige verplichtingen minder of meer waarschijnlijk is bij redelijke verwachtingen ten aanzien van de toekomst.¹ Voor de beoordeling van de solvabiliteit van een pensioenfonds is de omvang van het belegde vermogen ten opzichte van de verplichtingen van het fonds van belang alsmede de risico's waarmee het fonds kan worden geconfronteerd.

Hierna wordt uitgegaan van de onderstaande, sterk vereenvoudigde balans van een pensioenfonds.

Activa		Passiva	
Beleggingen	B	Voorziening pensioenverplichtingen	VPV
• Zakelijke waarden		Eigen vermogen	EV
• Vastrentende waarden		• Solvabiliteitsmarge (S_m)	
		• Saldo (S_a)	
Balanstotaal	B	Balanstotaal	$VPV + EV$

Figuur 1: Balans

In de bovenstaande balans, die gebaseerd is op de zogenaamde fondsvermogensmethode, wordt het eigen vermogen (EV) —met als synoniemen “buffer”, “surplus” en “weerstandvermogen”— van een pensioenfonds gevormd door de minimum solvabiliteitsmarge (S_m) plus het eventuele saldo (S_a).² Met andere woorden:

$$EV = S_m + S_a \quad (1)$$

De verplichtingen worden weergegeven door de voorziening pensioenverplichtingen (VPV). De minimum solvabiliteitsmarge is bedoeld om het verwachte neerwaartse risico inzake de financiële positie van een pensioenfonds op te vangen.

* Adres: GBF, Postbus 7855, 1008 CA Amsterdam. Email: J.Koning@GBF.NL.

¹M. Hilbrand, “De financiering van pensioenen”, in C. Petersen (1990), blz. 206.

²J. Kapteijn, “Reële waarde van de voorziening pensioenverplichtingen in de externe jaarrekening van pensioenfondsen”, in J. Kuné (red.), *Waarde van pensioen; over waardering, zekerheidsstelling en toereikendheid*, Den Haag, 2002, blz. 74–75.

Bij de beoordeling van de solvabiliteit van pensioenfondsen speelt de dekkingsgraad (DG) een belangrijke rol. De dekkingsgraad geeft de verhouding tussen het belegde vermogen en de verplichtingen van een pensioenfonds weer:

$$DG = \frac{B}{VPV} \times 100\% = \left(1 + \frac{EV}{VPV}\right) \times 100\% \quad (2)$$

De dekkingsgraad van een pensioenfonds is dus groter dan 100%, indien er sprake is van een positief eigen vermogen.

In het kader van de formulering van solvabiliteitstoetsen voor pensioenfondsen is een eenduidige en transparante waardering van de activa en de passiva van essentieel belang. Daarom zal in paragraaf 2 worden ingegaan op de verschillende wijzen waarop de beleggingen en vooral de verplichtingen van pensioenfondsen kunnen worden gewaardeerd. Hierna zal in paragraaf 3 een korte beschrijving worden gepresenteerd van het huidige toezicht door de Pensioen- en Verzekeringskamer op de solvabiliteit van pensioenfondsen. Vervolgens zal in paragraaf 4 een aantal recente ontwikkelingen op het gebied van de solvabiliteitsnormering van pensioenfondsen worden besproken. Tot slot wordt in paragraaf 5 een korte samenvatting gegeven en worden enkele conclusies geformuleerd.

1 De waardering van de beleggingen en de verplichtingen

Ten aanzien van de beleggingen kan worden gesteld dat pensioenfondsen heden ten dage vrij algemeen de opvatting huldigen dat de diverse beleggingscategorieën tegen marktwaarde/actuele waarde dienen te worden gewaardeerd.

Inzake de waardering van de verplichtingen van pensioenfondsen is geen sprake van een eenduidige waardering. Er kunnen in dit verband een aantal problemen worden onderscheiden. In de eerste plaats doet zich de vraag voor of bij de huidige waardering (d.w.z. de waardering op een bepaald tijdstip t) van de verplichtingen alleen sprake dient te zijn van de huidige opgebouwde pensioenrechten of ook de in de toekomst nog op te bouwen rechten. Wat de reikwijdte van de pensioenverplichtingen betreft kunnen schematisch gezien 4 fasen worden onderscheiden; deze fasen zijn in het onderstaande schema kort opgesomd:

SCHEMA: De verschillende fasen van de opgebouwde en toekomstige pensioenaanspraken binnen de VPV.³

- Fase 1:** *Opgebouwde rechten/nominaal huidige deelnemersbestand*
- contante waarde opgebouwde pensioenen actieve deelnemers
 - contante waarde slapersrechten
 - contante waarde ingegane pensioenen
- Fase 2:** *Opgebouwde rechten/geïndexeerd huidige deelnemersbestand*
- variant a1: als fase 1 + indexatie opgebouwde rechten
ingegane pensioenen met prijsindex
- variant a2: als fase 2/variant a1, maar indexatie opgebouwde rechten
actieven met maximum (loonindex, prijsindex)
- variant b: als fase 2/variant a1 of a2, maar tevens verhoging opgebouwde
rechten actieve deelnemers met verwachte salarisverhoging
tot pensioendatum (d.w.z. incl. ontwikkeling salaris)
- Fase 3:** *Opgebouwde en toekomstige rechten huidige deelnemersbestand*
Als fase 2/variant b, inclusief in de toekomst op te bouwen
rechten uit hoofde van huidige dienstverbanden
- Fase 4:** *Opgebouwde en toekomstige rechten huidige en
toekomstige deelnemersbestand*
Als fase 3, inclusief toekomstige opbouw in verband
met toekomstige bestandsontwikkeling

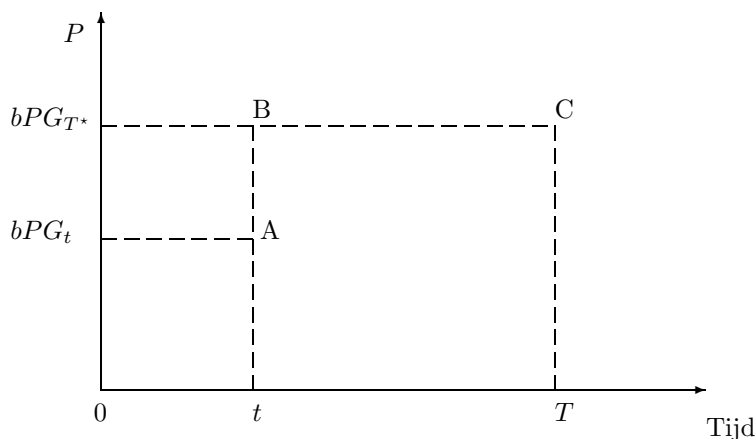
Figuur 2

In geval van fase 1 is het evident dat bij de bepaling van de verplichtingen de opgebouwde pensioenrechten op een minimumniveau worden gewaardeerd. In geval van fase 2 worden de op tijdstip t opgebouwde pensioenaanspraken van de actieve deelnemers en de opgebouwde slapersrechten tot de pensioengerechtigde leeftijd alsmede de ingegane pensioenen van de gepensioneerden geïndexeerd met het verwachte indexatiepercentage. Bij de slapersrechten en de ingegane pensioenen wordt veelal prijsindexatie toegepast. Bij de opgebouwde rechten van de actieve deelnemers kan waarde vaste respectievelijk welvaartsvaste indexatie worden onderscheiden (variant a2). Bij de indexatie van laatstgenoemde pensioenaanspraken kan naast de generieke indexatie tevens verhoging van de opgebouwde rechten met de verwachte salarisontwikkeling tot de pensioengerechtigde leeftijd worden toegepast (variant b). In geval van fase 3 worden in de pensioenverplichtingen niet alleen de op tijdstip t opgebouwde rechten geïncorporeerd, maar ook de door de actieve deelnemers tot de pensioendatum nog op te bouwen pensioenaanspraken. Bij de eerste 3 fasen zijn alleen de pensioenaanspraken van het op tijdstip t bestaande deelnemersbestand aan de orde. In geval van fase 4 worden niet alleen de pensioenaanspraken van het huidige deelnemersbestand in de pensioenverplichtingen betrokken, maar ook de

³E. Ponds, "Waardering pensioenverplichtingen", in: J. Kuné (red.), op. cit., blz. 10.

nog op te bouwen rechten van de deelnemers waarvan wordt verwacht dat ze na tijdstip t aan de pensioenregeling gaan deelnemen.

In de onderstaande figuur zijn de 3 eerstgenoemde fasen voor een individuele deelnemer weergegeven, waarbij volledigheidshalve wordt opgemerkt dat hierbij is uitgegaan van een eindloonregeling en is geabstraheerd van het sterfterisico.



Figuur 3

In geval van fase 1 geldt dat de op tijdstip t opgebouwde pensioenrechten de resultante zijn van het product van enerzijds de opbouwperiode tot tijdstip t en anderzijds het opbouwpercentage gerelateerd aan de nominale pensioengrondslag (salaris – franchise) op tijdstip t .

In geval van fase 2 geldt dat de op tijdstip t opgebouwde pensioenrechten de resultante zijn van het product van enerzijds de opbouwperiode tot tijdstip t en anderzijds het opbouwpercentage gerelateerd aan de —met het verwachte indexatiepercentage— geïndexeerde pensioengrondslag op de pensioendatum T .

In geval van fase 3 geldt dat de op tijdstip t opgebouwde pensioenrechten plus de tot de pensioendatum T nog op te bouwen pensioenrechten de resultante zijn van het product van enerzijds de opbouwperiode tot de pensioendatum T en anderzijds het opbouwpercentage gerelateerd aan de —met het verwachte indexatiepercentage— geïndexeerde pensioengrondslag op de pensioendatum T .

De voorgaande 3 fasen kunnen in de volgende formules worden weergegeven:

$$\begin{aligned} \text{Fase 1: } P_{t(1)} &= tb(S_t - F_t) = tbPG_t && (\text{Rechthoek } OtAbPG_t) \\ \text{Fase 2: } P_{t(2)} &= tb(S_{T^*} - F_{T^*}) = tbPG_{T^*} && (\text{Rechthoek } OtBbPG_{T^*}) \\ \text{Fase 3: } P_{t(3)} &= Tb(S_{T^*} - F_{T^*}) = TbPG_{T^*} && (\text{Rechthoek } OTCbPG_{T^*}) \end{aligned}$$

$$S_t = S_0 \prod_{i=1}^t (1 + p_{l,i})$$

$$S_{T^*} = S_t \prod_{ij=1}^t (1 + p_{l^*j})$$

$P_{t(1,2,3)}$	= pensioenrecht op tijdstip t in fase 1, fase 2 resp. fase 3
S_t	= salaris op tijdstip t
S_{T^*}	= salaris op tijdstip T
F_t	= franchise op tijdstip t
F_{T^*}	= verwachte franchise op tijdstip T
PG_t	= pensioengrondslag op tijdstip t
PG_{T^*}	= verwachte pensioengrondslag op tijdstip T
$p_{l,i}$	= nominale loonvoet
p_{l^*j}	= verwachte nominale loonvoet
b	= opbouwpercentage
T	= pensioendatum

Bij de contante waardebepaling kunnen een drietal factoren worden onderscheiden, te weten de relevante kasstromen, de actuariële kanssystemen en de pensioendisconteringsvoet(en).

Wat de kasstromen betreft, kan worden opgemerkt dat de uitgaande kasstromen betrekking hebben op de pensioenuitkeringen die naar verwachting vanaf de pensioendatum zullen plaatsvinden. Deze uitkeringen vloeien te zijner tijd voort uit de desbetreffende pensioenrechten van de actieve deelnemers en de slapers en regarden uiteraard ook de reeds ingegane pensioenen. De inkomende kasstromen hebben betrekking op de tot de pensioendatum nog te ontvangen pensioenpremies ten behoeve van de actieve deelnemers.

De actuariële kanssystemen hebben met name betrekking op de actuariële grondslagen voor sterfte, invaliditeit, gehuwdheidsfrequenties, afkoop en administratiekosten. Inzake de sterftekansen — tegenwoordig vaker aangeduid als overlevingskansen — kan in het algemeen worden gesteld dat actuelere sterftetafels langdurigere kasstromen impliceren.

Bij de keuze van de pensioendisconteringsvoet(en) kan een onderscheid worden gemaakt tussen enerzijds dezelfde disconteringsvoet voor alle toekomstige perioden — bijvoorbeeld een vaste 4% rekenrente — en anderzijds een voor elke periode eigen, marktconforme disconteringsvoet. Indien wordt geopteerd voor de laatstgenoemde mogelijkheid, wordt dus een rentetermijnstructuur gehanteerd bij het contant maken van de toekomstige kasstromen.

Omdat er voor pensioenverplichtingen geen markt bestaat waarop de toegezegde pensioenaanspraken kunnen worden verhandeld, dient de marktwaarde van de verplichtingen bij benadering te worden vastgesteld. Hierbij kan aansluiting worden gezocht bij (een combinatie) van financiële titels met een vergelijkbaar kasstroomprofiel. In dit verband wordt wel gesproken van een zogenaamde “perfect hedge portfolio” (PHP). Dit is een beleggingsportefeuille waarvan de kasstromen per periode een perfecte match hebben met de verzekeringstechnische kasstromen die besloten liggen in de aangegane verplichtingen van een pensioenfonds. De beleggingsrendementen van de PHP worden als pensioendisconteringsvoeten gehanteerd bij het contant maken van de verwachte kasstromen. Indien daadwerkelijk conform

de PHP wordt belegd, zijn de disconteringsvoeten die in het kader van de waardering van de pensioenverplichtingen worden gehanteerd per periode gelijk aan de (verwachte) beleggingsrendementen. De ontwikkeling in de waarde van het belegde vermogen is in dat geval steeds gelijk aan de ontwikkeling in de waarde van de verplichtingen. In de mate waarin de feitelijke beleggingsportefeuille afwijkt van de PHP is er sprake van een “mismatch risico”, zijnde de standaarddeviatie van de “excess return”. Het “excess”-rendement bestaat in dit verband uit het verschil tussen het feitelijk beleggingsrendement en de PHP-rendementen, waarmee de verwachte kasstromen bij de waardering van de pensioenverplichtingen contant zijn gemaakt.

Indien de pensioentoezegging strikt nominaal is en dus elke vorm van indexatie is uitgesloten—in termen van het fasenschema van figuur 3 is dan uitsluitend fase 1 van toepassing—ligt het voor de hand om de pensioendisconteringsvoeten af te stemmen op de uit de (staats)obligatiekoersen af te leiden nominale rentetermijnstructuur. Indien er daarentegen sprake is van een geïndexeerde pensioentoezegging—in termen van het fasenschema van zijn dan tevens de fasen 2 en eventueel 3 van toepassing—ligt het in de rede om de pensioendisconteringsvoeten af te stemmen op de reële rentetermijnstructuur die kan worden gedestilleerd uit de koersen van index(staats)obligaties.⁴ In dit verband dient er wel op te worden gewezen dat in de constellatie waarin sprake is van loonindexatie indexobligaties een imperfecte match opleveren, omdat bij deze financiële titels prijsindexatie en geen loonindexatie plaatsvindt.⁵

Een interessant aandachtspunt inzake de beleggingsrendementen als grondslag voor een marktconforme waardering van de pensioenverplichtingen betreft de vraag of hierbij ook gebruik mag worden gemaakt van reële aandelenrendementen, indien aandelen een onderdeel vormen van de beleggingsportefeuille.⁶ Indien deze vraag bevestigend wordt beantwoord, impliceert dit ceteris paribus een lagere waardering van de pensioenverplichtingen, omdat aandelen vanwege een positieve risicopremie op langere termijn een hoger rendement worden geacht op te leveren dan obligaties.

⁴In A. Quix, J. Kapteijn en H. Sallé, “Schets van een financieel-economische benadering voor de bepaling van overschotten en tekorten bij pensioenfondsen”, in: J. Kuné (red.), op. cit., blz. 53 wordt er op gewezen dat er eigenlijk 2 mogelijkheden voorhanden zijn, namelijk: (1) discontering van de verwachte kasstromen, inclusief indexering met de nominale rente(termijnstructuur), of (2) discontering van de verwachte kasstromen, exclusief indexering met de reële rente(termijnstructuur).

⁵In dit verband kan nog worden opgemerkt dat Nederland nog geen indexobligaties kent. Buiten Nederland bestaan betrekkelijk kleine markten voor indexobligaties. In de Euro-zone kent Frankrijk een tweetal langlopende indexobligaties, waarvan de indexatie is gekoppeld aan de Franse prijsinflatie. In 2001 heeft de Franse overheid een indexobligatie geëmitteerd, waarvan de indexatie is gekoppeld aan de Europese prijsinflatie. Zie E. Ponds, op. cit., blz. 10–12.

⁶Zie in dit verband H. Klein Haneveld, “De kosten van pensioenen” en E. Ponds, “Waardering verplicht”, in *Economisch Statistische Berichten*, 8 juni 2001, blz. 488–492 en E. Ponds, “Waardering pensioenverplichtingen”, in: J. Kuné, op.cit., blz. 20–25.

2 Solvabiliteitsnormering: huidige toezicht Pensioen- en Verzekeringskamer (PVK)

Het toezicht van de PVK op de solvabiliteit van de pensioenfondsen is gebaseerd op de Pensioen- en Spaarfondsenwet (PSW) en op de uitwerking daarvan in de Actuariële Principes Pensioenfondsen. In 1997 heeft de toenmalige Verzekeringskamer (VK) de Actuariële Principes Pensioenfondsen 1997 (APP/1997) geïntroduceerd.⁷ Bij deze introductie is van de zijde van de VK aangegeven dat op redelijk korte termijn een herziening van de actuariële principes ter hand zou worden genomen; hierdoor worden de APP/1997 veelal aangeduid als APP1. Nadat de PVK in maart 2000 in de vorm van een uitgangspuntennotitie de Nieuwe Actuariële Principes (NAP) ter consultatie aan onder andere de drie koepelorganisaties van pensioenfondsen en verzekeringsmaatschappijen voorlegde, laat de PVK in maart 2001 aan belanghebbenden weten de NAP te willen vervangen door een zogenaamd Financieel Toetsingskader (FTK). Volgens planning van de PVK is het streven er op gericht het FTK met ingang van 1 januari 2006 in te voeren.

Op grond van het voorgaande kan worden gesteld dat momenteel nog steeds de APP1 van kracht zijn.⁸ De kern van de APP1 bestaat uit een door de actuaire uit te voeren toereikendheidstoets op de voorziening pensioenverplichtingen (VPV). Inzake de VPV schrijven de APP1 een minimumpositie voor. De VPV dient in het standaardgeval van voorwaardelijke indexering namelijk minimaal gelijk te zijn aan de contante waarde van alle opgebouwde rechten, waarbij als disconteringsvoet een vaste rekenrente van maximaal 4% mag worden gehanteerd.⁹ De meeropbrengsten boven de rekenrente worden geacht beschikbaar te zijn voor indexering van de opgebouwde rechten. In termen van het fasenschema van figuur 3 dient de VPV dus in elk geval betrekking te hebben op fase 1. De in het pensioenfonds aanwezige middelen — zijnde het belegde vermogen — moeten op elk moment ten minste gelijk te zijn aan de ondergrens van de VPV. Het belegde vermogen dient te worden gewaardeerd tegen actuele waarde/marktwaarde. Voor het risico van neerwaartse fluctuaties in de marktwaarde van het belegde vermogen, dient een zogeheten weerstandsvermogen te worden gevormd. De omvang van het weerstandsvermogen is in de APP1 niet voorgeschreven. Wel wordt in de APP1 aangegeven dat er een zodanig weerstandsvermogen aanwezig dient te zijn dat onderdekking van de minimumpositie van de VPV zich niet zal voordoen. Als norm voor het neerwaartse risico van de beleggingen dient te worden uitgegaan van historische data en de naar redelijkheid te verwachten fluctuaties in de marktwaarde van de beleggingen. Uitgaande van het empirische feit dat de volatiliteit van het totaalrendement van aandelen gemiddeld groter is dan de volatiliteit van het totaalrendement van vastrentende waarden, zal

⁷Pensioen- en Verzekeringskamer, *Actuariële Principes Pensioenfondsen*, Apeldoorn, februari 1997.

⁸Na de afsluiting van deze bijdrage heeft de PVK in de vorm van een brief d.d. 30 september 2002 aan alle pensioenfondsen “de belangrijkste uitgangspunten voor de toetsing van hun financiële positie (nog eens) op een rij gezet. De uitgangspunten bevatten naast bestaand beleid een aantal aanvullingen.” In het kort wordt de onderhavige brief in Bijlage 1 geresumeerd.

⁹Volledigheidshalve kan nog worden opgemerkt dat uitsluitend in het geval dat indexering volstrekt is uitgesloten, dus indien slechts nominale aanspraken zijn toegezegd, een hogere rekenrente dan 4% mag worden gehanteerd.

het weerstandsvermogen dus hoger dienen te zijn naarmate de beleggingsportefeuille van een pensioenfonds een relatief grotere component aandelen bevat. Op grond van het voorgaande kan worden gesteld dat, rekening houdend met de grootte van het neerwaartse risico van de beleggingen, de dekkingsgraad op elk moment tenminste 100% zal bedragen. Een uitzondering op deze dekkingseis vormt de situatie waarin sprake is van exceptioneel ongunstige ontwikkelingen op de financiële markten. In een dergelijke situatie is een beperkte, tijdelijke onderdekking niet bij voorbaat uitgesloten, mits er sprake is van een voorwaardelijke indexatie.

Tot slot van deze paragraaf worden nog een tweetal kanttekeningen geplaatst. In de eerste plaats kan worden vastgesteld dat de in de APP1 voorgestane wijze van waardering van de pensioenverplichtingen afwijkt van de wijze van waardering van de beleggingen. Een consistente waardering van de verplichtingen impliceert dat de beleggingsrendementen van de desbetreffende PHP als pensioendisconteringsvoeten worden gehanteerd, hetgeen bij de 4% rekenrente van de APP1 evident niet het geval is. In de tweede plaats kan worden geconstateerd dat de beoordeling van de solvabiliteit in sterke mate is geconditioneerd door de in de PSW vervatte discontinuïteitsbenadering. Volgens deze benadering dient een pensioenfonds ten alle tijde in staat te zijn de pensioenverplichtingen, zonder korting van de aanspraken, zelfstandig —dit wil zeggen zonder hulp van de desbetreffende onderneming respectievelijk bedrijfstak— af te wikkelen dan wel over te dragen aan een verzekeringsmaatschappij. Een dergelijk afwikkelscenario kan actueel worden, indien de sponsor van het pensioenfonds —de onderneming in geval van een ondernemingspensioenfonds of de aangesloten ondernemingen in geval van een bedrijfstakpensioenfonds— zou wegvallen. Van diverse zijden is er op gewezen dat voor pensioenfondsen, anders dan voor een verzekeringsmaatschappij, een continuïteitsbenadering meer in de rede ligt.¹⁰ In een dergelijke lange termijn benadering van de solvabiliteit wordt uitgegaan van continuïteit van de relatie tussen pensioenfonds en onderneming/bedrijfstak. In het nieuwe FTK, dat in de volgende paragraaf in het kort zal worden gepresenteerd, is wel een continuïteitselement opgenomen.

3 Solvabiliteitsnormering: recente ontwikkelingen

In deze paragraaf zal eerst een korte beschrijving van de (uitgangspunten voor) een Financieel Toetsingskader van de PVK worden gegeven. Hierna worden de hoofdlijnen weergegeven van de voorstellen van het Actuarieel Genootschap inzake de solvabiliteitsnormering, zoals vervat in het rapport “Pensioenzekerheid dynamisch getoetst”.

¹⁰Zie in dit verband onder andere Vereniging van Bedrijfspensioenfondsen/Stichting voor Ondernemingspensioenfondsen/Pricewaterhouse N.V., *Continu op koers*, Utrecht, januari 2000 en Actuarieel Genootschap, *Pensioenzekerheid dynamisch getoetst. Aanbevelingen en een werkmodel voor de inrichting van de beoordeling van de solvabiliteit van pensioenfondsen*, oktober 2001.

3.1 Financieel Toetsingskader (FTK)

Het FTK is nog in ontwikkeling en is in eerste instantie in de vorm van de brochure “De Uitgangspunten voor een Financieel Toetsingskader” gepresenteerd.¹¹ De uitgangspuntenbrochure beschrijft het methodisch kader voor de toereikendheidstoets van pensioenfondsen en verzekeringsmaatschappijen. In dit methodisch kader wordt uitgegaan van een consistente waardering van activa en passiva, namelijk actuele waarde. Dit betekent dus dat de pensioenverplichtingen niet meer op basis van een vaste 4% rekenrente worden gewaardeerd, maar op basis van pensioendisconteringsvoeten die zijn afgeleid van de rendementen van financiële titels met een vergelijkbaar kasstroomprofiel.

De toereikendheidstoets kent een drietal deeltolsten, namelijk:

- (a) Minimumtoets;
- (b) Solvabiliteitstoets;
- (c) Continuïteitstoets.

Zoals hiervoor reeds is aangegeven dient de PVK de uitgangspunten inzake de 3 deeltolsten nog te vertalen naar in de praktijk hanteerbare deeltolsten. Daarom kan hierna alleen worden gerefereerd aan de algemene beschrijving van de 3 deeltolsten, zoals deze zijn gepresenteerd in de uitgangspuntenbrochure.

De *minimumtoets* bevat een tweetal componenten. De eerste component geeft aan dat op het toetsmoment de VPV minimaal gelijk dient te zijn aan de actuele waarde in het economische verkeer. Hierbij wordt inzake de reikwijdte van de pensioenverplichtingen uitgegaan van de rechtens afdwingbare pensioenaanspraken. De invulling van het vaststellen van de actuele waarde van de opgebouwde rechten met een voorwaardelijke indexatie is voor de PVK nog onderwerp van verdere studie. In termen van het fasenschema van figuur 3 is nog niet duidelijk of in geval van voorwaardelijke indexering de pensioenverplichtingen zich uitstrekken tot en met fase 2 (variant a). De tweede component geeft aan dat de actuele waarde van de aanwezige activa, zijnde de marktwaarde/actuele waarde van het belegde vermogen, op het toetsmoment tenminste gelijk dient te zijn aan de VPV. Het doel van de minimumtoets is dat op balansdatum de pensioenverplichtingen zonder korting van de pensioenaanspraken op marktconforme condities kunnen worden overgedragen.

De *solvabiliteitstoets* is er op gericht de omvang van de actuele waarde van de aanwezige middelen zodanig te bepalen, dat deze —in geval van een ongunstig scenario, bezien over een 1-jaars periode— aan het einde van de desbetreffende periode met een grote mate van waarschijnlijkheid de actuele waarde van de verplichtingen zal overtreffen. Volledigheidshalve wordt nog opgemerkt dat de pensioenverplichtingen bij de onderhavige toets niet alleen betrekking hebben op de rechtens afdwingbare pensioenaanspraken maar ook de aanspraken uit hoofde van voorwaardelijke indexering. Wanneer zich na balansdatum een ongunstig scenario voltrekt, dient een pensioenfonds de pensioenverplichtingen 1 jaar na balansdatum

¹¹Pensioen- en Verzekeringskamer, *De Uitgangspunten voor een Financieel Toetsingskader*, Apeldoorn, september 2001.

zonder korting van de aanspraken op marktconforme condities te kunnen overgedragen. Om een dergelijk 1-jaars afwikkelingscenario te kunnen effectueren dient op balansdatum de actuele waarde van de beleggingen minimaal gelijk te zijn aan de actuele waarde van de verplichtingen plus de benodigde solvabiliteitsmarge (weerstandvermogen). Bij de bepaling van deze marge dient, bezien over een 1-jaars periode, rekening te worden gehouden met:

- mismatch risico;
- marktrisico en kredietrisico;
- structurele afwijking van de verwachte kosteninflatie;
- slecht verzekeringsresultaat;
- tegenpartijrisico herverzekeraar;
- groot specifiek verzekeringsrisico.

Van belang is nog te vermelden dat de PVK in de uitgangspunten-brochure stipuleert: “Hoewel ... de actuele waarde van de aanwezige middelen met een acceptabele maar kleine kans één jaar na balansdatum lager kan zijn dan de actuele waarde van de verplichtingen, mag hieruit niet de conclusie worden getrokken dat een negatieve solvabiliteit (ook al is het tijdelijk) zou mogen voortbestaan. De actuele waarde van de aanwezige middelen dient immers gelijk te zijn aan de actuele waarde van de verplichtingen.”¹²

In termen van de dekkingsgraad kan de solvabiliteitstoets met behulp van de Solvency at Risk (SaR) methode als volgt kort worden toegelicht. Vooraf zij nog opgemerkt dat de SaR-methode een variant is van de Value at Risk (VaR) methode. De VaR is een asymmetrische risicomaatstaf die is ontwikkeld om de blootstelling aan het neerwaartse marktrisico te kwantificeren. De VaR is het maximale verlies —uitgedrukt in een geldbedrag— dat in een bepaalde periode, gegeven een gekozen betrouwbaarheidsinterval, als gevolg van het marktrisico kan worden verwacht. Een VaR-waarde is afhankelijk van volgende 3 factoren:

- simulatiemethode
- betrouwbaarheidsinterval
- lengte beschouwde periode.

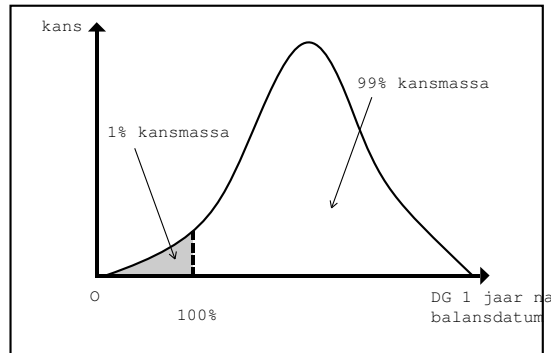
Gegeven de simulatiemethode —waarbij in hoofdzaak de variantie-covariantie methode, de historische simulatie en de Monte Carlo simulatie kunnen worden onderscheiden— kan worden gesteld dat de VaR ceteris paribus groter is naarmate het betrouwbaarheidsinterval hoger is en de beschouwde periode langer is.¹³

De SaR-methode behelst in het kader van de solvabiliteitstoets een zodanige bepaling van de solvabiliteitsmarge dat de dekkingsgraad met een voldoende kleine kans op het tijdstip 1 jaar na balansdatum niet kleiner is dan 100%. De onderstaande figuur geeft het voorgaande voor een betrouwbaarheidsinterval van 99% grafisch weer.

Indien een betrouwbaarheidsgraad van 99% een SaR-waarde van bijvoorbeeld 20% ten opzichte van de VPV vergt, dient de dekkingsgraad op balansdatum dus 120%

¹²Pensioen- en Verzekeringskamer, *De Uitgangspunten voor een Financieel Toetsingskader*, Apeldoorn, september 2001, blz. 12.

¹³Zie voor een heldere uiteenzetting van de VaR-methode: R. Polfiet, *Value at Risk, Verlies als maatstaf*, Antwerpen/Apeldoorn, 1999.



Figuur 4

te zijn, namelijk:

$$\text{Balansdatum : DG} = \frac{B}{VPV} \times 100\% = \left(1 + \frac{S_m}{VPV}\right) \times 100\% = 120\%.$$

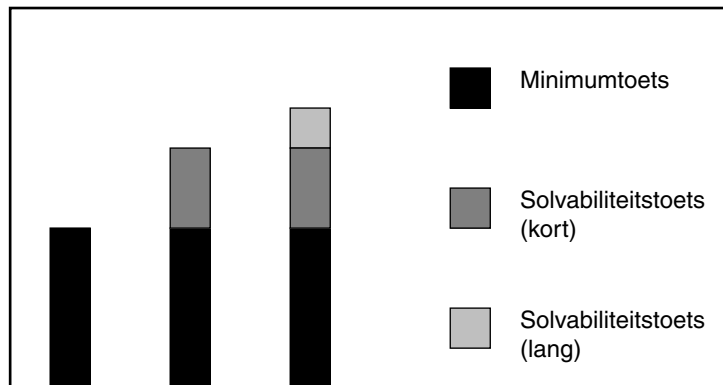
De initiële dekking van 120% bewerkstelligt dat de toekomstige dekking 1 jaar na balansdatum met een kans van 99% niet kleiner zal zijn dan 100%; met andere woorden er is een kans van 1% dat bij een initiële dekking van 120% de dekking 1 jaar na balansdatum kleiner zal zijn dan 100%. Opgemerkt kan nog worden dat de 1% kans op onderdekking als zodanig geen informatie verschaft over de complementaire omvang van de onderdekking.

Duidelijk zal zijn dat de minimumtoets en de solvabiliteitstoets als gemeenschappelijk kenmerk hebben dat beide toetsen met een tijdshorizon van 1 jaar de afwikkelpositie van een pensioenfonds normeren.

De *continuïteitstoets* brengt de toekomstige solvabiliteitsontwikkeling in beeld op basis van continuïteit van een pensioenfonds. In het kader van een lange termijn horizon wordt bij de vaststelling van de pensioenverplichtingen, naast de opgebouwde rechten, tevens rekening gehouden met de in te toekomst nog op te bouwen pensioenaanspraken van zowel het huidige als het toekomstige deelnemersbestand. In termen van het fasenschema van figuur 3 strekt de VPV zich bij de onderhavige deelttoets dus tot en met fase 4 uit. Verder wordt ook rekening gehouden met de toekomstige premieinkomsten. Bovendien worden bij de continuïteitstoets ook bijsturinginstrumenten zoals het premiebeleid, het indexeringsbeleid en het beleggingsbeleid in de beschouwing betrokken.

De continuïteitstoets —die binnen de toereikendheidstoets een nieuw element vormt— heeft in vergelijking met de beide andere deelttoetsen de meeste uitwerking. Op basis van de uitgangspunten-brochure kan in het kort worden gesteld dat de PVK vanuit een continuïteitsperspectief de lange termijn solvabiliteitsrisico's beoogt te normeren, resulterende in vermogensbandbreedtes waarbinnen financiële tegenslagen kunnen worden opgevangen.

In de onderstaande figuur zijn de implicaties van de 3 deelttoetsen voor de omvang van de solvabiliteit grafisch weergegeven.



Figuur 5

In aanvulling op de bovenstaande figuur kan nog worden opgemerkt dat de PVK er in de uitgangspuntenbrochure op wijst dat een pensioenfonds op elk van de 3 deoltoetsen van de toereikendheidstoets een positief resultaat dient te tonen.¹⁴ Hiervoor is gebleken dat een pensioenfonds voortdurend dient te voldoen aan de minimumtoets. Bovendien dient een fonds op elke balansdatum over een zodanig solvabiliteitsmarge te beschikken dat naar alle waarschijnlijkheid 1 jaar na balansdatum tenminste wordt voldaan aan de minimumtoets. Indien dit onverhoopt niet het geval is, dient het pensioenfonds blijkbaar onverwijld zodanige maatregelen te nemen dat de onderdekking volledig wordt geredresseerd. Inzake de continuïteitstoets doet zich de vraag voor of deze toets (uitsluitend) antwoord moet geven op de vraag of niet alleen na 1 jaar, maar in alle jaren van de beschouwde lange termijn periode wordt voldaan aan de minimumtoets.¹⁵ Indien dat inderdaad het geval zou zijn, is de “minimumpositie” dermate stringent dat de resterende beleidsruimte relatief beperkt is.

3.2 Actuariel Genootschap: Pensioenzekerheid dynamisch getoetst

In het belangwekkende rapport van de Actuariel Genootschap (AG) “Pensioenzekerheid dynamisch getoetst”, hierna kortheidshalve als het AG-rapport¹⁶ aangeduid,

¹⁴Pensioen- en Verzekeringskamer, *De Uitgangspunten voor een Financieel Toetsingskader*, Apeldoorn, september 2001, blz. 8.

¹⁵Zie in dit verband ook: D. De Jong en A. Hussem, “De inrichting van de solvabiliteitsbeoordeling van pensioenfonds: toetsing op uitkeringszekerheid”, in J. Kuné, op. cit., blz. 123.

¹⁶Na de afsluiting van deze bijdrage heeft het Actuariel Genootschap (AG) op 8 januari 2003 een pilotstudie gepubliceerd: “Prijs van zekerheid hand in hand met solvabiliteitsbeoordeling, Pilotstudie Pensioenzekerheid dynamisch getoetst.” In deze studie wordt verslag gedaan van het testen van het eerder uitgebrachte rapport “Pensioenzekerheid dynamisch getoetst” (1 november 2001) op een dwarsdoorsnede van de Nederlandse pensioenfonds op technische juistheid van het concept, werkbaarheid voor grote en kleine pensioenfonds en communicatie/interpretatie van de toetsuitkomsten. Het AG komt tot de conclusie dat de pilot geslaagd is.

wordt een volledig uitgewerkt toetsinstrumentarium geformuleerd dat in hoofdzaak uit de volgende 3 onderdelen bestaat:

- een 3-dimensionele maat voor uitkeringszekerheid;
- indexeringsindicatoren;
- een werkhypothese voor continuïteit.

De maat voor uitkeringszekerheid knoopt aan bij de benadering dat de centrale en voor de deelnemer aan een pensioenregeling enig relevante vraag inzake de financiële positie van een pensioenfonds de vraag betreft wat de zekerheid is dat de beloofde uitkering op het beloofde moment kan worden nagekomen. De op een bepaald moment aanwezige dekking van de pensioenaanspraken zal door een deelnemer op basis van de uit die dekking resulterende zekerheid omtrent de pensioenuitkeringen worden beoordeeld. Het gaat hierbij hoofdzakelijk om de zekerheid inzake de nominale uitkeringen in verband met de in het verleden opgebouwde pensioenrechten (fasenschema figuur 2: fase 1). Indien het feitelijke vermogen van een pensioenfonds groter is dan het voor de nominale afwikkeling benodigde vermogen, is er in het afwikkelingscenario ruimte voor indexatie van de in het verleden opgebouwde rechten. De indexeringsindicatoren van het toetsingsinstrumentarium geven de ruimte voor indexatie aan.

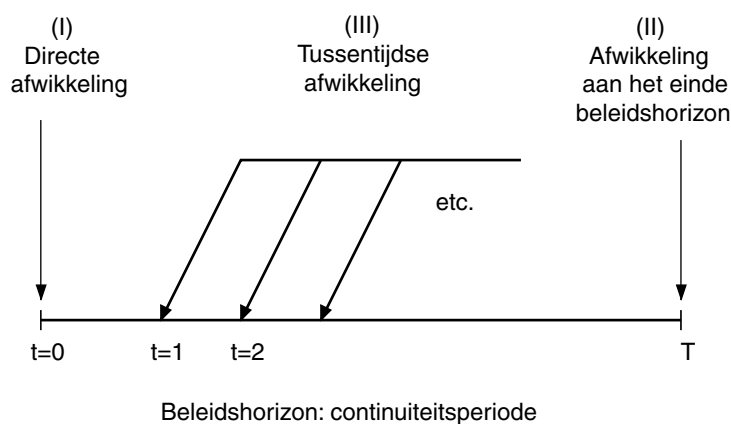
Van belang is dat in het AG-rapport wordt benadrukt dat de doelstelling van een pensioenfonds continuïteit van de uitvoering van de arbeidsvoorwaarde pensioen is. In beginsel geldt inzake de continuïteit een oneindige tijdshorizon, waarbinnen een pensioenfonds altijd haar verplichtingen zal kunnen nakomen. In vergelijking met een levensverzekeringsmaatschappij beschikt een pensioenfonds namelijk over een aantal extra beleidsinstrumenten voor de sturing van de financiële positie. Het betreft hier het premiebeleid, indexeringsbeleid, het beleid inzake de inhoud van de pensioenregeling en het beleggingsbeleid (asset-mix). Ook al is continuïteit van de uitvoering de doelstelling van een pensioenfonds, het wegvallen van de sponsor(s), waardoor een pensioenfonds zou moeten afwikkelen, kan niet worden uitgesloten. Daarom wordt bij de vormgeving van het continuïteitsstreven (nooit afwikkelen) rekening gehouden met een tussentijdse noodzaak tot afwikkelen.

In de onderstaande figuur zijn de 3 modaliteiten van fictieve afwikkeling, die in het AG-rapport worden onderscheiden, schematisch weergegeven.

Bij de bepaling van de *zekerheidsmaat voor de uitkeringen* gaat het in eerste instantie om de hiervoor al aangeduide nominale uitkeringen in verband met de in verleden opgebouwde pensioenaanspraken. In het kader van de directe afwikkeling, dit wil zeggen de onmiddellijke afwikkeling op tijdstip t_0 (afwikkeling I/figuur 6), verloopt de vormgeving van de zekerheidsmaat als volgt. Op basis van een veronderstelde economische toekomst inzake de beleggingsrendementen en de loon- en prijsinflatie worden een groot aantal scenario's —in het AG-rapport gaat het om 500 scenario's— gegenereerd, waarbinnen het mogelijke verloop van het belegde vermogen als gevolg van de verwachte uitkeringen en de beleggingsinkomsten wordt geprojecteerd.¹⁷ In de onderstaande figuur zijn ter illustratie een drietal mogelijke projecties van het vermogen weergegeven.

In geval van scenario 1 is er op het eindtijdstip —zijnde het tijdstip, waarop de laatste uitkering wordt gedaan— sprake van een positief vermogen. Scenario's waarvoor

¹⁷In het AG-rapport worden de volgende basis economische toekomstige gehanteerd:



Figuur 6

geldt dat er op het eindtijdstip nog vermogen over is, worden als overschotsscenario's aangeduid. In geval van scenario 2 is de omvang van het vermogen op het eindtijdstip precies voldoende om de laatste uitkering te kunnen doen. In geval van scenario 3 is het vermogen reeds voor het eindtijdstip, namelijk op tijdstip t^* , nihil geworden. Alhoewel er na tijdstip t^* dus eigenlijk geen uitkeringen meer kunnen plaatsvinden, wordt in dergelijke tekortsenario's de afwikkeling toch als het ware voortgezet om te kunnen bepalen hoe groot het tekort op het eindtijdstip zou zijn geworden. Tevens wordt voor alle tekortsenario's tijdstip t^* , namelijk het tijdstip waarop het vermogen nihil is geworden, vastgelegd.

Van de totale simulatie van 500 scenario's worden de volgende 3 kengetallen berekend:

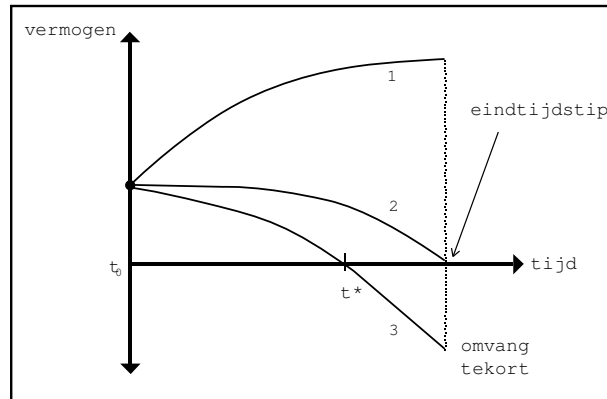
- in hoeveel scenario's treedt een tekort op (levert gerelateerd aan het totaal aantal scenario's de tekortfrequentie op)?
- hoe snel in de tijd treden de eventuele tekorten op?
- hoe groot is de omvang van de optredende tekorten?

De in de diverse scenario's optredende eindsaldi (eindtekorten respectievelijk eindoverschotten) worden per scenario gediscoteerd naar een aanvangssaldo op tijdstip

	min.	1%	5%	med.	gem.	95%	99%	max.	stand. dev.
looninflatie	2,3	3,0	3,6	4,7	4,8	6,1	6,7	6,9	0,8
rendementen									
vastrentend	5,1	5,4	5,8	6,7	6,7	7,6	7,9	8,1	0,6
zakelijk	5,2	5,8	7,1	10,0	10,0	13,0	14,0	14,6	1,7

Economische toekomst: alle scenario's (Uiteraard is de spreiding in geval van alle scenario's en alle jaren substantieel hoger.)

Volledigheidshalve wordt nog vermeld dat in het AG-rapport naast de basis economische toekomst, 2 alternatieve toekomst worden uitgewerkt. Tevens wordt er nog op gewezen dat inzake de basis economische toekomst een aantal gevoeligheidsanalyses worden doorgerekend, waarin de implicaties van een 2% lagere risicopremie van aandelen respectievelijk aandelenrendementen die in de eerste 2 jaren -25% bedragen, zijn nagegaan alsmede de implicaties van een looninflatie van 6,5% (in plaats van 4,8%) worden gezien.



Figuur 7

t_0 , waarbij als disconteringsvoet het in het desbetreffende scenario resulterende beleggingsrendement wordt gehanteerd. Deze aanvangssaldi worden in een percentage van de op tijdstip t_0 opgebouwde pensioenaanspraken uitgedrukt. Wanneer er in een scenario sprake is van een aanvangstekort van bijvoorbeeld 10%, dan wil dit dus zeggen dat als de pensioenaanspraken met onmiddellijke ingang (= tijdstip t_0) met 10% worden verlaagd, het eindtekort in het desbetreffende scenario volledig wordt geëlimineerd, hetgeen impliceert dat er op het eindtijdstip precies voldoende vermogen beschikbaar is om de laatste uitkering te verrichten.

In het AG-rapport worden 2 zekerheidsmaten geformuleerd, te weten: *ongewogen zekerheidsmaat* = $1 - \text{tekortfrequentie}$ en *gewogen zekerheidsmaat* = $1 - \text{omvanggewogen tekortfrequentie}$

In het AG-rapport wordt ter illustratie het volgende cijfervoorbeeld gegeven:

Voorbeeld.
 Tekortfrequentie = 20%
 Gemiddelde rechtenkorting = 5% (van de tekortgevallen)

Ongewogen zekerheid = $(1 - 0.2) \times 100\% = 80\%$

Gewogen zekerheid = $\frac{400 + 100(1 - 0.05)}{500} \times 100\% = 99\%$

Figuur 8

De ongewogen zekerheid van 80% geeft aan dat voor 400 van de 500 scenario's geldt dat de uitkeringen zonder rechtenkorting volledig kunnen worden geëffectueerd, omdat in deze 400 scenario's het vermogen op het eindtijdstip voldoende is voor de betaling van de laatste uitkering. Dit impliceert evenwel niet dat er in de overige 100 scenario's volstrekt geen uitkeringen plaatsvinden. Voor de 100 tekortscenario's geldt immers tot tijdstip t^* dat de uitkeringen volledig kunnen worden verricht en daarna niet meer. Zoals eerder aangegeven is deze constellatie equivalent met een

constellatie waarin vanaf het aanvangstijdstip t_0 tot het eindtijdstip de uitkeringen uit hoofde van een gemiddelde korting van 5% van de opgebouwde rechten kunnen plaatsvinden. In de gewogen zekerheidsmaat tellen dus niet alleen die scenario's mee, waarin het wat de uitkeringen betreft helemaal goed gaat, maar ook de scenario's waarin het voor een deel goed gaat.

Het niveau van de (gewogen) zekerheid is afhankelijk van onder andere de volgende factoren:

- De inhoud van de pensioenregeling. Deze factor vindt zijn weerslag in de hoogte van de verplichtingen; hoe hoger de verplichtingen, des te lager ceteris paribus de zekerheid.
- De duration (gemiddelde duur uitkeringen tegen een discontovoet van 0%) van de verplichtingen. Hoe hoger de duration, des te hoger ceteris paribus de zekerheid; met andere woorden een jong fonds heeft ceteris paribus een hogere zekerheid dan een oud gerijpt fonds.
- De initiële (tijdstip t_0) dekkingsgraad. Hoe hoger de dekkingsgraad bij aanvang van de afwikkeling, des te hoger de zekerheid.
- De gekozen beleggingsmix. Ten aanzien van deze factor kunnen geen algemene uitspraken worden gedaan.¹⁸

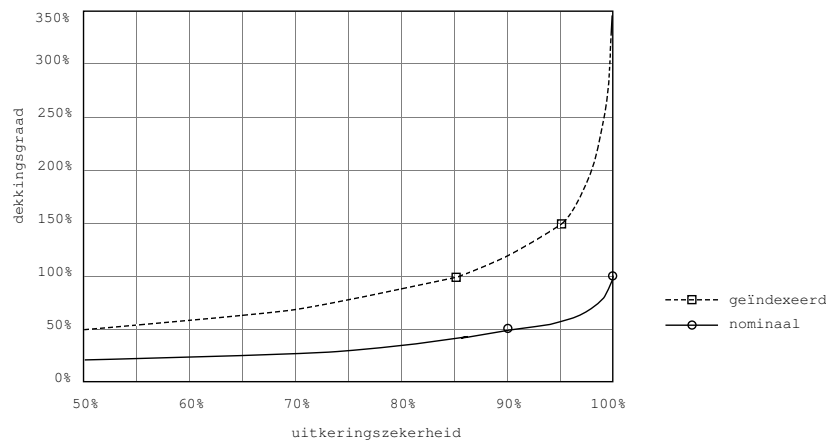
In het AG-rapport wordt ter illustratie een tabel met een aantal zekerheidsuitkomsten van de directe afwikkeling van een zogenaamd voorbeeldfonds gepresenteerd.¹⁹ Deze tabel is als Bijlage 2 in deze bijdrage opgenomen. In de onderstaande figuur is voor het in de bijlage vermelde voorbeeldfonds —met als uitgangssituatie ultimo 1999— de relatie weergegeven tussen enerzijds de klassieke zekerheidsmaat, de 4%-dekkingsgraad, en anderzijds de nieuwe AG-zekerheidsmaat, de (gewogen) zekerheidsmaat voor de uitkeringen.

Uit figuur 9 blijkt dat in geval van nominale afwikkeling een dekkingsgraad van 100% overeenkomt met een uitkeringszekerheid van 100%. In geval van afwikkeling met onvoorwaardelijke indexering kan worden geconstateerd dat een dekkingsgraad van 100% correspondeert met een uitkeringszekerheid van minder dan 100%, namelijk ca. 85%. Een verhoging van de uitkeringszekerheid tot 100% blijkt een dekkingsgraad van meer dan 300% te vergen. In geval van onvoorwaardelijke indexering neemt de zekerheid dus in vergelijking met de nominale afwikkeling sterk af. Verder blijkt dat vanaf een niveau van uitkeringszekerheid van ca. 95% blijkt dat de laatste procentpunten extra uitkerings-zekerheid een substantieel extra vermogen en dus een sterke verhoging van de premie vergen.

Nominale afwikkeling binnen een vooraf gestelde norm voor de (gewogen) uitkeringszekerheid kan worden beschouwd als het minimum dat een pensioenfonds aan de deelnemers moet kunnen bieden. Het maximale is dat een pensioenfonds, naast nominale afwikkeling van de in het verleden opgebouwde rechten, de uitkeringen onvoorwaardelijk zal aanpassen aan de loon-/prijsinflatie. Een tussenweg

¹⁸Er worden in het AG-rapport nog 2 andere factoren genoemd, namelijk (1) de gemodelleerde economische toekomst en (2) de kanssystemen in de fondsdemografie. Voor deze beide factoren wordt kortheidshalve verwezen naar het AG-rapport, blz. 19.

¹⁹Naast het voorbeeldfonds, dat standaard een beleggingsmix van 40% vastrentende waarden en 60% zakelijke waarden kent, worden nog een zestal andere fondstypologieën geanalyseerd; de typologie betreft de verschillen in dekkingsgraad beleggingsmix, gemiddelde uitkeringsduur en het laatste jaar van uitkering.



Figuur 9

is voorwaardelijke indexering, namelijk indexatie indien de financiële positie van het pensioenfonds het toelaat. In het kader van de in figuur 6 aangegeven directe afwikkeling wordt voorwaardelijke indexering volgens de volgende beslissingsregel geoperationaliseerd: er wordt in enig jaar geïndexeerd, indien het vermogen in dat jaar groter is dan nodig voor nominale afwikkeling, gegeven een vooraf gestelde norm voor uitkeringszekerheid. In dit verband kan nog worden opgemerkt dat voor een eenmaal toegezegde indexering dezelfde uitkeringszekerheid wordt gewenst als voor nominale pensioenrechten.

De zekerheidsmaat voor de uitkeringen kan zowel in het kader van de directe afwikkeling (afwikkeling I/figuur 6) als in het kader van afwikkeling binnen de premisse van continuïteit (afwikkeling II en III/figuur 6) worden toegepast.

Naast de uitkeringszekerheid inzake uitkeringen die voortkomen uit de over het verleden opgebouwde nominale pensioenaanspraken, verschaft het toetsinstrumentarium in de vorm van indexeringsindicators informatie over de indexeringsruimte van een pensioenfonds. Deze ruimte heeft betrekking op de omvang van het vermogen boven het voor de nominale uitkeringszekerheid benodigde vermogen, dat voor indexering kan worden aangewend. Evident is dat de indexeringsruimte afhankelijk is van de aard van het indexeringsstreven. De indexeringsindicatoren beogen de indexeringsruimte binnen de financiële positie aan te geven, waarbij als referentiepunt het maximaal mogelijke van de onvoorwaardelijke indexering wordt gehanteerd. In het AG-rapport worden de volgende indexeringsindicatoren gepresenteerd:

- de kans dat onvoorwaardelijke indexering gedurende de gehele afwikkelperiode kan worden volgehouden
- het gemiddelde aantal jaren dat vanaf het begin van de afwikkeling onafgebroken zonder aftopping wordt geïndexeerd (waarna in het geheel niet meer zal worden geïndexeerd)
- het met de uitkeringen gewogen gemiddelde indexeringspercentage
- het percentage van de opgetreden inflatie, dat gemiddelde en gewogen kan worden

doorgegeven.

Hierna wordt nu nog het derde element van het toetsinstrumentarium, de werkhypothese voor continuïteit, in hoofdlijnen weergegeven. In het kader van deze hypothese wordt bestuur van een pensioenfonds geacht een tijdshorizon —namelijk de periode van tijdstip t_0 tot tijdstip T (zie figuur 6)— te bepalen, waarvoor een financiële opzet wordt vormgegeven. De lengte van deze beleidshorizon is met name afhankelijk van de bestandsopbouw van het pensioenfonds en de business van de sponsor(s). Tijdens gekozen continuïteitsperiode vindt tot tijdstip T opbouw van pensioenrechten plaats; er is dus sprake van een opbouwfase. Verder geldt dat in de opbouwfase wordt geïndexeerd, hetgeen impliceert dat de premie een indexeringscomponent bevat.²⁰ In de continuïteitsperiode worden de uitkeringen en de premie- en beleggingsinkomsten gegeven verondersteld en wordt een ontwikkeling aangenomen van toe- en uittredende deelnemers. In termen van het fasenschema van figuur 2 is in de continuïteitsperiode dus fase 4 van toepassing. Tijdens de continuïteitsperiode wordt verondersteld dat geen afwikkeling plaatsvindt, maar aan het einde van de gekozen horizon, dus op tijdstip T , wordt het pensioenfonds geacht fictief af te wikkelen (afwikkeling III/figuur 6). Er dient daarom op tijdstip t_0 een zodanige financiële opzet te worden geïnitieerd dat op tijdstip T , gegeven een gekozen norm van (gewogen) uitkeringszekerheid, kan worden afgewikkeld. Dit impliceert een benodigd vermogen, een doelvermogen, ten bedrage van de verwachtingswaarde van de op tijdstip T opgebouwde pensioenrechten. Duidelijk zal zijn dat een hoger gekozen zekerheidsniveau een hoger doelvermogen op tijdstip T en dus een hogere pensioenpremie in de continuïteitsperiode impliceert. Evident is dat het doelvermogen en de pensioenpremie in de continuïteitsperiode ook hoger zullen zijn naarmate er bij de afwikkeling vanaf tijdstip T in sterkere mate —in de zin van meer en/of langduriger— met indexering rekening wordt gehouden.

De werkhypothese van continuïteit behelst als element niet alleen de bepaling van een adequaat doelvermogen op tijdstip T en de complementaire bepaling van de pensioenpremie in de continuïteitsperiode, maar ook een eventuele discontinuïteitsopslag op de continuïteitspremie. Dit tweede element heeft betrekking op het risico van een tijdens de continuïteitsperiode optredende, tussentijdse afwikkeling (afwikkeling II/figuur 6). De impact van tussentijdse discontinuïteit wordt getraceerd door middel van “as if”-toetsen, waarbij wordt nagegaan of bij tussentijdse afwikkeling tijdens de opbouwfase voldoende nominale uitkeringszekerheid aanwezig is en in welke mate nog kan worden geïndexeerd. Indien de uitkomsten van deze toetsen een onvoldoende niveau van uitkeringszekerheid opleveren, is een discontinuïteitsopslag op de continuïteitspremie nodig. Of deze opslag daadwerkelijk in de financiële opzet moet worden opgenomen, hangt in principe af van de mate waarin tussentijdse afwikkeling voor het desbetreffende pensioenfonds een voorzien risico vormt.

In het voorgaande is duidelijk geworden dat aan uitkeringszekerheid een prijs is verbonden. Van de zijde van het Actuariel Genootschap wordt mede op grond

²⁰Volledigheidshalve kan nog worden opgemerkt, dat ook kan worden gewerkt met een benodigd vermogen ter grootte van de opgebouwde nominale aanspraken, hetgeen impliceert dat in de premie dan geen component voor indexering is begrepen. Indien deze minimumvariant in de beschouwingen wordt betrokken, wordt de indexeringscomponent in de premie zichtbaar.

hiervan er voor gepleit dat de norm voor het minimumniveau van de gewogen nominale uitkeringszekerheid —met andere woorden, de minimum zekerheid voor de uitkeringen, voortkomend uit de over het verleden opgebouwde nominale pensioen-aanspraken— vanuit het maatschappelijk verkeer, te weten het overleg tussen sociale partners en de overheid, zou dienen te worden bepaald. Hierbij zou aansluiting kunnen worden gezocht bij andere risico's waarvoor de acceptabele zekerheidsmaat maatschappelijk is bepaald. Hierbij kan worden verwezen naar bijvoorbeeld het Deltaplan, dat voorziet in een uitstekende kwaliteit van de dijken in Nederland, maar overstroming onder extreme omstandigheden niet kan uitsluiten, tenzij de dijken tegen enorme extra kosten verder worden verhoogd. Een analoge afweging is, zoals aan de hand van het AG-rapport duidelijk is geworden, ook bij de uitkeringszekerheid van pensioenen aan de orde.

4 Slot

In deze bijdrage hebben we aangegeven dat een adequate bepaling en beoordeling van de solvabiliteit van pensioenfondsen vereist dat een aantal aandachtspunten expliciet(er) aan de orde dienen te komen. In de eerste plaats is het van belang om bij de bepaling van de VPV de reikwijdte van de pensioenverplichtingen expliciet af te stemmen op de beleidsmatig gekozen tijdshorizon van het pensioenfonds. In de tweede plaats is duidelijk geworden dat een consistente, marktconforme waardering van zowel de beleggingen als de pensioenverplichtingen een noodzakelijke voorwaarde vormt voor een adequate bepaling van de solvabiliteit van pensioenfondsen. Bij de bespreking van uitgangspunten het FTK is gebleken dat de PVK voornemens is om een dergelijke consistente, marktconforme waardering voor te schrijven.

Van belang is verder te memoreren dat de solvabiliteit in de toereikendstoets van het FTK wordt genormeerd aan de hand van de dekkingsgraad. Alhoewel de continuïteitstoets van het FTK nog vrijwel volledig door de PVK moet worden uitgewerkt, is er in paragraaf 4.1 op gewezen dat bij de onderhavige toets, naar het zich thans laat aanzien, de vraag centraal staat of niet alleen na 1 jaar maar in alle jaren van de beschouwde lange termijn periode de actuele waarde van de pensioenverplichtingen tenminste volledig is gedekt door de actuele waarde van de beleggingen. Op basis van het in paragraaf 4.2 besproken AG-rapport is gebleken dat het toetsinstrumentarium van het AG is geformuleerd vanuit de continuïteitsdoelstelling van pensioenfondsen. Hierbij wordt als zekerheidsmaat niet de eendimensionele dekkingsgraad, maar de voor de deelnemers en gepensioneerden relevantere driedimensionele uitkeringszekerheid gehanteerd. Belangrijk is te memoreren dat een zelfde dekkingsgraad kan resulteren in verschillende niveaus van uitkeringszekerheid. Verder is van belang dat een, tijdens de door het bestuur van het pensioenfonds gekozen continuïteitsperiode, eventueel optredende discontinuïteit van een pensioenfonds in het toetsinstrumentarium is geïncorporeerd in de vorm van een discontinuïteitsopslag op de continuïteitspremie. Naast informatie omtrent de uitkeringszekerheid beoogt het toetsinstrumentarium, in de vorm van indexeringsindicatoren, informatie te genereren inzake de indexeringsruimte van een pensioenfonds.

Het toetsinstrumentarium van het AG verschaft, in vergelijking met de dekkingsgraad, adequatere informatie inzake de solvabiliteit van pensioenfondsen. Het onderhavige instrumentarium verschaft niet alleen uitgebreidere informatie, maar de informatie is ook transparanter. Hierbij zij nog gememoreerd dat op basis van de onderhavige informatie een duidelijke relatie kan worden gelegd tussen het niveau van uitkeringszekerheid enerzijds en de prijs van deze zekerheid anderzijds. In dit kader doet zich dan wel nog de vraag voor of de norm voor de uitkeringszekerheid de resultante dient te zijn van maatschappelijk overleg tussen overheid en sociale partners, of dat het bestuur van een pensioenfonds deze norm eigenstandig —uiteraard rekening houdend met de preferenties van sociale partners in de desbetreffende onderneming/bedrijfstak— dient te bepalen.

Bijlage 1

1. 1. BEPALING PASSIVA

- (a) Voorziening pensioenverplichtingen (VPV)
 - Rekenrente $\leq 4\%$
- (b) Reserve algemene risico's (Rar)
 - $Rar \cong 5\% \times VPV$
- (c) Reserve voorgenomen pensioenaanpassing (Rvpa)
 - Rvpa \Rightarrow
 - transparant maken voorwaardelijke indexerings-“verplichting”
 - expliciet maken in communicatie van voorwaardelijkheid indexering
 - geen vermogensoverschot, indien bestendige voorwaardelijke indexering onvoldoende is gecureerd
- (d) Reserve beleggingsrisico's (Rbelris)
 - d.1. Zakelijke waarden

$$R_{belris}^{ZW} \geq \max \left\{ \begin{array}{l} \text{benchm}^t - 0,6 \times \text{benchm}^{\text{MAX48MND}}, \\ \text{benchm}^t - 0,9 \times \text{benchm}^{\text{MIN12MND}} \end{array} \right\}$$

d.2. Vastrentende waarden

$$\begin{aligned} R_{belris}^{VRW} &\geq 0,10 \times \text{Actuele waarde vastrentende waarden} \\ &\quad \text{i.g.v. markttrente van 4\%} \\ &\geq 0,05 \times \text{Actuele waarde vastrentende waarden} \\ &\quad \text{i.g.v. markttrente van 5\%} \end{aligned}$$

2. 2. STURINGSMIDDELEN

- (a) Premieniveau
Kostendeekkende premie (= actuarieel & uitvoeringskosten & opslag beleggingsbuffer voor nieuwe beleggingsmiddelen).
- (b) Indexering
Transparante besluitvorming en communicatie inzake indexeringsbeleid

3. 3. MAATREGELEN ONTOEREIKENDE VERMOGENSPOSITIE

- (a) Onderdekking

$$\text{Dekkingsgraad}_I \text{ (DGI)} = \frac{\text{Netto activa}}{\text{VPV} + \text{Rar}} \times 100\% \geq 100\%$$

ofwel

$$\frac{\text{Netto activa}}{\text{VPV}} \times 100\% \geq 105\%$$

Indien $DG_I < 100\%$:

- Bestuur moet onmiddellijk PVK inlichten;
- Er dient binnen 3 maanden een herstelplan inzake de redressering van de onderdekking, met een herstelperiode van maximaal 1 jaar, aan de PVK te worden voorgelegd.

- (b) Reservetekort

$$\text{Dekkingsgraad}_{II} \text{ (DG}_{II}) = \frac{\text{Netto activa}}{\text{VPV} + \text{Rar} + \text{Rbelris}(+\text{Rvpa?})} \times 100\% \geq 100\%$$

Indien $DG_{II} < 100\%$:

- Bestuur moet PVK actief en snel de PVK informeren (ook in geval van dreigend reservetekort);
- Er dient binnen 3 maanden een herstelplan inzake de redressering van het reservetekort, met een herstelperiode van ca. 2-8 jaar en waarbij gedurende de gehele periode voortgang moet kunnen worden getoond, aan de PVK te worden voorgelegd.

Bijlage 2

TOELICHTING (bij tabel op volgende pagina):

Kolom IA : Nominale afwikkeling

De ongewogen zekerheid van 99,6% impliceert dat van de 500 scenario's er 0,4% ofwel 2 scenario's zijn met een tekort. De omvang van de tekorten in deze 2 scenario's vergt een gemiddelde rechtenkorting van 4% en een maximale rechten-Korting van 6%.

Kolom IB : Nominale afwikkeling

Omdat vastrentende waarden beter matchen met een nominale uitkeringstroom

4 situaties van directe afwikkeling voorbeeldfonds

	(IA)	(IB)	(IIA)	(IIB)
Gegevens voorbeeldfonds				
• 4% dekkingsgraad bij afwikkeling	100%	100%	100%	100%
• Beleggingsmix (% vastrentende waarden)	40%	90%	40%	40%
• Gemiddeld reëel rendement	9%	7.1%	4%	4%
UITKOMSTEN:				
<i>1. Uitkeringszekerheid</i>				
• Ongewogen	99.6%	100%	45%	71%
• Gewogen	100%	100%	85%	93%
<i>2. Omvang van overschotten (+) of (-) als % van rechten</i>				
• Minimum tekort	-6%	24%	-85%	-80%
• 1% tekort	6%	34%	-72%	-63%
• 5% tekort	29%	42%	-54%	-40%
• Gemiddeld tekort	-4%	nvt	-27%	-23%
• Gemiddeld overschot	97%	58%	0%	30%
• 1% overschot	182%	74%	63%	112%
• 5% overschot	213%	79%	100%	160%
• Maximum overschot	254%	84%	137%	208%
<i>3. Tekorttijdstip</i>				
• Minimum tekorttijdstip	36	nvt	13	15
• 1% tekorttijdstip	nvt	nvt	16	19
• 5% tekorttijdstip	nvt	nvt	19	24
• Gemiddeld tekorttijdstip	40	nvt	29	32
• Maximum tekorttijdstip	44	nvt	59	55

is er in geval van een beleggingsmix met 90% vastrentende waarden sprake van een nominale uitkeringszekerheid van 100%, hetgeen fractioneel hoger is dan in kolom IA. Hier staat tegenover dat de overschotten zich op een lager niveau bevinden dan in kolom IA. De grotere nominale uitkeringszekerheid gaat dus gepaard met een geringere indexeringscapaciteit. Indien de beleggingsmix van kolom IA dus als going concern mix zou worden gehanteerd, zou dit een hogere continuïteitspensioenpremie impliceren.

Kolom IIA : Afwikkeling met indexering (looninflatie)

In deze constellatie van onvoorwaardelijke indexering is er —alhoewel er sprake is van dezelfde startdekkingsgraad— ten opzichte van kolom IA sprake van een substantieel lagere uitkeringszekerheid. Verder kan worden geconstateerd dat er thans ten opzichte van kolom IA veel grotere tekorten optreden, namelijk gemiddeld -27% ten opzichte van -4% in kolom IA.

Kolom IIA : Afwikkeling met indexering (looninflatie)

De hogere startdekkingsgraad van 130% in plaats van 100% resulteert ten opzichte van kolom IIA in een hogere uitkerings-zekerheid en lagere tekorten. De relatieve tekortreductie blijkt evenwel geringer te zijn dan de relatieve toename van de dekkingsgraad.

Referenties

- Actuariel Genootschap (2001), *Pensioenzekerheid dynamisch getoetst. Aanbevelingen en een werkmodel voor de inrichting van de solvabiliteitsbeoordeling van pensioenfondsen*, Woerden, 1 november 2001.
- Actuariel Genootschap (2003), *Prijs van zekerheid hand in hand met solvabiliteitsbeoordeling, Pilotstudie Pensioenzekerheid dynamisch getoetst*, Woerden, 8 januari 2003.
- Boender, C., S. Van Hoogdalem, M. Vos en H. Steehouwer (2000), "Solvabiliteit en fair value van pensioenverplichtingen", *Economisch Statistische Berichten*, 6 september 2000.
- Klein Haneveld, H. (1999), *Solvabiliteitscriteria voor pensioenfondsen*, Capelle a/d IJssel.
- Klein Haneveld, H. (2001), "De kosten van pensioenen", *Economisch Statistische Berichten*, 8 juni 2001.
- Ponds, E. (2001), "Waardering verplicht", *Economisch Statistische Berichten*, 8 juni 2001.
- Pensioen- en Verzekeringkamer (2001), *De Uitgangspunten voor een Financieel Toetsingskader*, Apeldoorn, september 2001.
- Petersen, C. (1990), *Pensioenen: Uitkeringen, Financiering & Beleggingen*, Leiden/Antwerpen.
- Polfiliet, R. (1999), *Value at Risk, Verlies als maatstaf*, Antwerpen/Apeldoorn.
- Siegelaar, G. (2000), "Van waardering naar toetsing. Over belangenafweging, risicobeoordeling en transparantie bij pensioenfondsen", *Economisch Statistische Berichten*, 6 september 2000.
- Vereniging van Bedrijfspensioenfondsen/Stichting voor Ondernemingspensioenfondsen/Pricewaterhouse N.V. (2000), *Continu op koers*, Utrecht, januari 2000.
- Verzekeringkamer (1997), *Actuariële Principes Pensioenfondsen*, Apeldoorn, februari 1997.

MOBILITEIT EN PUBLIEKE DOELEN GOED VERZOENBAAR

Wim Hafkamp* en Pieter van Driel†

Erasmus Universiteit Rotterdam

De opleiding econometrie stond nog in de kinderschoenen, begin jaren 70. Geen handboeken, maar handgeschreven dictaten, of preciezer: tractaten. Dalmulder en Plasmans c.s. wijdden ons in in de wondere wereld van de modellenbouw, waar alles draaide rond die ene mysterieuze variabele Y . Van De Klundert behandelde in zijn colleges model 69C van het Centraal Planbureau. En Pieter Ruys, jong wetenschappelijk medewerker, gaf aansprekende colleges wiskundige economie. Dat is al weer zo'n 30 jaar geleden. Publieke goederen, externe effecten, social choice, allocatiemechanismen. Het was een even wondere wereld. Mensen werden er aangeduid als 'agents', die zo een Nash-evenwicht indoken, tenzij je als wiskundig econoom hun spelregels veranderde. Als je dat goed deed, met quasi-goederen die niet verhandeld werden op niet-markten, en verpakte in optimal control, dan was het met behulp van het bang-bang theorema van Uzawa mogelijk om de agents tot Pareto-optimaliteit te brengen. Ging het ook ergens over, dat was de vraag waar we als studenten mee worstelden. Menig docent ging de discussie uit de weg, maar Pieter niet. Of het nu om vragen ging op het gebied van onderwijs, werkloosheid of milieu en energie. Toen ik op zoek ging naar een brug tussen de Tilburgse econometrie en de ruimtelijke- en milieu-economie zette hij me op het spoor van Peter Nijkamp, mijn latere promotor. Gedurende twee doctoraal-jaren begon ik wetenschappelijk plezier te krijgen in de spanningsboog die ik zag ontstaan tussen de wiskundige economie en de milieu-economie. Pieter stond voor de 'zuivere' theorie, de precieze formulering, het abstracte. Peter Nijkamp stond voor het empirische, toetsende, modellerende. Zo werd Peter tweede begeleider bij mijn afstudeerwerkstuk. En toen ik in Amsterdam ging werken werd Pieter co-promotor bij mijn proefschrift. In latere jaren heb ik me vooral ontwikkeld in de milieu-economie, de milieukunde en het milieubeleid. Ik ben eigenlijk wel benieuwd wat Pieter Ruys vindt van die structurele decompositie die Pieter van Driel en ik plegen in de onderstaande bijdrage. Na 30 jaar gaat het nog steeds om publieke doelen, hoe ze te stellen, en hoe ze te bereiken.

*Adres: Faculteit der Sociale Wetenschappen, Erasmus Universiteit Rotterdam, Postbus 1738 3000 DR Rotterdam, Nederland. *Email*: hafkamp@fsw.eur.nl.

†Adres: Postbus 30949, IPC 105, 2500 GX Den Haag, Nedrerland. *Email*: pieter.vandriel@minvrom.nl

1 Inleiding

In deze bijdrage willen we bespreken welke publieke doelen de overheid als het gaat om verkeer en vervoer vooral heeft te behartigen, hoe moet ze dat doen en vanuit welke principes. We gaan na in hoeverre de samenhang tussen deze doelen eisen stelt aan te ontwikkelen overheidsbeleid, en doen enkele beleidsvoorstellen. Onze analyse is vooral gericht op personenmobiliteit. We maken daarbij ondermeer gebruik van enkele adviezen van de VROM-raad, over mobiliteit ‘Mobiliteit met Beleid’ (2000), over milieu en economie (binnenkort te verschijnen) en over externe veiligheid (samen met de Raad voor Verkeer en Waterstaat, volgend jaar te verschijnen).

2 Overheidsdoelen

Van de overheid wordt verwacht dat hij onze ‘collectieve’ belangen behartigt - belangen van burgers en de samenleving als geheel waarvan niet in redelijkheid gevraagd kan worden dat andere maatschappelijke actoren ze zelf regelen. Bij verkeer en vervoer gaat het dan om mobiliteit en bereikbaarheid, maar ook om een gezond milieu, veiligheid, ruimtelijke kwaliteit, en continuïteit in de energievoorziening. De overheid kan dat op een aantal manieren proberen te doen. Hij kan partijen ‘organiseren’ om waar nodig samenwerking tot stand te brengen. Hij kan de randvoorwaarden voor het vrije spel vastleggen (regelgeving). Hij kan investeren of investeringen uitlokken. Hij kan actoren verantwoordelijkheden toekennen - rechten op het gebruik van collectieve voorzieningen, maar ook verplichtingen en aansprakelijkheid. Hij kan toezien op eerlijke concurrentie en waken tegen het afwentelen van kosten en risico’s op derden.

Welke weg de overheid ook kiest, telkens zouden de volgende principes zoveel mogelijk in acht moeten worden genomen. Hij moet de grote diversiteit in preferenties erkennen, en mensen zoveel mogelijk vrij laten in hun keuzen van leefstijl. Dat is ook relevant als het gaat om regio’s. Zo kan Zuid-Limburg vanuit andere preferenties andere oplossingsrichtingen zoeken dan de Noordvleugel van de Randstad. Hij moet uitgaan van principes zoals ‘de gebruiker betaalt’, ‘de vervuiler betaalt’, en de actoren niet onnodig op kosten jagen. Bij regelgeving dient hij de betrokkenen zoveel mogelijk ruimte geven daaraan tegen lage kosten te voldoen. En uiteraard moet de overheid zelf als het even kan ook efficiënt optreden, dus innovatief zijn, en uitbesteden als anderen het beter kunnen.

3 Mobiliteitsreductie geen publiek doel

Het streven naar minder mobiliteit heeft vele jaren het ruimtelijk beleid belast. Nu Nederland, vervolgens heel Europa, en zelfs de wereld binnen bereik is gekomen, is het een illusie te denken dat ruimtelijk beleid het aantal (auto)kilometers zou kunnen reduceren. Het type stedelijke ontwikkeling bleek dan ook nauwelijks van invloed op de vervoersstromen en modal split. Veeleer bleek, omgekeerd, de inrichting van het verkeer- en vervoerssysteem (de weg, de afrit, het ov-knooppunt) en

het aanbod van modaliteiten bepalend voor waar mensen heen kunnen, en vervolgens heen willen gaan. De zorgelijkheid over mobiliteitsgroei kan maar beter worden gestaakt, zonder te vervallen in het andere uiterste, 'mobiliteit is fun'. Mobiliteit is een hulpbron net als 'inkomen' en 'energie' waarmee men de meest uiteenlopende doelen kan bereiken. In een globaliserende samenleving is het aantal doelen in feite oneindig; de enige rem is inkomen en tijd (al eeuwen lang wil men niet meer dan gemiddeld ruim een uur per dag in mobiliteit steken - dat zet oude vormen van OV op een onoverbrugbare achterstand t.o.v. de auto). De groei van mobiliteit is dan ook een functie van bevolkingstoename, welvaartsgroei per hoofd en aanbod van snellere vervoermiddelen zoals vliegen en HSL. Zoals het niet zinnig is voor de overheid te streven naar 0-groei en BNP-krimp, zo is het ook niet zinnig te streven naar beperking van mobiliteit¹ noch naar ont koppeling van groei van economie en mobiliteit (een voorstel van de Europese Commissie)². Kernpunt is dat zowel de toegang tot mobiliteit moet worden veilig gesteld, als ook publieke doelen die in de knel komen door mobiliteit in zijn verschillende vormen. Die problemen moeten nauwkeurig benoemd worden, elk heeft z'n eigen schaal, aard, oorzaak en dynamiek, en dus ook z'n eigen aanpak. We lopen de belangrijkste langs.

4 Verkeersveiligheid

Verkeersveiligheid mag gerust één van de meest onderschatte onderdelen van het verkeer- en vervoersbeleid genoemd worden. De immateriele schade valt niet te berekenen; en de economische schade wordt geraamd op ruim 5,5 miljard euro per jaar, en dat is veel meer dan de economische schade door milieuverontreiniging of congestie. Pleit dat voor beperking van mobiliteit? Geenszins. Terwijl afgelopen twintig jaar het aantal voertuigkilometers in ons land steeg met 75%, is het aantal doden bijna gehalveerd van 2000 in 1980 (in 1970: 3200..!) tot minder dan 1100 in het jaar 2000. Verdere 'ontkoppeling' vraagt overduidelijk opnieuw om maatwerk. Zo blijkt de snelweg waar de meeste kilometers gereden worden, relatief het meest veilig. De meeste doden vallen op provinciale wegen, de meeste gewonden binnen de bebouwde kom, en 20% van het aantal doden valt in de leeftijdsgroep der 18-24 jarigen. De overheid moet bij dit maatwerk het voortouw nemen: eisen aan kwalificaties van bestuurders, aan verkeersregels, aan voertuigen, aan het ontwerp van wegen en kruisingen; aan verkeersbeleid in de bebouwde kom en benutting van rand- en rondwegen - dit ongeacht het wel of niet verzelfstandigen van een deel van het wegennet. Ook moeten kwesties van aansprakelijkheid van bestuurders; eigenaren van voertuigen en andere verkeersdeelnemers beter geregeld worden³. Handhaving, voorlichting, inzet van politie en hulpdiensten blijven cruciaal.

¹Een analogie is dat het streven naar minder energiegebruik alleen de energie-efficiency betreft, niet de mate van energiediensten zelf.

²Het Witboek beoogt relatieve ont koppeling: tragere groei van mobiliteit, daarbinnen een verschuiving naar rail.

³De overheid mag best een poging doen uit te leggen waarom onlangs na wetswijziging de aansprakelijkheid gelegd is bij bestuurders van voertuigen na een aanrijding met fietsers/voetgangers.

5 Milieu

Op milieugebied gaat het om verschillende doelen. We beginnen met de risico's verbonden aan het vervoer gevaarlijke stoffen, in de context van steeds intensiever ruimtegebruik en de stedelijke ontwikkeling rond knooppunten van vervoer. De risico's leggen grote beperkingen op aan het ruimtegebruik rond transportassen en knooppunten. Dit indirect ruimtebeslag kan heel goed bij de veroorzakers in rekening gebracht worden: 'gebruiker betaalt'. Ook de beheerders van vervoersnetwerken - rail, weg en water- zouden verantwoordelijkheden moeten krijgen voor de veiligheid met bijbehorende bevoegdheden, bijvoorbeeld op het gebied van routing, quoter-ing, tijdsvensters, en maximum-snelheden. Eén van de belangrijke risico's betreft transport van LPG. Om die te verminderen geldt een zoneringsbeleid rond wegen. Dat gaat gepaard met hoge kosten voor overheid en samenleving. Indien hier 'de gebruiker moet betalen' dan verdient de fiscale bevoordeling van LPG - ooit ondermeer gerechtvaardigd door milieuvordelen; maar nu niet meer - heroverweging.

De overige milieu-aspecten van mobiliteit hebben vooral te maken met uitlaatgassen, geluidshinder, en effecten van infrastructuur op natuur en landschap. De VROM-raadsadviezen benadrukken hoe belangrijk het is om een onderscheid te maken naar het ruimtelijk schaalniveau waarop milieu-effecten zich voordoen.

Bij NO_x, en bij emissies die de gezondheid bedreigen - overwegend veroorzaakt door de korte ritten in het stedelijk gebied - heeft Europese regelgeving en technische vooruitgang de oplossing in zicht gebracht. Hoewel tussen 1980 en 2010 het aantal personen-voertuigkilometers bijna verdubbeld zal zijn, zullen in deze periode vrijwel alle emissies in totaal afnemen met 80% of meer - ruwweg factor 4⁴, en die daling zal na 2010 doorgaan totdat heel het wagenpark voldoet aan de normen die vanaf 2005 voor nieuwe auto's gelden. Intussen wordt hard gewerkt aan nieuwe concepten die wat betreft gebruik in stedelijk gebied als nul-emissie-auto kunnen worden beschouwd: hybride auto's, auto's op waterstof of perslucht. Tegen deze achtergrond is modal shiftbeleid - dat vooral perspectief heeft in stedelijk gebied - om milieuredenen vrijwel niet meer van belang (veel belangrijker zijn schonere bussen); wel om andere redenen zoals hinder en congestie. Wat er in dichtbevolkte centra nog rest aan emissies met gezondheidsrisico's, moet, in combinatie met een adequate inrichting van verkeer en vervoer in de bebouwde kom als overwinbaar beschouwd worden. De geluidhinder, die beduidend minder sterk afnam, lijkt door een nieuw type banden en fluisterasfalt beheersbaar te blijven; en ook hier wijst Europese regelgeving de weg.

Van de CO₂-emissies, die nog wel elk jaar toenemen, komen we minder makkelijk af. Het grootste deel wordt veroorzaakt tijdens lange ritten en het interlokaal verkeer. Daarom kunnen opties zoals lokaal beleid, ruimtelijk beleid en meer openbaar vervoer, die vooral op de schaal van de (netwerk)stad relevant zijn, hier weinig betekenen⁵, temeer omdat het verschil in energieprestatie van stedelijk OV en auto

⁴afname 1980-2010, NO_x, CO, VOS, PM10 : 80% of meer. Bron: RIVM, Verkeer en vervoer in de Nationale Milieuverkenning 5, rapport 408129014, p. 107 e.v. Op Europese schaal is nog fikse normgeving en handhaving nodig: grote steden in m.n. Zuid-Europese landen kampen met veel ernstiger problemen dan die in Noord-West Europa.

⁵Leg wat dat betreft die rondweg dus maar aan: goed voor veiligheid, goed voor bereikbaarheid, goed voor binnenstedelijke rust.

buiten de spits beperkt is, en meer OV over het algemeen niet minder autoverkeer betekent. Zelfs als het zou lukken (de groei van) het aantal stedelijke autokilometers significant te beperken, dan heeft dat op de totale CO₂-uitstoot weinig betekenis; en gevreesd moet worden dat dit geringe resultaat afgaande op de Breverwet ongedaan gemaakt zal worden door een toename van mobiliteit naar bestemmingen elders, in binnen- en buitenland. Hoewel alle beetjes helpen, is de kosteneffectiviteit van dit type aanpak niet erg opwekkend.

Klimaatbeleid vraagt in deze sector dan ook om een andere aanpak. Gelukkig is bij dit milieuvraagstuk de beleidsruimte groot, want voor de mondiale atmosfeer maakt het niet uit in welke sectoren en landen de reducties gerealiseerd worden, en voor drastische energiebesparing en de omschakeling naar een koolstofarme economie is veel tijd beschikbaar. Om op een kosteneffectieve manier aan 'Kyoto' te voldoen is beleid nodig minstens op Europese schaal, met zoveel mogelijk gebruik van economische instrumenten. Dan kunnen dwars door alle sectoren heen die beschikbare mogelijkheden voor reductie worden geselecteerd die het minste kosten (vgl. rapport commissie Voigtlander). In het verkeer, waar het gebruik van fossiel brandstof - lees uitstoot van CO₂ - al zeer zwaar wordt belast, zijn voorlopig weinig kosteneffectieve reducties te vinden. Er kunnen zoveel reducties tegen lagere kosten buiten de verkeerssector gerealiseerd worden⁶, dat tot 2010 de emissies in het personenverkeer zelfs nog met 20% kunnen stijgen. Dat geeft respijt voor een krachtig Europees innovatiebeleid om op termijn voor verkeer en vervoer kosteneffectieve CO₂-reducties te kunnen realiseren, waarbij het convenant van de Europese Commissie met autoindustrie een belangrijke eerste stap is⁷. Bij zo'n strategie wordt voldaan aan het principe dat de overheid actoren niet onnodig op kosten jaagt en mensen zoveel mogelijk vrij moet laten in hun keuzen van leefstijl, ook wat betreft hun voorkeuren voor (auto)mobiliteit.

Deze zoektocht naar een andere brandstoftechnologie is ook opportuun om geopolitieke overwegingen en omdat goed toegankelijke oliebronnen vanaf 2010-20 wel eens geleidelijk minder rijkelijk zouden kunnen gaan vloeien. Voor het overige is de ooit gevreesde generieke uitputting van niet-vernieuwbare natuurlijke hulpbronnen niet zo relevant gebleken⁸, noch voor continuïteit in verkeer en vervoer, noch in het algemeen.

6 Ruimtelijke kwaliteit en bereikbaarheid

De bereikbaarheid is een publiek doel waarmee het zeker niet goed gaat. Gezien het voorgaande is er wat de milieuhygiënische, de 'grijze' kant, betreft eigenlijk geen beletsel om wegen, knooppunten en parkeerplaatsen bij te bouwen. Maar ook

⁶Economic evaluation of sectoral reduction objectives in the EU; P. Capros, N. Kouvaritakis, L. Mantzos. Een studie voor de Europese Commissie, DG Milieu (http://europa.eu.int/comm/environment/enveco/climate_change/sectoral_objectives.htm). Zie ook: M.Harmelink, K. Blok, D. de Jager, C. Hendriks, Emissiereductie: welke gassen en sectoren? ESB-Dossier Klimaatbeleid, 11 okt. 2001.

⁷Afgesproken is dat het energieverbruik voor nieuwe auto's gemiddeld in 2010 per km 25% lager zal liggen.

⁸Dit wordt besproken in het binnenkort te verschijnen advies van de VROM-raad over 'milieu en economie'.

de 'groene' kant, of woonkwaliteit en welzijns-kant van de milieuproblematiek is belangrijk. Wegenbouw, door aanleg of verbreding, met uitbreiding van knooppunten, en de onvermijdelijke aanpalende bebouwing (van benzinstations en wegresterants, tot keuken-, licht- en autopaleizen op 'zichtlocatie') zijn een aanslag vooral op leefkwaliteit in en rond de steden, en hier en daar ook op het landschap, m.n. de openheid ervan⁹. Hoewel elk groot project tegenwoordig wel te maken heeft met de habitat-richtlijn van de EU, staat niet zozeer de biodiversiteit op het spel, als wel de habitat van Nederlanders die openheid, vergezicht, rust en groen waarderen (in contrast met het stedelijk gebied). De kwaliteit van de leefomgeving is een evident publiek doel van groot belang, ook al is dat voor de overheid niet eenduidig vast te stellen, en moet er ook hier zeker ruimte voor pluriformiteit zijn. In de ogen van veel mobilisten zijn zichtlocaties verworpen tot 'afzichtslocaties', een variant op ongewenst drukwerk. Een nieuwe mobiliteitsethiek zou hier veel goed kunnen maken¹⁰.

Om het doel van bereikbaarheid te dienen moet dan ook uiterst behoedzaam worden omgegaan met bijbouwen van wegcapaciteit. Vanuit een afweging van publieke doelen hoort een sterke voorkeur uit te gaan naar een betere benutting van de capaciteit van bestaande wegen en rijstroken door verkeersgeleiding, vergroting van de capaciteit op bestaand asfalt door dynamische rijstroken e.d. Dat zal de schaarste echter niet wegnemen, zeker niet gezien de verwachte groei van het verkeer in de stedelijke netwerken. Is toename van asfalt dan toch nodig, dan liefst binnen bestaand ruimtebeslag door ondergronds, bovengronds of anderszins 'slim' te gaan. De aanwijzing van nieuwe tracs, met de daarbijbehorende slopende procedures is dan niet nodig. Hoe dan ook: een vergroting van de capaciteit heeft alleen voordeel op locaties waar de baten van uitbreiding echt opwegen tegen de hoge maatschappelijke kosten.

Daarop krijgen we echter geen zicht bij het huidige systeem van fiscale regelingen waarbij de automobilist, gegeven het energiegebruik van z'n auto, evenveel betaalt per kilometer ongeacht waar en wanneer hij rijdt, en met welke effecten hij de plaatselijke omgeving belast. Dat is niet eerlijk en genereert ook geen prikkel tot vermindering van die maatschappelijke kosten. Internationaal groeit het inzicht dat kosten en baten van vervoer en de vraag naar en aanbod van schaarse infrastructuur beter op elkaar afgestemd moeten worden. Dat vraagt om inzet van instrumenten als de kilometerheffing naar tijd en plaats¹¹ die de effecten op de leefkwaliteit in en

⁹Het ruimtebeslag door landbouw en grootschalige landbouwtechnieken is veel relevanter. De verkeersinfrastructuur beslaat 4% van grondoppervlak, en ruim 3% van het totale oppervlak van ons land. De toename sinds 1981 beslaat 0,15% van het totale grondoppervlak. Tussen 1986-96 kwam er bijna 16.000 km verharde weg bij - voor een groot deel in samenhang met nieuwe woonwijken -, en 200 km snelweg. Het effect van de infrastructuur op het landschap wordt vooral bepaald door de fijnmazigheid, de uitstraling van verkeerslawaai e.d. en minder door de wegbreedte.

¹⁰Vg. Mobiliteit met Beleid, H. 5.

¹¹Getting prices right. Results from the transport research programme. European Commission Commission. <http://europa.eu.int/comm/transport/extra/home.html>. Deze brochure geeft een beknopt overzicht van een groot aantal onderzoeksprogramma's van de laatste jaren, w.o. Trenen II, STRAN, Quits, Afford, PATS, PRIMA, Capri, PETS, Imprint, Unite. Zie ook: Why reform transport prices, A review of European Research, C. Nash, B. Matthews, 2001; alsmede een aantal landenstudies zoals voor Engeland (Surface Transport Costs & Charges, Sansom, Nash e.a., Leeds 2001), voor België (External Costs of Transportation, Mayeres, Proost, Int Panis, De Nocker e.a.,

rond de steden goed weerspiegelt. Diverse recente studies, met name in opdracht van de Europese Commissie, laten zien dat de prijs die nu in het stedelijk verkeer betaald wordt veel lager ligt dan wat men per kilometer aan maatschappelijke kosten veroorzaakt. Daarentegen betaalt de automobilist in het interstedelijk verkeer bij de huidige tarieven, en rekening houdend met de vaste lasten, iets meer dan de maatschappelijke kosten¹². Er is dus een ingrijpende herziening nodig van de lastendruk op mobiliteit. Dit wordt een complexe operatie, allereerst omdat de huidige fiscale behandeling van mobiliteit een lappendeken is. Met of zonder ‘kwartje van kok’, Nederlanders begrijpen het al lang niet meer. Ze wantrouwen de overheid, en elk voorstel dat ruikt naar lastenverzwaring. Het huidige stelsel (BPM, MRB, BTW, accijns, etc.) weerspiegelt een genealogie van rationales uit de tweede helft van de vorige eeuw. Het vergt een forse politieke en maatschappelijke discussie om een nieuwe rationale te creëren. In dat licht is het niet erg dat het kabinet besloten heeft om de invoering van de kilometerheffing uit te stellen. Net als bij de ‘Oort-Stevens-operatie’ voor de loon- en inkomstenbelasting zullen hiervoor jaren nodig zijn; maar is wel zaak hiermee een begin te maken, waartoe ook de Europese commissie oproept¹³.

Bij een herziening van de lastendruk op mobiliteit zal eens te meer blijken hoe belangrijk modern openbaar vervoer is voor stedelijke kwaliteit en bereikbaarheid, vooral op strategische verbindingen. Bereikbaarheid is ook een kwestie van samenwerking tussen verkeerssysteem en ruimtelijke ontwikkeling: de overheid moet investeren in OV niet om redenen van ‘milieu’ maar als voorwaarde voor een gewenst type stedelijke ontwikkeling. De van oudsher geldende sociale doelstelling hoeft daarbij niet uit het oog te worden verloren, maar vraagt veeleer om een andere wijze van financiering. Enkele voorbeelden: regionale bereikbaarheidsheffingen en subsidiëring openbaar vervoer langs de vraagkant in plaats van de aanbodkant.

7 Conclusies

Door stapeling van publieke doelen is de complexiteit van de sturingsvraag lange tijd overschat. Veilig, schoon en fluide zijn echter goed te scheiden publieke doelen. Bezinning daarop laat zien dat de overheid over voldoende instrumenten beschikt om de publieke zaak ook op dit gebied adequaat te dienen. Veiligheid is nog steeds zwaar onderschat, de overheid heeft echter geen gebrek aan instrumenten. Externe veiligheid in het bijzonder verdient meer aandacht ondermeer door wegbeheerders verantwoordelijkheden te geven, en bevoegdheden. De milieuhygiëne blijkt met emissie-eisen oplosbaar. Knelpunt is hier vooral de CO₂-emissie. Daarvoor is echter juist geen behoefte aan apart beleid per sector, maar aan een kosteneffectieve aanpak op systeemniveau gericht op de transitie naar een koolstofarme energievoorziening, waarbij gelijktijdig gewerkt wordt langs twee gescheiden routes: verhandelbare emisierchten (ev. koolstofheffing) en innovatiebeleid. Ook de congestie valt aan te

2001), voor Nederland (J.Dings e.a., Efficiente prijzen voor het verkeer, Delft 1999).

¹²How large is the gap between present and efficient transportprices in Europe, J.v.d. Bergh, e.a., Leuven, 2001 (<http://www.econ.kuleuven.ac.be/ew/academic/energimil/>).

¹³Witboek, Het Europese vervoersbeleid tot het jaar 2010: tijd om te kiezen, Brussel 2010 (http://europa.eu.int/comm/dgs/energy_transport/).

pakken langs twee routes: langs de prijzenkant richting kilometerheffing naar tijd en plaats, en langs de aanbodskant vooral door verbeterde benutting van capaciteit, op diverse innovatieve manieren, in welke, w.b. de infrastructuur, nieuwe eigendoms- of beheersconstructie dan ook. Er is volop aanleiding om de systematiek van fiscale tarieven in het verkeer grondig te herzien en zodanig opnieuw op te bouwen dat Nederlanders weer weten wanneer ze wat betalen, en waarvoor, met adequate incentives om maatschappelijke kosten te reduceren, in plaats van ze af te wentelen. Een budget-neutrale operatie herziening lastendruk mobiliteit.

DISCRETE TIJD BENADERING VAN ECONOMIEËN IN CONTINUË TIJD*

Cees Withagen[†]

Vrije Universiteit en Universiteit van Tilburg

Begin 2002 ontving ik een uitnodiging tot het schrijven van een kort essay ter gelegenheid van het afscheid van Pieter Ruys. Zonder ook maar even na te denken heb ik de uitnodiging aanvaard. Vanzelfsprekend moet Pieter Ruys met een *liber promovendorum* geëerd worden voor de betekenis die hij voor de Nederlandse economiebeoefening, en in het bijzonder de wiskundige economie, heeft gehad. Daarnaast heeft Pieter in mijn professionele leven een belangrijke rol vervuld. In het eerste jaar van mijn studie aan de toenmalige Katholieke Hogeschool doceerde Pieter Ruys statistiek; voor colleges aan het handjevol wiskundig economen was er Dalmulder. Pieter was niet rechtstreeks bij mijn afstuderen betrokken, omdat de promotiewerkzaamheden veel tijd van hem vergden. Maar wel herinner ik me enkele tochten met Pieter naar Leuven/Louvain-La-Neuve in de wankele Ami van Claus Weddepohl, waar ik kennis mocht maken met grootheden zoals Drèze en Barten. Na mijn afstuderen werd ik al snel werkzaam op het gebied van de wiskundige economie, aan de Universiteit van Amsterdam. Nadat Pieter in 1980 benoemd was aan de Technische Universiteit Eindhoven, ben ik daar in 1982 aangesteld. Een periode van nauwe samenwerking volgde. In 1984 was Pieter, met Claus Weddepohl, mijn promotor. Sinds enkele jaren verblijf ik weer part-time aan de Universiteit van Tilburg en kom Pieter dus regelmatig tegen.

Waarover moet zo'n essay gaan? De opdracht was om zoveel mogelijk thema's uit het werk van de aanstaande emeritus als leidraad te nemen. Wat is kenmerkend voor de inhoud van het werk van Pieter? Een aantal steekwoorden valt me gemakkelijk te binnen: publieke goederen, algemeen evenwicht, diensteneconomie. Methodisch stonden Pieter vaak nieuwe vormen van aanpak voor ogen: dualiteit en convexe analyse speelden een belangrijke rol. En het ontbrak nooit aan weidse vergezichten. Het zal niet meevallen al deze kenmerken in een enkele bijdrage te verenigen. Wat de filosofische aspecten betreft laat ik het helemaal afweten. De aanpak die ik zal volgen, is qua methode wel onconventioneel. Ik zal een bijdrage leveren aan algemeen evenwichtsmodelleren en laten zien dat dualiteit een belangrijk en handig begrip is in het oplossen van evenwichtsvraagstukken. Dat geschiedt aan de hand van een model geïnspireerd door de milieueconomie, mijn huidige werkterrein. Het

*Dit artikel is gebaseerd op werk dat gedaan is met dr. Jan van Geldrop, verbonden aan de Technische Universiteit Eindhoven.

[†]Adres: Departement Algemene Economie, Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg. *Email*: cwithagen@feweb.vu.nl.

gaat om de vraag of er een algemeen evenwicht bestaat in een economie waarin het gebruik van een grondstof uit een natuurlijke hulpbron van belang is voor het nationaal product. Bij natuurlijke hulpbronnen gaat het vaak over een lange horizon, hetgeen aanleiding geeft tot oneindig dimensionale goederenruimtes. Dat levert in wiskundig economisch opzicht een interessante problematiek op. Uit puur economisch oogpunt is het model wellicht niet zo interessant, omdat er geen sprake is van substituten, technische vooruitgang en omdat de technologie die verondersteld wordt, op voorhand uitsluit dat de economie duurzaam is. Maar dat biedt wel inzicht in de voorwaarden waaraan voldaan moet zijn voor duurzame ontwikkeling. En de gehanteerde methodologie zal een breder toepassingsgebied kunnen vinden. Het bestudeerde model is in continuë tijd voor een oneindige horizon. Wat we doen is eerst te kijken naar een economie, die in essentie analoog is, maar die voor een eindige horizon en discrete tijd is gedefinieerd. Daarna wordt de limiet van deze economie behandeld voor een oneindige horizon. Tot slot wordt de limiet van deze laatste economie bekeken voor steeds kleinere tijdsintervallen.

1 Het model

De beschrijving van een algemeen evenwichtsmodel bevat standaard de volgende elementen:

- goederen
- consumenten
- producenten

Dat geldt ook voor het model dat ik wil analyseren. Ik begin dus met de *goederen*. Het model is dynamisch van aard, met $t = 0$ als het begintijdstip. Er zijn twee *voorraadgrootheden*. Ten eerste is er een voorraad kapitaal, aangeduid met k . De voorraad aan het begin is $\omega_1 > 0$. Ten tweede is er een voorraad van een niet-vernieuwbare natuurlijke hulpbron. De beginvoorraad hiervan is $\omega_2 > 0$. De voorraden kunnen in de loop van de tijd veranderen. Kapitaal kan toenemen als gevolg van positieve netto investeringen, kapitaal neemt af als de netto investeringen negatief zijn; de niet-vernieuwbare natuurlijke hulpbron kan per definitie alleen maar afnemen. De bruto investeringen op tijdstip t bestaan uit de netto investeringen $\dot{k}(t)$ ($= dk(t)/dt$) en de afschrijvingen, die een vast deel μ van de bestaande voorraad vormen. De laatste zijn *stroomgrootheden*, evenals de winning uit de natuurlijke hulpbron, die wordt aangeduid met v . Kapitaal kan worden ingezet voor twee doeleinden. In de eerste plaats in een productieproces dat consumptiegoederen voortbrengt, in de tweede plaats in de winning van de grondstof uit de natuurlijke hulpbron.

De consumptiekant van de economie is eenvoudig. Er zijn twee alternatieve wijzen van aanpak. De eerste is aan te nemen dat er een enkele representatieve consument bestaat die oneindig lang leeft. De tweede aanpak is dat er een oneindig aantal consumenten is, die elk een enkel moment leven en zich als een dynastie gedragen. Wiskundige gezien is het onderscheid niet relevant. In het vervolg zal

voor de eerste interpretatie gekozen worden. De consument heeft de beginvoorraden kapitaal en niet-vernieuwbare hulpbron in zijn bezit. Hij heeft initieel geen voorraad van het consumptiegoed. De welvaart van de consument op een bepaald moment hangt niet af van de voorhanden voorraden, noch van de hoeveelheid die uit de natuurlijke hulpbron wordt gewonnen, maar wel van het consumptiegoed. Het instantane nut van de consument wordt weergegeven door een nutsfunctie $u: \mathbb{R}_+ \rightarrow \mathbb{R}$ met de volgende eigenschappen:

- U1.** u is continu op \mathbb{R}_+ en tweemaal continu differentieerbaar op \mathbb{R}_{++} ;
- U2.** u is strikt stijgend en strikt concaaf;
- U3.** $u'(x) \rightarrow \infty$ indien $x \downarrow 0$;
- U4.** Er is een $\eta > 0$ zodat voor alle $c > 0$ geldt $-u''(c)c/u'(c) = \eta(c) \geq \eta$.

Hierin geven accenten afgeleiden aan en definiëren we $\mathbb{R} = (-\infty, \infty)$ als de ruimte der reële getallen, $\mathbb{R}_+ := [0, \infty)$ en $\mathbb{R}_{++} := (0, \infty)$. De veronderstellingen komen erop neer dat de nutsfunctie “mooi” is. Een voorbeeld van een nutsfunctie die aan de veronderstellingen voldoet, is de Bernoulli functie $u(c) = c^{-\eta}$ met $\eta > 0$. Tenslotte nemen we aan dat de consument een positieve tijdsvoorkeurvoet heeft, aangegeven door $\rho > 0$.

Wat de *productiekant* van de economie betreft, wordt er ook een aantal vereenvoudigende veronderstellingen gemaakt. De eerste betreft de exploitatie van de natuurlijke hulpbron. De hoeveelheid kapitaal die nodig is om een eenheid grondstof te winnen, is een constante, die aangeduid wordt met α . De productie van het consumptiegoed vindt, zoals gezegd, plaats met behulp van kapitaal en de gewonnen grondstof. De productiefunctie wordt aangeduid met $F: \mathbb{R}_+^2 \rightarrow \mathbb{R}_+$ en heeft de volgende eigenschappen:

- F1.** F is continu op \mathbb{R}_+^2 en continu differentieerbaar op \mathbb{R}_{++}^2 ;
- F2.** F is strikt stijgend en strikt concaaf op \mathbb{R}_{++}^2 ;
- F3.** $F_2(x, v) \rightarrow \infty$ voor alle $x > 0$ en $v \downarrow 0$;
- F4.** $F(0, v) = F(x, 0) = 0$ voor alle $(x, v) \in \mathbb{R}_+^2$;
- F5.** F is lineair homogeen;
- F6.** Er bestaan k en v zodanig dat $F(k - \alpha v, v) > \mu k$.

In eigenschap F2 staat F_2 voor de afgeleide van F naar het tweede argument, de grondstof; F_1 is op analoge wijze gedefinieerd. De grondstofsector en de sector die het consumptiegoed produceert, zullen als één enkele sector worden behandeld. Dat gebeurt zonder verlies van algemeenheid en vereenvoudigt de notatie. Er vindt in de voorraden alleen handel plaats aan het begin, dus op $t = 0$. De productieverzameling Y van de economie valt nu als volgt te karakteriseren.

$$y = (-y_1, -y_2, z) \in Y \text{ dan en slechts dan als}$$

- a. $(y_1, y_2) \in \mathbb{R}_+^2$ en $z(t) \in \mathbb{R}_+$ voor alle $t \geq 0$,
- b. er bestaan absoluut continuë $k: \mathbb{R}_+ \rightarrow \mathbb{R}_{++}$ en Lebesgue integreerbare $v: \mathbb{R}_+ \rightarrow \mathbb{R}_+$ zodanig dat
- i. $k(t) \geq \alpha v(t)$ voor alle $t \geq 0$.
 - ii. $\dot{k}(t) = F(k(t) - \alpha v(t), v(t)) - \mu k(t) - z(t)$ voor alle $t \geq 0$ met $k(0) = y_1$.
 - iii. $\int_0^\infty v(s) ds \leq y_2$.

Er wordt dus, zonder verlies van algemeenheid, gedaan alsof het bedrijf bij aanvang voorraden kapitaal en natuurlijke hulpbron aanschafte en daarmee een productieplan voor de hele toekomst vastlegt, dat inhoudt dat de productie gelijk is aan de finale afzet naar de consumenten plus de bruto investeringen (b-ii) en de totaal gewonnen hoeveelheid uit de natuurlijke hulpbron niet meer is dan er initieel is aangeschaft (b-iii).

Heeft deze economie een algemeen evenwicht? Alvorens het evenwicht nader te omschrijven, moet worden afgesproken wat de goederenruimte is. Dat is enigszins gecompliceerder dan gebruikelijk, om twee redenen. In de eerste plaats is er een oneindige tijdshorizon. In de tweede plaats willen we in continuë tijd werken, hetgeen formeel impliceert dat ook voor een eindige tijdshorizon de goederenruimte oneindig dimensionaal is. De veronderstellingen over de productiefunctie en de winning van de grondstof geven echter een indicatie van de ruimte waarin consumptiepaden zullen moeten liggen. Vanwege de homogeniteit van F geldt voor $v > 0$:

$$\frac{1}{v} [F(k - \alpha v, v) - \mu k] = F\left(\frac{k}{v} - \alpha, 1\right) - \mu \frac{k}{v} = \frac{k}{v} \left[F\left(1 - \alpha \frac{v}{k}, \frac{v}{k}\right) - \mu \right]$$

De uitdrukking in het linkerlid is van boven begrensd. Immers, indien dat niet het geval zou zijn, is k/v onbegrensd en is er een tegenspraak vanwege F4 (het meest rechtse lid wordt negatief). Er is, vanwege (b-iii) in de definitie van de productieverzameling en de eis dat een evenwichtsallocatie toegelaten is, dus een $\beta > 0$ zodanig dat

$$\int_0^\infty [F(k(s) - \alpha v(s), v(s)) - \mu k(s)] ds \leq \omega_1 + \beta \omega_2.$$

Aangezien in een evenwicht de netto productie gelijk moet zijn aan de consumptie plus de netto-investeringen, volgt uit bovenstaande ongelijkheid dat de totale consumptie in de economie van boven begrensd is. Het is dus 'redelijk' om de consumptieruimte te beperken tot de ruimte van positieve Lebesgue integreerbare functies, aan te duiden met L_+^1 . De goederenruimte wordt derhalve $\mathbb{R}^2 \times L_+^1$. De consumptieruimte is dan $X = 0 \times L_+^1$. De preferentierelatie wordt gerepresenteerd door

$$U(x) = U(0, 0, c) = \int_0^\infty e^{-\rho s} u(c(s)) ds.$$

De *marktprijzen* zijn q_1 en q_2 voor kapitaal en de hulpbron, die beide op $t = 0$ verhandeld worden. De prijs voor het consumptiegoed wordt aangeduid met p . Dit

is een functie van de tijd omdat de consumptiegoederen gedateerd zijn. We zullen eisen dat de prijs (functie) p begrensd is, dus in L_+^∞ ligt. De complete prijsvector is $\pi = (q_1, q_2, p)$.

(x, y, π) heet een **algemeen evenwicht** als $x \in X$, $y \in Y$, $\pi \in \mathbb{R}_+^2 \times L_+^\infty$ en

- i. $x \leq y + \omega$ en $\pi \cdot (x - y - \omega) = 0$ (marktevenwicht)
- ii. $\bar{y} \in Y \implies \pi \cdot \bar{y} \leq \pi \cdot y = 0$ (winstmaximalisatie)
- iii. $\bar{x} \in X$, $\pi \cdot \bar{x} \leq \pi \cdot \omega \implies U(\bar{x}) \leq U(x)$ (nutsmaximalisatie)

We hanteren de conventie $\pi \cdot y = -q_1 y_1 - q_2 y_2 + \int_0^\infty p(s)z(s) ds$.

In het vervolg van dit essay wordt aangetoond dat de bovenbeschreven economie een algemeen evenwicht bezit. Daartoe staan in principe meerdere wegen beschikbaar. Er bestaat literatuur over economieën met oneindig dimensionale goederenruimtes. (Zie bijvoorbeeld Mas-Colell en Zame (1991) voor een overzicht.) Het probleem is echter dat in die literatuur veronderstellingen worden gemaakt die in strijd zijn met de veronderstellingen die we met betrekking tot F en U willen opleggen. Het gaat met name om de begrensdheid van het marginaal product van productiefactoren en het marginale nut van consumptie. Daarom zijn we gedwongen een alternatieve aanpak te kiezen. Deze wordt hieronder uit de doeken gedaan, in drie stappen. Eerst wordt gekeken naar een analoge economie met een eindige horizon en discrete tijd (sectie 2), dan naar een economie met discrete tijd met een oneindige horizon (sectie 3) en tenslotte keren we in sectie 4 terug naar de oorspronkelijke economie.

2 Discrete tijd, eindige horizon

De eerste stap in de bewijsvoering is te kijken naar een economie die analoog is aan de economie die hierboven is beschreven, maar die is gedefinieerd voor discrete tijd en voor een eindige horizon. Deze economie is eindig dimensionaal, waardoor de bekende resultaten uit de standaard algemeen evenwichtstheorie kunnen worden toegepast.

De eindige economie werkt voor $T + 1$ perioden. Perioden worden aangeduid met $t = 0, 1, 2, \dots, T$. De lengte van een periode is ϑ . Een aantal ingrediënten van het model moeten opnieuw worden gedefinieerd.

De goederenruimte is $\mathbb{R}^2 \times \mathbb{R}_+^{T+1}$, de consumptieverzameling is $X(T, \vartheta) = 0 \times \mathbb{R}_+^{T+1}$ en de beginvoorraden zijn $\omega = (\omega_1, \omega_2, 0, 0, \dots, 0)$. De preferentierelatie wordt gerepresenteerd door

$$U(x) = U(0, 0, c_0, c_1, \dots, c_T) = \sum_{t=0}^T \left(\frac{1}{1 + \vartheta\rho} \right)^t \vartheta u(c_t).$$

De productieverzameling is $Y(T, \vartheta)$ met

$$y = (-y_1, -y_2, z_0, z_1, \dots, z_T) \in Y(T, \vartheta) \text{ dan en slechts dan als}$$

- a. $(y_1, y_2) \in \mathbb{R}_+^2$, $z_t \in \mathbb{R}_+^{T+1}$ voor alle $t = 0, 1, 2, \dots, T$.
- b. Er bestaan niet-negatieve rijen k_0, k_1, \dots, k_{T+1} en v_0, v_1, \dots, v_T zodanig dat
 - i. $k_t \geq \alpha v_t$ voor alle $t = 0, 1, \dots, T$;
 - ii. $k_{t+1} = k_t + \vartheta F(k_t - \alpha v_t, v_t) - \mu \vartheta k_t - \vartheta z_t$ voor alle $t = 0, 1, \dots, T$ met $k_0 = y_1$;
 - iii. $\vartheta \sum_{t=0}^T v_t \leq y_1$.

De prijzen zijn: $\pi = (q_1, q_2, p_0, p_1, \dots, p_T)$. De interpretatie van de variabelen spreekt voor zich: c_t bijvoorbeeld is de consumptievoet in het interval dat loopt van ϑt tot $\vartheta(t+1)$. De totale consumptie in het interval is dan ϑc_t .

Voor deze economie laat zich een algemeen evenwicht op de standaard manier definiëren. Dankzij de convexiteit van de preferentierelatie, de convexiteit van de productieverzameling en de begrensdheid van de verzameling van toegelaten allocaties bestaat er een evenwicht. Grootheden worden in het evenwicht aangeduid met T als bovenindex. De parametrisatie met betrekking tot de lengte van de periodes blijft voorlopig achterwege. Het algemeen evenwicht heeft enkele eigenschappen die in het vervolg benut zullen worden. De prijzen van het kapitaalgoed (q_1^T) en de consumptiegoederen (p_t^T) zijn alle positief. In het vervolg nemen we het kapitaalgoed, en daarmee de consumptie in de eerste periode, als de numéraire: $q_1^T = p_0^T = 1$. Deze prijzen zijn in een evenwicht nooit gelijk aan nul, omdat er, gezien de nutsfunctie, altijd positieve consumptie zal plaatsvinden. In elke periode is het gediscanteerde instantane nut uiteraard evenredig met de prijs van het consumptiegoed:

$$\left(\frac{1}{1 + \vartheta \rho} \right)^t u'(c_t^T) = \varphi^T p_t^T, \quad t = 0, 1, \dots, T.$$

Voor het vervolg is van groot belang dat bepaalde grootheden uniform begrensd zijn, dat wil zeggen dat de grenzen onafhankelijk zijn van het aantal periodes waarop de economie gedefinieerd is. Dit heeft in de eerste plaats betrekking op de totale productie, de kapitaalgoederenvoorraad en de totale consumptie die in de economie tot stand kunnen worden gebracht. Dit resultaat is analoog aan de begrenzing die we reeds in de oorspronkelijke economie hadden gevonden. Meer precies kan er aangetoond worden dat er een $\beta > 0$ bestaat, die onafhankelijk van T is, zodanig dat voor alle T geldt

$$k_{t+1}^T + \vartheta \sum_{s=0}^T z_s^T \leq \omega_1 + \beta \omega_2, \quad t = 0, 1, \dots, T.$$

Verder is de inkomensmultiplier φ^T uniform begrensd. Dat valt eenvoudig als volgt

in te zien. Definieer $\sigma = (\omega_1 + \beta\omega_2)/\vartheta$. Dan

$$\begin{aligned} \sum_{t=0}^{\infty} \vartheta \left(\frac{1}{1+\rho\vartheta} \right)^t u(\sigma) &\geq \sum_{t=0}^T \vartheta \left(\frac{1}{1+\rho\vartheta} \right)^t u(c_t^T) \\ &\geq \sum_{t=0}^T \vartheta \left(\frac{1}{1+\rho\vartheta} \right)^t u'(c_t^T) c_t^T = \varphi^T \sum_{t=0}^T \vartheta^t p_t^T c_t^T = \varphi^T [\omega_1 + q_2^T \omega_2]. \end{aligned}$$

Dus de inkomensmultiplier is uniform van boven begrensd. Zij is ook van onder begrensd omdat $\varphi^T p_0^T = \varphi^T = u'(c_0^T) \geq u'(\sigma)$. Hiermee is tevens de prijs van de hulpbron q_2^T uniform begrensd. Er is ook een zekere begrenzing van (gemodificeerde) prijzen. Dat wordt als volgt afgeleid. De winst bedraagt

$$\begin{aligned} \sum_{t=0}^T p_t^T \vartheta z_t^T - y_1^T - q_2^T y_2^T &= \\ &= \sum_{t=0}^T p_t^T [\vartheta F(k_t^T - \alpha v_t^T, v_t^T) + k_t^T - k_{t+1}^T - \mu \vartheta k_t^T] - y_1^T - q_2^T \vartheta \sum_{t=0}^T v_t^T. \end{aligned}$$

Voor de kapitaalgoederenvoorraad is er een inwendige oplossing (afgezien van k_{T+1}^T die vanzelfsprekend gelijk is aan nul). Winstmaximalisatie leidt tot

$$p_t^T \vartheta [F_1(k_t^T - \alpha v_t^T, v_t^T) - \mu] = p_{t-1}^T - p_t^T \quad (1)$$

$$p_t^T \vartheta [F_2(k_t^T - \alpha v_t^T, v_t^T) - \alpha F_1(k_t^T - \alpha v_t^T, v_t^T)] = \vartheta q_2^T \quad (2)$$

voor $t = 0, 1, \dots, T$ met $p_{-1}^T := 1$. Hieruit valt een aantal interessante conclusies te trekken. Aangezien de prijs van de hulpbron constant is (er wordt alleen maar aan het begin in de voorraad gehandeld) volgt uit bovenstaande vergelijkingen dat de procentuele verandering in het marginale product van de grondstof gelijk is aan het netto marginale product van kapitaal. In de milieueconomie staat dit fenomeen bekend als de *Regel van Hotelling*. Het achterliggende idee is dat de eigenaar en exploitant van de hulpbron in een evenwicht indifferent is tussen het exploiteren van de bron nu en op een ander tijdstip. Dat is het geval indien de winst per eenheid stijgt met een percentage dat gelijk is aan de rentestand. De winst per eenheid is de prijs verminderd met de winningskosten. Dat is dus de waarde van de grondstof in het productieproces F_2 verminderd met de waarde van het kapitaal dat per eenheid winning moet worden ingezet αF_1 ; de (netto) rente is het netto marginale product van kapitaal $F_1 - \mu$.

Van belang voor het vervolg is dat Regel van Hotelling bepaalde consequenties heeft voor de prijzen. Vanwege de homogeniteit van de productiefunctie hangen de afgeleiden naar de inputs alleen maar af van de kapitaal grondstofverhouding $\kappa = k/v$ en dat brengen we in het vervolg tot uitdrukking, in het gebruik van niet geheel correcte notatie. We schrijven $F_i(k_i^T - \alpha v_i^T, v_i^T) = F_i(\kappa_i^T)$, $i = 1, 2$. Definieer $\hat{\kappa}$ door $F_1(\hat{\kappa}) = \mu$. Deze $\hat{\kappa}$ bestaat dankzij de gemaakte veronderstellingen en is positief. Neem een willekeurige $0 < t < T$ en stel $\kappa_t^T > \hat{\kappa}$. Vanwege de concaviteit van F

geldt dan dat $F_1(\kappa_t^T) < \mu$ en $p_{t-1}^T < p_t^T$. Maar dan volgt uit de tweede conditie voor winstmaximalisatie (2) dat $\kappa_{t-1}^T > \kappa_t^T > \hat{\kappa}$ en derhalve $p_{t-2}^T < p_{t-1}^T$, indien van toepassing. Het geval $\kappa_t^T < \hat{\kappa}$ levert een soortgelijk resultaat. Derhalve tendeert de kapitaal-grondstofverhouding monotoon naar $\hat{\kappa}$ en de prijzen convergeren dus ook, en wel naar van de oplossing van (2). Dit convergentieresultaat is onafhankelijk van de horizon T en de lengte van de perioden ϑ .

We definiëren tenslotte $r_t^T = (1 - \mu\vartheta)p_t^T$, $t = -1, 0, 1, \dots, T$. Deze grootheid is uniform begrensd.

3 Discrete tijd, oneindige horizon

Het bestaan van een algemeen evenwicht in de eindige economie met een oneindige horizon kan in twee stappen bewezen worden. In de eerste stap wordt op grond van de uniforme begrensdheid van allocaties en prijzen aangetoond dat deze grootheden limieten hebben. In de tweede stap wordt bewezen dat de limieten een algemeen evenwicht vormen.

Stap 1

Variabelen die duiden op een evenwicht in de eindig dimensionale economie, worden in het vervolg geïndexeerd door T en uitgebreid met nullen voor $t > T$. Bijvoorbeeld $x^T = (0, 0, c_1^T, c_2^T, \dots, c_T^T, 0, 0, \dots)$. Een belangrijke eigenschap is de volgende. Er bestaan $\hat{x} = (0, 0, \hat{c}_0, \hat{c}_1, \dots)$, $\hat{y} = (-y_1, -y_2, \hat{z}_0, \hat{z}_1, \dots)$, $\hat{k} = (\hat{k}_0, \hat{k}_1, \dots)$, $\hat{v} = (\hat{v}_0, \hat{v}_1, \dots)$, $\hat{\varphi}$ en $\hat{r} = (\hat{r}_0, \hat{r}_1, \dots)$ en een deelrij T_i zodanig dat

$$x^{T_i} \rightarrow \hat{x}, y^{T_i} \rightarrow \hat{y}, v^{T_i} \rightarrow \hat{v}, \varphi^{T_i} \rightarrow \hat{\varphi}, r^{T_i} \rightarrow \hat{r}$$

waar de convergentie puntsgewijs is. Ook de prijzen q_2^t en p_t^T convergeren in het licht van de definitie van r . Eén en ander is het gevolg van de toepassing van de Stelling van Alaoglu, waarover onder voor de niet-ingewijde lezer even wordt uitgeweid.

Intermezzo 1 (Stelling van Alaoglu) *Zij X een reële genormeerde ruimte. De gesloten eenheidsbol in de duale ruimte van X is zwak ster compact. In de toepassing die we voor ogen hebben is X gelijk aan ℓ_1 , de ruimte van vectoren $h = (h_0, h_1, \dots)$ met $\sum_{i=0}^{\infty} h_i < \infty$. De duale ruimte van ℓ_1 is ℓ_{∞} , de ruimte van begrensde rijen. We zeggen dat een rijtje $\{x^t\}$ in ℓ_{∞} zwak ster naar het element x convergeert als voor elke $h \in \ell_1$ geldt $\sum_{i=0}^{\infty} h_i x_i^t \rightarrow \sum_{i=0}^{\infty} h_i x_i$. De toepassing van de stelling op de rijtjes van evenwichtsallocaties en prijzen is nu vanzelfsprekend.*

Stap 2

Vervolgens moet worden aangetoond dat de limietallocatie en de limietprijzen ook daadwerkelijk een algemeen evenwicht vormen in de oneindig dimensionale economie. De vraag naar het kapitaalgoed en het consumptiegoed wordt altijd gedekt door het aanbod. Dat is immers het geval voor elke periode voor elke eindige horizon. Tevens ligt de limiet- consumptiebundel in de consumptieruimte, omdat die gesloten is. Dat de budgetrestrictie geldt is het gevolg van Lebesgue's 'dominated convergence'

theorema. Immers $\vartheta p_t^T c_t^T \leq \vartheta \left(\frac{1}{1+\rho\vartheta} \right)^t \frac{1}{\varphi^T} u(\sigma)$ en de inkomensmultiplier is uniform begrensd van nul. Nutsmaximalisatie volgt uit:

$$\sum_{t=0}^{\infty} \left(\frac{1}{1+\rho\vartheta} \right)^t [u(\hat{c}_t) - u(c_t)] \geq \hat{\varphi} [\hat{p}_t \hat{c}_t - \hat{p}_t c_t] \geq 0.$$

Het is rechttoe rechtaan om aan te tonen dat de limiet-productievector de winst maximaliseert. Minder eenvoudig is het om aan te tonen dat de aanschafwaarde van de hulpbron in de limietallocatie gelijk is aan de waarde van het gebruik in het productieproces: $\hat{q}_2 \hat{y}_2 = \hat{q}_2 \sum_{t=0}^{\infty} \vartheta \hat{v}_t$. Ik wil de lezer het bewijs niet onthouden. Als de prijs gelijk is aan 0, dan is het resultaat triviaal. Dus neem aan dat de prijs positief is. Op het eind van de voorgaande paragraaf is aangetoond dat de kapitaal-grondstofverhouding monotoon is en tendeert naar $\hat{\kappa}$. Omdat de grondstofprijs in de limiet gelijk is aan \hat{q}_2 , convergeert ook de prijs van het consumptiegoed. In de limiet, voor $t \rightarrow \infty$, geldt $\hat{p} [F_2(\hat{\kappa}) - \alpha F_1(\hat{\kappa})] = \hat{q}_2$. Als de evenwichtsprijs van de oneindige economie naar een constante tendeert, gaat de consumptie in de tijd naar nul. Ook de kapitaalgoederenvoorraad convergeert naar nul. Als er immers een $\varepsilon > 0$ zou bestaan met $\hat{k}_t > \varepsilon$ voor t groot genoeg, dan blijft dit gelden voor alle perioden hierna. Maar dat is in tegenspraak met het feit dat de kapitaal-grondstofverhouding naar een positieve constante convergeert en dat de grondstofvoorraad eindig is. Er geldt

$$\begin{aligned} \vartheta \hat{q}_2 \sum_{t=0}^T \hat{v}_t &= -y_1 + \sum_{t=0}^T \vartheta \hat{p}_t \hat{z}_t + \hat{p}_T \hat{k}_T = -y_1 + \sum_{t=0}^T \vartheta \hat{p}_t \hat{c}_t + \hat{p}_T \hat{k}_T = \\ &= \hat{q}_2 \omega_2 + \hat{p}_T \hat{k}_T \rightarrow \hat{q}_2 \omega_2. \end{aligned}$$

Alvorens over te gaan naar de oorspronkelijke economie in continuë tijd kan een aantal belangrijke eigenschappen worden afgeleid. In de eerste plaats is er de uniforme begrensdheid van de consumptie. Uniformiteit met betrekking tot de horizon is al aan de orde geweest. Het gaat hier om de uniformiteit met betrekking tot de lengte van de periodes. Bij het bewijs speelt de veronderstelling over de elasticiteit van het marginale nut, veronderstelling U4, een cruciale rol. Het bewijs is gestoeld op het feit dat dankzij deze veronderstelling de procentuele verandering in de consumptie beperkt is. Een gevolg hiervan is dat een willekeurig hoog consumptieniveau in een bepaalde periode een hoog consumptieniveau in volgende periodes induceert (vanwege de beperkte afname). Maar dan is een eenmalig hoog consumptieniveau in tegenspraak met de eigenschap dat de totale consumptie over de tijd beperkt is. Verder kan worden aangetoond dat ook de inkomensmultiplier begrensd is, onafhankelijk van de lengte van de periodes.

4 Oneindige horizon, continuë tijd

De overgang van een economie met een oneindige horizon en discrete tijd naar de oorspronkelijke economie met een oneindige horizon en continuë tijd voltrekt zich

door de lengte van de periodes naar nul te laten gaan. Een handig hulpmiddel hierbij is de karakteristieke functie. Voor een variabele ℓ definiëren we:

$$\ell^\vartheta(s) = \sum_{t=0}^{\infty} \hat{\ell}_t^\vartheta \chi_t(s).$$

met $\chi_t(s) = 1$ als $\vartheta t \leq s \leq \vartheta(t+1)$ en $\chi_t(s) = 0$ elders. Dus $\ell^\vartheta(s)$ geeft de waarde die ℓ aanneemt in de discrete tijdseconomie met periodelengte ϑ aan het begin van de periode waarin s ligt.

Het zou te ver voeren om voor alle relevante variabelen de puntsgewijze convergentie te bewijzen. Daarom beperken wij ons hier tot de convergentie van de kapitaalgoederenvoorraad, waarmee tevens een algemeen beeld geschetst wordt van de gevolgde methodologie.

Voor het algemeen evenwicht voor de discrete tijd economie met een oneindige horizon is bekend dat

$$k_{t+1}^\vartheta = k_t^\vartheta + \vartheta [F(k_t^\vartheta - \alpha v_t^\vartheta, v_t^\vartheta) - \mu k_t^\vartheta - z_t^\vartheta], \quad k_0^\vartheta = \omega_1.$$

Neem een willekeurige s en bepaal, voor gegeven ϑ , t zodanig dat $\vartheta t \leq s \leq \vartheta(t+1)$. De omvang van de kapitaalgoederenvoorraad aan het begin van de periode waarin s ligt is gelijk aan de beginvoorraad ω_1 vermeerderd met de integraal van alle productie tot aan s , maar verminderd met de productie tussen het begin van de periode waarin s ligt en s zelf. Derhalve

$$\begin{aligned} k^\vartheta(s) - \left[\omega_1 + \int_0^s (F(k^\vartheta(\tau) - \alpha v^\vartheta(\tau), v^\vartheta(\tau)) - \mu k^\vartheta(\tau) - z^\vartheta(\tau)) d\tau \right] &= \\ = -(s - t\vartheta) [F(k_t^\vartheta - \alpha v_t^\vartheta, v_t^\vartheta) - \mu k_t^\vartheta - z_t^\vartheta] & \end{aligned}$$

Er geldt nu

$$\left| k^\vartheta(s) - \left[\omega_1 + \int_0^s (F(k^\vartheta(\tau) - \alpha v^\vartheta(\tau), v^\vartheta(\tau)) - \mu k^\vartheta(\tau) - z^\vartheta(\tau)) d\tau \right] \right| \leq \vartheta A$$

voor alle $s \geq 0$ en voor zekere A omdat $F(k_t^\vartheta - \alpha v_t^\vartheta, v_t^\vartheta) - \mu k_t^\vartheta - z_t^\vartheta$ uniform begrensd is, zoals eerder aangetoond. Ook de integrand is uniform begrensd. Derhalve heeft deze een zwak ster limiet en convergeert $k^\vartheta(s)$ puntsgewijs (Alaoglu).

Naar analogie valt ook de puntsgewijze convergentie van $r^\vartheta(s)$ naar zekere $r(s)$ aan te tonen. De definitie van r impliceert dan meteen dat $p^\vartheta(s) = e^{\mu t} r(s) \rightarrow p(s)$. Omdat φ^ϑ uniform begrensd is, treedt ook voor deze variabele puntsgewijze convergentie op en dientengevolge voor de consumptie. Tenslotte volgt de convergentie van de input uit de natuurlijke hulpbron, omdat winstmaximalisatie impliceert dat $p_t^\vartheta [F_2(k_t^\vartheta - \alpha v_t^\vartheta, v_t^\vartheta) - \alpha F_1(k_t^\vartheta - \alpha v_t^\vartheta, v_t^\vartheta)] = q_2^\vartheta$ en de afgeleiden alleen afhangen van de kapitaal-grondstofverhouding en kapitaal puntsgewijs convergeert.

De laatste stap in deze exercitie is aan te tonen dat de limieten ook het evenwicht vormen in de oorspronkelijke economie met continuë tijd en een oneindige horizon. Het spreekt vanzelf dat de vraag naar kapitaal gelijk is aan het aanbod. Tevens is de vraag naar de natuurlijke hulpbron niet groter dan het aanbod, en als het aanbod

groter is, dan is de prijs gelijk aan nul. Dit volgt uit het feit dat deze eigenschappen ook al golden in de discrete tijd economie met de oneindige horizon. De prijzen van de consumptiegoederen zijn altijd positief en de vraag naar consumptiegoederen is gelijk aan het aanbod. De budgetrestrictie

$$\int_0^{\infty} p(t)c(t) dt = \omega_1 + q_2\omega_2$$

geldt vanwege het feit dat $p^\vartheta(t)c^\vartheta(t) \leq e^{-\rho t}u(\sigma)/\varphi^\vartheta$ en Lebesgue's 'dominated convergence' theorema. Dan volgt onmiddellijk dat $\pi \cdot y = 0$. Tevens wordt het nut gemaximaliseerd. Het enige dat nog moet worden bewezen is dat y de winst maximaliseert. We weten (zie (1) en (2)) dat

$$\begin{aligned} p(t) [\mu - F_1(k(t) - \alpha v(t), v(t))] &= \dot{p}(t) \text{ en} \\ p(t) [F_2(k(t) - \alpha v(t), v(t)) - \alpha F_1(k(t) - \alpha v(t), v(t))] &= q_2. \end{aligned}$$

Met gebruikmaking van bovenstaande vergelijkingen, de homogeniteit van de productiefunctie F en $\dot{k} = F - \mu k - z$ volgt

$$q_2 \int_0^T v(t) dt = p(T)k(T) - k_0 + \int_0^T p(t)z(t) dt \rightarrow q_2 y_2,$$

omdat $p(T)k(T) \rightarrow 0$. Derhalve wordt er niet meer van de bron uitgeput dan de bron initieel bevatte. Dit is een voldoende voorwaarde voor winstmaximalisatie.

Hiermee lijkt het bestaan van een algemene evenwicht in de oorspronkelijke economie aangetoond. Strikt formeel zijn we er nog niet. Er is immers nog niet bewezen dat de kapitaalgoederenvoorraad differentieerbaar is, hetgeen in de definitie van algemeen evenwicht was opgelegd. Deze eis is echter equivalent aan de eis dat het grondstofgebruik (in de limiet) stuksgewijs continu is in de tijd en dat is voor de hand liggend.

5 Conclusie

Aangetoond is dat het bestaan van een algemeen evenwicht voor een specifieke economie gedefinieerd voor continuë tijd met een oneindige horizon valt te bewijzen in een serie van drie stappen: het bestaan van een evenwicht in een analoge economie in discrete tijd met een eindige horizon, het bestaan van een evenwicht in een economie met discrete tijd en een oneindige horizon en tenslotte in de oorspronkelijke economie. In deze analyse is ampel gebruik gemaakt van dualiteit. De vraag kan gesteld worden wat de voordelen van deze methode zijn ten opzichte van meer gebruikelijke methoden zoals elders ontwikkeld. In elk geval kunnen meer traditionele methoden in het onderhavige geval niet worden toegepast. Tegengeworpen zou ook kunnen worden dat dit een zeer specifiek model is en dat de bredere toepasbaarheid van de methode niet is aangetoond. Dat is ongetwijfeld zo, en daar ligt een onderwerp voor nader onderzoek.

Referentie

Mas-Colell, A., en W. Zame (1991), “Equilibrium Theory in Infinite Dimensional Spaces”, in W. Hildenbrand en H. Sonnenschein (eds.), *Handbook of Mathematical Economics, Volume 4*, North-Holland, Amsterdam, 1835–1898.

UITBREIDINGEN VAN WAARGENOMEN KEUZES

Ton Storcken*

Universiteit van Maastricht

Zij die het voorrecht hebben om een geruime tijd in de wetenschappelijke omgeving van Pieter te mogen vertoeven, leren op zijn minst twee van zijn abstracte stokpaardjes kennen: Tripolariteit en Dualiteit. Het artikel dat hierna volgt, maakt duidelijk dat het keuzegedrag van een agent, zoals dat standaard micro-economisch gemodelleerd wordt, een dualiteit tussen voorkeuren over alternatieven en voorkeuren over verzamelingen van alternatieven vooronderstelt. Op basis van die dualiteit kan men dat standaard micro-economisch agentengedrag karakteriseren en zo zijn beperkingen aan het licht brengen.

Dit soort fundamenteel onderzoek heeft altijd Pieters aandacht gehad, getuige het Research Memorandum “Extensions of Choice Behaviour” dat wij samen aan de Economische Faculteit van de KUB in 1992 schreven. Eigenlijk is het volgende een vertaald uittreksel hiervan. Het lag altijd in ons beider bedoeling om dit verhaal te herschrijven. Het leven neemt echter zijn beloop waardoor dit nooit gebeurd is. Dit verhaal is een eerste aanzet om dit oorspronkel verhaal aan te passen: Wetend dat de kritische geest van Pieter alleen maar de meest essentiële zaken van het volgende overeind zal laten. Op deze manier hoop ik u een blik op onze samenwerking te geven.

1 Inleiding

In vele economische modellen wordt er vanuit gegaan dat agenten “rationeel” handelen, wat betekend dat zij volledig geïnformeerd zijn over hun beschikbare alternatieven en zij hieruit op basis van hun individuele voorkeur het beste alternatief kiezen. In dit artikel zullen we beide veronderstellingen, “volledige informatie” en “beste kiezen”, niet zonder meer aannemen. Dit zal leiden tot een formele theorie die waarneembare keuzes van een agent koppelt aan voorkeuren over alternatieven via voorkeuren over verzamelingen van alternatieven.

Alle voorkeuren ten spijt, zelfs navraag naar deze zal ze niet direct bij een agent waarneembaar maken. Vandaar dat men op basis van de waarneembare keuzes van een agent deze voorkeuren, als zij er al zijn, slechts kan proberen te herleiden. Om dit te kunnen veronderstelt men dat de agent de waargenomen keuzes

* *Adres:* Department of Quantitative Economics, Maastricht University, P.O. Box 616, NL 6200 MD Maastricht. *Email:* t.storcken@ke.unimaas.nl

baseert op het kiezen van beste alternatieven, gegeven de keuzesituatie, volgens een volledige voorkeurrelatie over alle alternatieven. Dit idee van de “Revealed Preference Theory”, geformuleerd door Samuelson (1938), leidt er toe dat de te achterhalen voorkeurrelatie een zwakke ordening van alle alternatieven is. Ook moge duidelijk zijn dat, op zijn minst theoretisch, niet alle denkbare waar te nemen keuzes van agenten volgens het hierboven uiteengezette model te beschrijven zijn. In Richter (1966) staat dan ook een noodzakelijke en voldoende voorwaarde, congruentie genaamd, voor de waargenomen keuzes opdat deze in het hierboven beschreven model passen. Deze voorwaarde is geformuleerd in termen van *gebleken voorkeur*: Als alternatieven x en y aanwezig zijn in een keuzesituatie en alternatief x gekozen wordt, dan zegt men dat de *zwakke preferentie van x boven y blijkt*. Als echter bovendien y niet gekozen wordt, dan zegt men dat de *sterke preferentie van x boven y blijkt*. Congruentie betekent nu dat een rij van opeenvolgende zwakke gebleken preferenties niet strijdig mag zijn met een sterke gebleken preferentie.

Naast congruentie zijn er ook andere klassieke condities voor de waarneembare keuzes van een agent welke in specifieke omstandigheden dit gedrag zoals hierboven uiteengezet modelleerbaar maken als beste keuzes ten opzichte van een (gebleken) voorkeur. Hierbij valt bijvoorbeeld te denken aan het “Weak Axiom of Revealed Preference” van Samuelson (1938): sterke gebleken voorkeuren zijn niet strijdig met zwakke gebleken voorkeuren; of aan het “Strong Axiom of Revealed Preference” van Houthakker (1950) en Ville (1951-1952 en eerder in 1946): rijen van opeenvolgende sterke gebleken voorkeuren zijn niet strijdig met andere sterke gebleken voorkeuren, oftewel de sterke gebleken voorkeur is acyclisch; of aan de “Independence of Irrelevant Alternatives” van Arrow (1948), de keuze bij keuzesituatie zeg X wordt indien mogelijk overgeorven bij deelkeuzesituatie zeg Y , bijvoorbeeld de wereldkampioen korfbal is ook de winnaar van de Benelux.

In dit artikel vooronderstellen we *a priori* niet dat de waargenomen keuzes tot stand komen door het kiezen van de besten ten opzichte van een nog eventueel te blijken (volledige) voorkeur. Immers niet alle keuzes zijn resultaat van zo’n maximalisatie. Bijvoorbeeld de kennis over de Nederlandse tweedehands automarkt zal zelfs bij gebruikmaking van het internet niet volledig zijn. Een potentiële koper zal dan ook altijd de voor hem meest gunstige keuze vastleggen onder nevenvoorwaarden, bijvoorbeeld dat er zich binnen de laatste twee weken geen beter aanbod voordeed. Het aanwezig zijn van drop bij een buitenlandse kruidenier zou uw keuze tussen merkloze kaas en worst kunnen beïnvloeden. Ook kunnen keuzes onderhevig zijn aan impliciete regels van fatsoen of mode, welke ten minste op het eerste oog niet goed te vangen lijken met het hier eerder beschreven model. Het beste kiezen, en daarmee onlosmakelijk verbonden een (eventueel nog te blijken volledige transitieve) voorkeurrelatie, is voor ons dan ook geen voor zichzelf sprekend grondbeginsel. We zullen hier dan ook nieuwe grondbeginselen ontwikkelen, die dit eventueel onderbouwen. Hiermee worden de beperkingen van deze geblekenvoorkeuretheorie duidelijk: geen van bovenstaande voorbeelden zijn met deze theorie te analyseren. Wel vinden we zo een diepere onderbouwing van die theorie en eventueel aanpassingsmogelijkheden om haar op een breder scala van problemen toe te passen.

Uitgangspunt zijn de waargenomen keuzes bij de diverse keuzesituaties van een agent. We zullen dit in het vervolg de waargenomen keuzes noemen. Laat A

de verzameling alternatieven zijn. Dan hebben we voor verschillende deelverzamelingen van A (dit zijn de keuzesituaties) de keuzes van een agent waargenomen. Laat \mathbf{B} deze verzameling van deelverzamelingen (oftewel keuzesituaties) zijn. De waargenomen keuzes zijn nu in principe vast te leggen met een afbeelding C van \mathbf{B} naar $P(A)$, de machtsverzameling van A , dit is de verzameling van alle deelverzamelingen van A . Afbeelding C voegt nu aan iedere keuzesituatie B uit \mathbf{B} de keuze van de agent bij die situatie toe: $C(B)$. Dus $C(B)$ is weer een deel van B . We sluiten niet uit dat er geen of meer dan één alternatief gekozen kan worden. Het doel is nu om de waargenomen keuzes te vertalen voor nieuwe situaties B die niet binnen \mathbf{B} liggen. Dus in concreto zouden we de afbeelding C op \mathbf{B} willen uitbreiden tot een afbeelding \tilde{C} op $\tilde{\mathbf{B}}$, waarbij \mathbf{B} een deelverzameling is van $\tilde{\mathbf{B}}$ en $\tilde{\mathbf{B}}$ zo veel mogelijke deelverzamelingen van A bevat. Deze uitbreiding van C naar \tilde{C} zullen we realiseren op basis van twee vertaalslagen. Eén om voorkeuren tussen deelverzamelingen van A naar voorkeuren tussen alternatieven uit A te vertalen en één om voorkeuren tussen alternatieven uit A naar voorkeuren tussen deelverzamelingen van A te vertalen.

Ofschoon in beginsel deze twee vertaalslagen onafhankelijk van elkaar opgezet worden, volgt uit hun verdere uitwerking dat ze elkaars duale zijn. Een eigenschap die zwakker is dan elkaars inverse. Net zoals de vertaalslagen Nederlands-Engels en Engels-Nederlands niet elkaars inverse zijn. Bijvoorbeeld het Nederlandse woord “braaf” kan vertaald worden met het Engelse woord “kind” dat weer terugvertaald kan worden met ook het Nederlandse woord “soort”. Wel zal blijken dat de oorspronkelijke waargenomen keuzes in de terugvertaling zit en juist de extra hoeveelheid informatie die door deze heen en weer vertaling ontstaat tot een uitbreiding van die waargenomen keuzes leidt. Ook zal blijken dat herhalen van deze dubbele vertaalslag niet zinvol is: elke volgende dubbele vertaalslag levert precies dezelfde informatie op als de eerste. Met andere woorden we krijgen op deze manier in één keer de maximaal haalbare uitbreiding van de waargenomen keuzes.

De vertaalslagen worden middels vijf natuurlijke condities uniek vastgelegd. Deze condities zijn: contextonafhankelijkheid, onvooringenomenheid, naamsonafhankelijkheid, behoudendheid en representativiteit. We zullen deze vijf condities verder omschrijven met betrekking tot een vertaler van voorkeuren tussen deelverzamelingen van A naar voorkeuren tussen alternatieven uit A . De interpretatie van deze condities voor de andere vertaler van voorkeuren tussen alternatieven uit A naar voorkeuren tussen deelverzamelingen uit A is analoog en laten we derhalve achterwege.

contextonafhankelijkheid betekent dat het al dan niet tot de vertaling behoren van een voorkeurrelatie over deelverzamelingen van A , zeg \mathfrak{R}^1 , niet afhangt van een eventueel grotere voorkeurrelatie over deelverzamelingen, zeg \mathfrak{R}^2 , waartoe \mathfrak{R}^1 behoort. Een zelfde contextonafhankelijkheid treft men aan bij de beschrijving van de zwakke en sterke gebleken voorkeur. Immers of alternatief x zwak dan wel sterk boven alternatief y geprefereerd blijkt te zijn hangt in beginsel van slechts een keuzesituatie af, waarin x gekozen en y aanwezig is en al dan niet gekozen wordt. Andere situaties waarbij y gekozen en x aanwezig is en al dan niet gekozen wordt, of een rij van situaties waarin y gekozen wordt en alternatief z_1 aanwezig is, z_1 gekozen en z_2 aanwezig is, ... en tenslotte z_{108} gekozen en x aanwezig is, spelen hierbij geen

rol van betekenis om de gebleken voorkeur tussen x en y te bepalen.

Onvooringenomenheid betekent dat er geen *a priori* restricties op de vertalingen van voorkeuren tussen deelverzamelingen van A gelegd worden. De vertaling van de lege relatie dient dus elke voorkeur over alternatieven uit A , welke in een vertaling op kan treden, toe te laten, aangezien deze lege relatie geen enkele restrictie op voorkeuren tussen deelverzamelingen van A legt.

Deze twee condities samen impliceren paarsgewijsheid. Beschouwt met een voorkeurrelatie, \mathfrak{R} , over deelverzamelingen van A , zo bestaat deze uit paren, zeg (X, Y) , van deelverzamelingen uit A . Zo'n paar zit in die relatie precies dan als X zwak boven Y geprefereerd wordt. Analooq zit een paar van alternatieven (x, y) in voorkeurrelatie R over alternatieven uit A , precies dan als x zwak boven y geprefereerd wordt. Paarsgewijsheid betekent dat een paar (x, y) in de vertaling van \mathfrak{R} zit precies dan als dit paar in alle vertalingen van alle paren (X, Y) van \mathfrak{R} zit. Oftewel, in de vertaling van een voorkeurrelatie over deelverzamelingen van A zit precies dat wat in de vertaling van alle paren uit die voorkeurrelatie zit.

Naamsonafhankelijkheid betekent dat de vertaling noch in zijn geheel noch gedeeltelijk afhankelijk is van zowel de naamgeving van de alternatieven uit A als van de deelverzamelingen van A . Bij de vertaling speelt de naamgeving van zowel de alternatieven als ook die van de deelverzamelingen van A geen rol. Dit samen met paarsgewijsheid betekent, dat of alternatievenpaar (x, y) in de vertaling zit van een voorkeurrelatie \mathfrak{R} over deelverzamelingen van A , volledig terug te voeren is op de verzamelingstheoretische “element van” relatie tussen x, y, X en Y voor alle paren (X, Y) uit \mathfrak{R} . Met andere woorden alles is terug te voeren tot vertalingen van éénelementsrelaties $\{(X, Y)\}$ over deelverzamelingen van A , welke gebaseerd zijn op de “element van relatie” tussen de alternatieven uit A en de deelverzamelingen X en Y van A . De laatste twee condities gaan over deze vertalingen van die éénelementsrelaties.

Behoudendheid betekent dat zoveel mogelijk voorkeuren tussen alternatieven in de vertaling zitten: enkel als daartoe een gegronde aanleiding bestaat zit een voorkeur tussen twee alternatieven niet in de vertaling. Indien twee alternatieven niet vergeleken kunnen worden, bijvoorbeeld, dan zullen deze in de vertaling zitten van zo'n éénelementsrelatie. Voorbeelden van onvergelykbaarheid zijn als een van de alternatieven in geen van de deelverzamelingen van de éénelementsrelatie voorkomt of als beide alternatieven in beide deelverzamelingen voorkomen.

Tenslotte, representativiteit betekent dat voor alternatieven x en y de vertaling van de éénelementsrelaties $\{(\{x\}, \{y\})\}$, $\{(\{x\}, \{x, y\})\}$ en $\{(\{x, y\}, \{y\})\}$ voor de handliggend is. Alle drie zijn op te vatten als een sterke voorkeur van x boven y . Representativiteit verlangt van de vertaling dan ook dat zij deze éénelementsrelatie in zo'n sterke voorkeur tussen x en y vertaalt.

Naast de eerder genoemde resultaten zullen we aantonen dat de heen en weer vertaling enkel aanleiding geeft tot de beschrijving van maximaliserend keuzegedrag. Ook zullen we het hier ingevoerde mechanisme vergelijken met het standaard “Revealed Preference Theory” model. De condities van “Weak” en “Strong Axiom of Revealed Preference” samen zijn voldoende om het keuzegedrag in preferentie maximaliserende termen te kunnen beschrijven. We kunnen dus concluderen dat

dit axiomatisch model een onderbouwing geeft voor het preferentie maximaliserend gedrag welk standaard in de “Revealed Preference Theory” gebruikt wordt.

Paragraaf 2 behandelt een verdere formalisering van het hierboven beschreven model en bespreekt de relatie tussen dit model en het standaard model uit de “Revealed Preference Theory”. In paragraaf 3 zullen we tenslotte suggesties aangeven hoe ons model eventueel aan te passen is om niet zuiver maximaliserend keuzegedrag toe te staan.

2 Van keuzes naar voorkeuren en terug

Laat A , $P(A)$, \mathbf{B} , $\tilde{\mathbf{B}}$, C en \tilde{C} dezelfde betekenis hebben als in de inleiding. Een voorkeurrelatie over deelverzamelingen van A , weergegeven met \mathfrak{R} , is een deelverzameling van het Cartesische product $P(A) \times P(A)$. Een voorkeurrelatie over alternatieven uit A , weergegeven met R , is een deelverzameling van het Cartesische product $A \times A$. De grafiek van de waargenomen keuzes C , weergegeven met $gr(C)$, gedefinieerd door $\{(C(B), B) \mid B \in \mathbf{B}\}^1$, is een relatie over deelverzamelingen van A . Gegeven dat $C(B)$ de keuze van de agent is in keuzesituatie B , is het natuurlijk te stellen dat de agent $C(B)$ zwak boven B prefereert. Daarmee is de grafiek van C te zien als een voorkeurrelatie over deelverzamelingen van A .

De vertaler van (voorkeur) relaties over deelverzamelingen van A naar (voorkeur) relaties over alternatieven uit A wordt aangegeven met Δ . Daar Δ een voorkeurrelatie \mathfrak{R} over deelverzamelingen van A zal vertalen in een (voorkeur) relatie R over alternatieven uit A en we deze vertaling uniek veronderstellen, is Δ een afbeelding van $P(P(A) \times P(A))$, de verzameling van alle relaties over deelverzamelingen van A , naar $P(A \times A)$, de verzameling van alle relatie over alternatieven uit A . Analoog is de vertaler van (voorkeur) relaties over alternatieven uit A naar (voorkeur) relaties over deelverzamelingen van A , weergegeven door Γ , een afbeelding van $P(A \times A)$ naar $P(P(A) \times P(A))$. De vertaling van \mathfrak{R} door Δ wordt weergegeven met $\Delta(\mathfrak{R})$ en die van R door Γ met $\Gamma(R)$.

De vertaler Δ heet *contextonafhankelijk* indien voor alle relaties \mathfrak{R}^1 , \mathfrak{R}^2 en \mathfrak{R}^3 over deelverzamelingen van A en alle alternatieven x en y uit A

$$\begin{aligned} (x, y) \in \Delta(\mathfrak{R}^1)^2 &\iff (x, y) \in \Delta(\mathfrak{R}^2) \\ &\text{impliceert} \\ (x, y) \in \Delta(\mathfrak{R}^1 \cup \mathfrak{R}^3) &\iff (x, y) \in \Delta(\mathfrak{R}^2 \cup \mathfrak{R}^3). \end{aligned}$$

Merk op dat “ $(x, y) \in \Delta(\mathfrak{R}^1) \iff (x, y) \in \Delta(\mathfrak{R}^2)$ ” equivalent is met óf beide $(x, y) \in \Delta(\mathfrak{R}^1)$ en $(x, y) \in \Delta(\mathfrak{R}^2)$ óf geen van deze twee. Met andere woorden óf bij zowel $\Delta(\mathfrak{R}^1)$ en $\Delta(\mathfrak{R}^2)$ wordt x zwak boven y geprefereerd, óf dit is bij geen van beiden het geval. Deze conditie is dan ook als volgt te interpreteren: als de zwakke voorkeur van x boven y een gelijkwaardige betekenis heeft in de vertalingen

¹Hier staat eigenlijk de omgedraaide grafiek relatie.

² $(x, y) \in \Delta(\mathfrak{R})$ betekent dat x zwak boven y wordt geprefereert bij $\Delta(\mathfrak{R})$ de vertaling van \mathfrak{R} .

van $\Delta(\mathfrak{R}^1)$ en $\Delta(\mathfrak{R}^2)$, dan heeft deze voorkeur een gelijkwaardige betekenis in de vertalingen $\Delta(\mathfrak{R}^1 \cup \mathfrak{R}^3)$ en $\Delta(\mathfrak{R}^2 \cup \mathfrak{R}^3)$. Oftewel de invloed van \mathfrak{R}^3 op de zwakke voorkeur van x boven y in de vertaling van $\mathfrak{R}^1 \cup \mathfrak{R}^3$ is hetzelfde als in de vertaling van $\mathfrak{R}^2 \cup \mathfrak{R}^3$. Dit is precies wat men van contextonafhankelijkheid zou verwachten.

Analoog heet vertaler Γ *contextonafhankelijk* indien voor alle relaties R^1, R^2 en R^3 over alternatieven uit A en alle deelverzamelingen X en Y van A

$$\begin{aligned} (X, Y) \in \Gamma(R^1) &\iff (X, Y) \in \Gamma(R^2) \\ &\text{impliceert} \\ (X, Y) \in \Gamma(R^1 \cup R^3) &\iff (X, Y) \in \Gamma(R^2 \cup R^3). \end{aligned}$$

De vertaler Δ heet *onvooringenomen* indien

$$\Delta(\emptyset) \supseteq \cup \{ \Delta(\mathfrak{R}) \mid \mathfrak{R} \text{ is een relatie over deelverzamelingen van } A \}.$$

In dat geval bevat $\Delta(\emptyset)$ al die voorkeuren over alternatieven die bij minstens één vertaling van een (voorkeur)relatie over deelverzamelingen A optreden. Aangezien de lege relatie \emptyset geen restrictie op de voorkeuren over deelverzamelingen van A legt, zal onvooringenomenheid betekenen dat al deze voorkeuren over alternatieven uit A die ooit in een vertaling optreden in de vertaling van $\Delta(\emptyset)$ zitten.

Analoog heet vertaler Γ *onvooringenomen* indien

$$\Gamma(\emptyset) \supseteq \cup \{ \Gamma(R) \mid R \text{ een relatie over alternatieven uit } A \}.$$

Vertaler Δ heet *paarsgewijs* indien

$$\Delta(\mathfrak{R}) = \cap \{ \Delta(\{(X, Y)\}) \mid (X, Y) \in \mathfrak{R} \}.$$

Dus voor alternatieven x en y uit A geldt dat (x, y) in de vertaling van \mathfrak{R} zit precies dan als (x, y) in de vertaling van alle voorkeuroparen (X, Y) uit \mathfrak{R} zit.

Vertaler Γ heet *paarsgewijs* indien

$$\Gamma(R) = \cap \{ \Gamma(\{(x, y)\}) \mid (x, y) \in R \}.$$

Het is duidelijk dat als vertaler Δ paarsgewijs is, dan is hij ook contextonafhankelijk. Daar met paarsgewijsheid voor alle relaties \mathfrak{R} over deelverzamelingen van A geldt:

$$\Delta(\emptyset) \cap \Delta(\mathfrak{R}) = \Delta(\mathfrak{R}) \text{ volgt } \Delta(\mathfrak{R}) \subseteq \Delta(\emptyset)$$

volgt ook onvooringenomenheid.

Omgekeerd impliceren onvooringenomenheid en contextonafhankelijkheid paarsgewijsheid. Met context onafhankelijkheid volgt

$$\cap \{ \Delta(\{(X, Y)\}) \mid (X, Y) \in \mathfrak{R} \} \subseteq \Delta(\mathfrak{R}).$$

Voor bewijzen die hier niet gegeven worden, zij verwezen naar het oorspronkelijke artikel van Ruys en Storcken (1992). We zullen hier alleen de omgekeerde

inclusie bespreken. Beschouw hiervoor alternatieven x en y uit A en relatie \mathfrak{R} over deelverzamelingen van A zodanig dat $(x, y) \in \Delta(\mathfrak{R})$. We tonen nu aan dat $(x, y) \in \Delta(\{(X, Y)\})$ voor een willekeurig paar (X, Y) uit \mathfrak{R} . Omdat het paar (X, Y) willekeurig gekozen is, is hiermee dan de omgekeerde inclusie aangetoond. Daar $(x, y) \in \Delta(\mathfrak{R})$ volgt met onvooringenomenheid dat $(x, y) \in \Delta(\emptyset)$. Dus

$$(x, y) \in \Delta(\mathfrak{R}) \iff (x, y) \in \Delta(\emptyset)$$

en met contextonafhankelijkheid

$$(x, y) \in \Delta(\mathfrak{R} \cup \{(X, Y)\}) \iff (x, y) \in \Delta(\emptyset \cup \{(X, Y)\}).$$

Daar $(X, Y) \in \mathfrak{R}$ en $(x, y) \in \Delta(\mathfrak{R})$ volgt met het voorgaande dat $(x, y) \in \Delta(\{(X, Y)\})$.

Deze en een analoge redenering voor vertaler Γ leert ons het volgende.

Stelling 1 *Vertaler Δ is paarsgewijs dan en slechts dan als hij contextonafhankelijk en onvooringenomen is. Vertaler Γ is paarsgewijs dan en slechts dan als hij contextonafhankelijk en onvooringenomen is.*

Deze propositie maakt duidelijk dat de vertaling $\Delta(\mathfrak{R})$ van \mathfrak{R} , een (voorkeur) relatie over deelverzamelingen van A , volledig bepaald wordt door de vertalingen $\Delta(\{(X, Y)\})$ van alle paren (X, Y) uit \mathfrak{R} . Dus voor het bepalen van de vertaling $\Delta(\mathfrak{R})$ van \mathfrak{R} kunnen we ons beperken tot het bepalen van de vertaling $\Delta(\{(X, Y)\})$ van paren (X, Y) uit \mathfrak{R} . Paarsgewijsheid maakt het mogelijk om het vertaalprobleem in elementaire delen op te lossen. De volgende drie condities zijn dan ook voor zulke elementaire delen geformuleerd.

Vertaler Δ , respectievelijk Γ , heet *naamsafhankelijk* indien voor alle alternatieven x, y, a en b uit A en alle deelverzamelingen X, Y, V en W van A

$$\begin{aligned} \text{als } x \in X &\iff a \in V \text{ en } x \in Y \iff a \in W \\ \text{en als } y \in X &\iff b \in V \text{ en } y \in Y \iff b \in W, \\ \text{dan } (x, y) \in \Delta(\{(X, Y)\}) &\iff (a, b) \in \Delta(\{(V, W)\}), \\ \text{respectievelijk } (X, Y) \in \Gamma(\{(x, y)\}) &\iff (V, W) \in \Gamma(\{(a, b)\}). \end{aligned}$$

Dus als x in preferentie paar (X, Y) dezelfde rol speelt als a in (V, W) en als y in (X, Y) dezelfde rol speelt als b in (V, W) , dan zit de zwakke voorkeur van x boven y in de vertaling van $\{(X, Y)\}$ precies dan als de zwakke voorkeur van a boven b in de vertaling van $\{(V, W)\}$ zit. Anders geformuleerd, een naamsverwisseling van zowel alternatieven als deelverzamelingen van alternatieven, leidt tot eenzelfde naamsverwisseling in de vertaling. Uit deze conditie volgt dat een paar (x, y) uitsluitend op basis van de “element van relatie” uit de verzamelingsleer in de vertaling $\Delta(\{(X, Y)\})$ van $\{(X, Y)\}$ zit en de eventuele aanwezigheid van een derde alternatief z speelt hierbij geen rol.

Vertaler Δ , respectievelijk Γ , heet *behoudend* indien voor alle alternatieven x en y

uit A en alle deelverzamelingen X en Y van A

$$(x, y) \in \Delta(\{(X, Y)\}), \text{ respectievelijk } (X, Y) \in \Gamma(\{(x, y)\}),$$

indien minstens één van de volgende vijf eigenschappen geldt:

$$x \notin X \cup Y, y \notin X \cup Y, \{x, y\} \cap X = \emptyset, \{x, y\} \cap Y = \emptyset \text{ of } \{x, y\} \subseteq X \cap Y.$$

Merk op dat indien $x \notin X \cup Y$, dan heeft $\{(X, Y)\}$ geen betrekking op x en als $\{x, y\} \cap X = \emptyset$, dan heeft X geen betrekking op de alternatieven x en y . Behoudendheid zegt dan dat er in die situaties geen reden is om de zwakke voorkeur van x boven y uit de vertaling van $\{(X, Y)\}$ te weren. Indien $\{x, y\} \subseteq X \cap Y$ dan is er wederom geen gegronde reden om die zwakke voorkeur buiten de vertaling van $\{(X, Y)\}$ te houden.

Vertaler Δ heet *representatief* indien voor alle alternatieven x en y uit A

$$(x, y) \in \Delta(\{(\{x\}, \{y\})\}), (x, y) \in \Delta(\{(\{x\}, \{x, y\})\}),$$

$$(x, y) \in \Delta(\{(\{x, y\}, \{y\})\}), (x, y) \notin \Delta(\{(\{y\}, \{x\})\}),$$

$$(x, y) \notin \Delta(\{(\{x, y\}, \{x\})\}) \text{ en } (x, y) \notin \Delta(\{(\{y\}, \{x, y\})\}).$$

Vertaler Γ heet *representatief* indien voor alle alternatieven x en y uit A

$$(\{x\}, \{y\}) \in \Gamma(\{(x, y)\}), (\{x\}, \{x, y\}) \in \Gamma(\{(x, y)\}),$$

$$(\{x, y\}, \{y\}) \in \Gamma(\{(x, y)\}), (\{y\}, \{x\}) \notin \Gamma(\{(x, y)\}),$$

$$(\{x, y\}, \{x\}) \notin \Gamma(\{(x, y)\}) \text{ en } (\{y\}, \{x, y\}) \notin \Gamma(\{(x, y)\}).$$

Zoals in de inleiding reeds verduidelijkt is, betekent representativiteit dat de vertaler voor de hand liggend is op deelverzamelingen van $\{x, y\}$. Merk hierbij bovendien op dat de preferentie $\{(\{x\}, \{y\})\}$ op deelverzamelingen van A zeer wel te indentificeren is met de sterke voorkeur van x boven y . Vandaar dat met representativiteit het paar (x, y) in de vertaling $\Delta(\{(\{x\}, \{y\})\})$ van $\{(\{x\}, \{y\})\}$ zit maar het paar (y, x) niet.

Vanaf nu nemen we aan dat beide vertalers Δ en Γ aan de vijf hiervoor beschreven condities voldoen.

Het is recht toe recht aan om na te gaan dat voor alle alternatieven x en y uit A en alle deelverzamelingen X en Y van A

$$(x, y) \in \Delta(\{(X, Y)\}) \iff \Theta(x, y, X, Y) \text{ en}$$

$$(X, Y) \in \Gamma(\{(x, y)\}) \iff \Theta(x, y, X, Y),$$

waarbij $\Theta(x, y, X, Y)$ gedefinieerd is als

$$[x \in (X - Y) \Rightarrow y \notin X] \text{ en } [y \in (Y - X) \Rightarrow x \notin Y].$$

Daar Δ paarsgewijs is volgt uit de dualiteitstheorie van Evers en van Maaren (1984) dat de duale afbeelding Δ^* van Δ zinvol is om nader onder de loep te nemen. Deze is een afbeelding van $P(A \times A)$, de verzameling van alle relaties over alternatieven

uit A , naar $P(P(A) \times P(A))$, de verzameling van alle relaties over deelverzamelingen van A en is als volgt gedefinieerd

$$\Delta^*(R) := \cup\{\mathfrak{R} \in P(P(A) \times P(A)) \mid R \subseteq \Delta(\mathfrak{R})\}.$$

Dus $\Delta^*(R)$ bevat alle relaties op de deelverzamelingen van A welke R in hun vertaling onder Δ bevatten. Nu, dan

$$\begin{aligned} \Delta^*(R) &= \cup\{\mathfrak{R} \in P(P(A) \times P(A)) \mid R \subseteq \cap\{\Delta(\{(X, Y)\}) \mid (X, Y) \in \mathfrak{R}\}\} \\ &= \cup\{\{(X, Y)\} \in P(P(A) \times P(A)) \mid R \subseteq \{\Delta(\{(X, Y)\})\}\} \\ &= \{(X, Y) \in P(A) \times P(A) \mid \Theta(x, y, X, Y) \text{ voor alle } (x, y) \in R\} \\ &= \Gamma(R). \end{aligned}$$

Dus is Γ de duale van Δ en vanaf nu schrijven we derhalve Δ^* in plaats van Γ .

Laat R een relatie zijn over alternatieven uit A . Dan volgt uit de dualiteitstheorie dat

$$R \subseteq \Delta\Delta^*(R).$$

Per definitie volgt dat $(\{y\}, \{x\}) \notin \Delta^*(R)$ precies dan als er een paar (a, b) is uit R met niet $\Theta(a, b, \{y\}, \{x\})$. Dus met $a \in (\{x\} - \{y\})$ en $b \in \{y\}$ of met $b \in (\{y\} - \{x\})$ en $a \in \{x\}$. Dus met $(a, b) = (x, y)$. Dan volgt uit $(x, y) \notin R$ dat $(\{y\}, \{x\}) \notin \Delta^*(R)$ en hiermee met representativiteit dat $(x, y) \notin \Delta\Delta^*(R)$. Dus $\Delta\Delta^*(R) \subseteq R$. Deze en de inclusie hierboven leveren de volgende stelling.

Stelling 2 *Laat R een relatie over alternatieven uit A zijn. Dan is $\Delta\Delta^*(R) = R$.*

Uit de dualiteitstheorie volgt verder dat voor iedere relatie \mathfrak{R} over deelverzamelingen van A dat

$$\mathfrak{R} \subseteq \Delta^*\Delta(\mathfrak{R}) \subseteq \Delta^*\Delta\Delta^*\Delta(\mathfrak{R}) = \Delta^*(\Delta\Delta^*(\Delta(\mathfrak{R}))) = \Delta^*\Delta(\mathfrak{R}).$$

Waarbij de laatste gelijkheid met Stelling 2 volgt. Dus $\mathfrak{R} \subseteq \Delta^*\Delta(\mathfrak{R}) = \Delta^*\Delta\Delta^*\Delta(\mathfrak{R})$. Nemen we voor \mathfrak{R} de grafiek, $gr(C)$, van de waargenomen keuzes C , dan zien we dat $\Delta^*\Delta$ deze uitbreidt en dat deze uitbreiding in één keer haar maximale omvang bereikt. Wel is $\Delta^*\Delta(gr(C))$ een relatie over deelverzamelingen van A , dus niet de grafiek van uitbreiding van de waargenomen keuzes. We besluiten deze paragraaf dan ook met een toelichting hoe uit $\Delta^*\Delta(gr(C))$ een uitbreiding \tilde{C} van C te herleiden valt en welke relatie deze uitbreiding met het standaard model binnen de "Revealed Preference Theory" heeft. Van de vier uitbreidingen die in het oorspronkel artikel staan zullen we er hier slecht twee nader toelichten. Hierbij moeten we ons begrippenapparaat uitbreiden met het volgende. Laat P^C de sterke gebleken voorkeurrelatie aangeven, gedefinieerd door

$$P^C := \{(x, y) \in A \times A \mid \text{er is een } B \text{ uit } \mathbf{B} \text{ met } x \in C(B) \text{ en } y \in (B - C(B))\}.$$

Deze bekende gebleken voorkeur uit "Revealed Preference Theory" speelt een belangrijke rol bij de beschrijving van de vertaling van de grafiek van de waargenomen

keuzes. Hiervoor hebben we bovendien twee operaties op relaties nodig, *complement nemen* en *omdraaien* van een relatie R over alternatieven uit A , welke respectievelijk als volgt gedefinieerd worden

$$\begin{aligned} cR &:= \{(x, y) \in A \times A \mid (x, y) \notin R\}, \\ vR &:= \{(x, y) \in A \times A \mid (y, x) \in R\}. \end{aligned}$$

Verder zijn de volgende twee deelverzamelingen van een keuzesituatie B uit $P(A)$ ten opzichte van een voorkeurrelatie R over alternatieven uit A van belang. De verzameling van *maximale* en de verzameling van *beste* alternatieven binnen B ten opzichte van R , zij zijn respectievelijk gedefinieerd door

$$\begin{aligned} \max(B, R) &:= \{x \in B \mid \text{voor alle } y \in B \text{ als } (y, x) \in R, \text{ dan } (x, y) \in R\}, \\ \text{best}(B, R) &:= \{x \in B \mid \text{voor alle } y \in B \quad (x, y) \in R\}. \end{aligned}$$

Beschouw nu de vertaling van de grafiek van de waargenomen keuzes

$$\begin{aligned} \Delta(\text{gr}(C)) &= \cap\{\Delta(\{(C(B), B)\}) \mid B \in \mathbf{B}\} \\ &= \{(x, y) \in A \times A \mid \Theta(x, y, C(B), B) \text{ voor alle } B \in \mathbf{B}\} \\ &= \{(x, y) \in A \times A \mid x \notin (B - C(B)) \text{ of } y \notin C(B) \text{ voor alle } B \in \mathbf{B}\} \\ &= cvP^C. \end{aligned}$$

Dus de vertaling van de waargenomen keuzes is het complement van de omgedraaide sterke gebleken voorkeur. Merk op dat het asymmetrisch deel van deze sterke gebleken voorkeur gelijk is aan het asymmetrisch deel van het complement van de omgedraaide sterke gebleken voorkeur. Verder verandert deze dubbele operatie cv het symmetrische deel in het onvergelykbare deel en het onvergelykbare deel in het symmetrische. Indien P^C asymmetrisch is, wat het geval is als aan het ‘‘Weak Axiom of Revealed Preference’’ voldaan is, dan is cvP^C volledig en in ordeningstermen gelijkwaardig met de sterke gebleken voorkeur P^C . Zie ook Storcken (1989).

Voor deelverzamelingen X en Y van A , met $X \subseteq Y$,

$$\begin{aligned} (X, Y) \in \Delta^*\Delta(\text{gr}(C)) &\Leftrightarrow \Theta(x, y, X, Y) \text{ voor alle } (x, y) \in cvP^C \\ &\Leftrightarrow x \notin (Y - X) \text{ of } y \notin X \text{ voor alle } (x, y) \in cvP^C \\ &\Leftrightarrow ((Y - X) \times X) \cap cvP^C = \emptyset \\ &\Leftrightarrow (X \times (Y - X)) \subseteq P^C. \end{aligned}$$

Daar $P^C = \cup\{C(B) \times (B - C(B)) \mid B \in \mathbf{B}\}$ zien we dus wederom, maar nu direct, in, dat $\text{gr}(C) \subseteq \Delta^*\Delta(\text{gr}(C))$. Verder volgt dat een paar (X, Y) , met $X \subseteq Y$, in $\Delta^*\Delta(\text{gr}(C))$ zit als alle alternatieven uit X sterk boven alle alternatieven uit $Y - X$ geprefereerd blijken te zijn.

Zoals reeds eerder opgemerkt is $\Delta^*\Delta(\text{gr}(C))$ niet noodzakelijk de grafiek van een uitbreiding van de waargenomen keuzes C . We bespreken nu twee uitbreidingen van deze waargenomen keuzes welke in $\Delta^*\Delta(\text{gr}(C))$ kunnen zitten. Verder geven

we voldoende en noodzakelijke voorwaarden opdat dit zo is. Deze mogelijke uitbreidingen zijn C^{\max} en C^{best} en zijn als volgt gedefinieerd voor een deelverzameling Y van A :

$$\begin{aligned} C^{\max}(Y) &:= \cup\{X \in P(Y) \mid (X, Y) \in \Delta^*\Delta(\text{gr}(C)) \\ &\text{en voor } V \text{ en } W \text{ met } V \subseteq W \subseteq X \\ &(V, W) \in \Delta^*\Delta(\text{gr}(C)) \text{ impliceert } (W - V, W) \in \Delta^*\Delta(\text{gr}(C))\}, \\ C^{\text{best}}(Y) &:= \cap \left\{ (Y - (X - Z)) \in P(Y) \mid \begin{array}{l} Z \text{ is niet leeg } Z \subseteq X \subseteq Y \\ \text{en } (Z, X) \in \Delta^*\Delta(\text{gr}(C)) \end{array} \right\}. \end{aligned}$$

$C^{\max}(Y)$ bestaat uit alternatieven die in een deel X van Y zitten welk ten opzichte van $\Delta^*\Delta(\text{gr}(C))$ zwak boven Y geprefereerd wordt en waarvoor geldt dat voor geneste subdelen $V \subseteq W$ van X , als V zwak boven W geprefereerd wordt ten opzichte van $\Delta^*\Delta(\text{gr}(C))$, dan wordt $W - V$ ook zwak boven W geprefereerd. $C^{\text{best}}(Y)$ bestaat uit die alternatieven waarvoor er geen $Z \subseteq X \subseteq Y$ bestaat zodanig dat ze in $X - Z$ zitten en Z zwak boven X geprefereerd wordt ten opzichte van $\Delta^*\Delta(\text{gr}(C))$. Uit het voorafgaande valt nu af te leiden (voor meer informatie verwijzen we naar het oorspronkelijke artikel) dat voor deelverzamelingen Y van A

$$\begin{aligned} C^{\max}(Y) &= \max(Y, cvP^C) \text{ en} \\ C^{\text{best}}(Y) &= \text{best}(Y, cvP^C). \end{aligned}$$

Het moge duidelijk zijn ofschoon C^{\max} en C^{best} op basis van de waargenomen keuzes gedefinieerd zijn deze niet automatisch een uitbreiding hiervan hoeven te zijn. We besluiten deze paragraaf met een discussie van noodzakelijke en voldoende voorwaarden voor de waargenomen keuzes opdat deze twee afgeleide afbeeldingen wel een uitbreiding van die keuzes zijn. We beginnen met C^{\max} . We zeggen dat de waargenomen keuzes C *blijk geven van dominante keuzes* indien voor alle keuzesituaties B uit \mathbf{B} alle deelverzamelingen X van B , met $(X, B) \in \Delta^*\Delta(\text{gr}(C))$, $X \subseteq C(B)$ indien voor alle deelverzamelingen V en W van X , met $V \subseteq W$,

$$(V, W) \in \Delta^*\Delta(\text{gr}(C)) \text{ impliceert dat } (W - V, W) \in \Delta^*\Delta(\text{gr}(C)).$$

Dus als X zwak beter is dan B , ten opzichte van $\Delta^*\Delta(\text{gr}(C))$, dan zit X in de waargenomen keuze bij B , als bovendien voor ieder tweetal geneste delen van X , zeg $V \subseteq W \subseteq X$, de zwakke preferentie van V boven W ten opzichte van $\Delta^*\Delta(\text{gr}(C))$ ook die van $W - V$ boven W impliceert. De volgende propositie laat zien dat deze voorwaarde noodzakelijk en voldoende is opdat C^{\max} een uitbreiding is van de waargenomen keuzes.

Stelling 3 *De volgende twee uitspraken zijn equivalent.*

- (i) voor alle $B \in \mathbf{B}$ $C^{\max}(B) = C(B)$,
- (ii) C geeft blijk van dominante keuzes.

De noodzakelijke en voldoende voorwaarden opdat C^{best} een uitbreiding is van de waargenomen keuzes, is in termen van bekende voorwaarden voor deze waargenomen keuzes, zoals de volgende propositie laat zien.

Stelling 4 *De volgende twee uitspraken zijn equivalent.*

- (i) $C^{\text{best}}(B) = C(B)$ voor alle $B \in \mathbf{B}$ en $\text{best}(Y, cvP^C) \neq \emptyset$ voor alle $Y \in P(A)$,
- (ii) C voldoet aan het “Weak Axiom of Revealed Preference” (dit betekent dat $R^C \cap vP^C = \emptyset$, waarbij $R^C := \cup\{C(B) \times B \mid B \in \mathbf{B}\}$),
aan het “Strong Axiom of Revealed Preferences” (dit betekent dat P^C acyclisch is) en $C(B) \neq \emptyset$ voor alle $B \in \mathbf{B}$.

We zien dat de waargenomen keuzes van \mathbf{B} naar $P(A)$ uitbreidbaar zijn met een beste keuze gedrag ten opzichte van een preferentie, indien de waargenomen keuzes aan het “Weak” en “Strong Axiom of Revealed Preference” voldoen en bij de waargenomen keuzes altijd iets gekozen wordt. Als aan het “Strong Axiom of Revealed Preference” voldaan is dan is P^C acyclisch, dus asymmetrisch. Nu, dan is cvP^C volledig. Uit de acycliciteit van P^C volgt dan dat het asymmetrisch deel van cvP^C acyclisch is. De volledigheid en de acycliciteit van het asymmetrisch deel van cvP^C garanderen dat er altijd een niet lege beste verzameling alternatieven is voor ieder deel van A . We zien dus dat onze benadering hier volledig overeenkomt met het standaard model uit de “Revealed Preference Theory”.

3 Conclusies

We hebben in paragraaf 2 een axiomatische onderbouwing van het standaardmodel in de “Revealed Preference Theory” gegeven, welke onder de voorwaarden van het “Weak” en “Strong Axiom of Revealed Preference” vergelijkbare resultaten oplevert: de waargenomen keuzes zijn te modelleren met maximaliserend keuzegedrag. Ofschoon het hier gepresenteerde model een ruimere keur aan keuzegedrag schijnt toe te laten, welke geïllustreerd wordt met C^{max} naast C^{best} , blijft grof gezien het gedrag preferentiemaximalisatie. Wezenlijk hier is dat de preferentie niet noodzakelijk transitief hoeft te zijn.

Indien de keuze afhankelijk is van het al dan niet aanwezig zijn van alternatieven, welke bijvoorbeeld informatie over de kwaliteit van andere alternatieven verondersteld wordt te onthullen, dan faalt het hier ingevoerde model. Een aanpassing dan wel vervanging van de naamsonafhankelijkheidsconditie kan hier eventueel een uitweg zijn, want deze conditie introduceert juist deze onafhankelijkheid tussen alternatieven.

Het aanpassen van representativiteit levert geen zinvolle modellen op: zie hiervoor het oorspronkelijke artikel. Ook behoudendheid is een conditie die volgens mij zeer natuurlijk is. Het afzwakken van contextonafhankelijkheid en onvoorin genomenheid zou modellen kunnen opleveren waarbij globale informatie van de waargenomen keuzes een rol kan spelen. Bij het hier beschreven model is vanwege de paarsgewijsheid enkel lokale informatie van de waargenomen keuzes relevant. Bijvoorbeeld, laat een consument negen van de tien keer een artikel van merk a kopen en één van de tien keer eenzelfde soort artikel van merk b, c, d of e . Dit zou deze consument kunnen doen om bijvoorbeeld te variëren of om kwaliteiten van de merken te testen. Dan zou men verwachten dat hij merk a strikt boven de

andere vier merken verkiest. Met de hier besproken modellen, waarbij herhaalde keuzesituaties van elkaar onderscheidbaar gemaakt zouden worden, zal blijken dat hij indifferent is tussen deze merken. Globale informatie, waarbij het aantal keren dat een consument voor een merk kiest relevant is, is hier nodig om deze strikte preferentie te onderkennen.

Tot slot zij opgemerkt dat de vijf condities welke de vertalers vastleggen logisch onafhankelijk zijn, zoals in Ruys en Storcken (1992) aangetoond wordt.

Referenties

- Arrow, K.J. (1948), "The Possibility of a Universal Social Welfare Function", Project RAND, RAD(L)-289, 26 October 1948, Santa Monica, California (hctographed).
- Arrow, K.J. (1959), "Rational choice functions and orderings", *Economica*, NS 26, 121-127.
- Evers, J.J.M. en H. van Maaren (1985), "Duality principles in mathematics and their relations to conjugate functions", *Nieuw Archief voor de Wiskunde*, Vol 4, 23-68.
- Houthakker, H.S. (1950), "Revealed preference and the utility function", *Economica*, NS17, 159-174.
- Richter, M.K. (1971), "Rational Choice", In: *Preferences, Utility and Demand*, Chipman J.S., L.Hurwicz, M.K. Richter en H.F. Sonnenschein, Hartcourt, 29-58.
- Ruys, P.H.M., en A.J.A. Storcken (1992), "Extensions of choice behaviour", KUB, CentER Discussionpaper, nr.9212.
- Samuelson, P.A. (1938), "A note on the pure theory of consumer's behaviour", *Economica*, NS 5, 61-67 en 353-354.
- Sen, A.K. (1971), "Choice functions and revealed preference", *Review of Economic Studies*, Vol 38, 307-317.
- Storcken, A.J.A. (1989), *Possibility Theorems for Social Welfare Functions*, Proefschrift, KUB.
- Ville, J. (1951-1952), "The existence-condition of a total utility function", *Review of Economic Studies*, Vol 19, 123. Reprint from *Anales de l'Université de Lyon*, 1946.
- Wakker, P.P. (1989), *Additive Representations of Preferences*, Dordrecht: Kluwer Academic Publishers.

NETWERKEN, SOCIALE RATIONALITEIT EN ECONOMISCHE INSTITUTIES*

Robert P. Gilles[†]

Virginia Polytechnic Institute & State University

Zoals opgemerkt in de bijdragen in deze bundel is het een genoegen om samen met Pieter Ruys de economische wetenschap te beoefenen. De beste eigenschap van Pieter — niet alleen als wetenschapper, maar ook als intellectueel mens — is dat hij altijd interessante vraagstukken op zodanige wijze weet te formuleren dat ze tot vernieuwende gezichtspunten leiden. Hierbij is een factor dat Pieter één van de meest abstracte denkers is die ik onder economen ontmoet heb.

De inspiratie van Pieter komt van zijn ervaringen in de niet-academische wereld, waaronder zijn raadslidmaatschap van D'66 in de Tilburgse raad, zijn betrokkenheid met de reorganisatie van het familiehotel in Zeist en zijn adviserende functies bij het St. Elisabeth Ziekenhuis in Tilburg en diverse ministeries in Den Haag. Vanuit die ervaringen heeft Pieter gedurende zijn carrière enkele zeer diepe en vernieuwende vraagstukken geformuleerd. Zijn wetenschappelijk werk kent belangrijke vernieuwende thema's zoals neergelegd in zijn modellen over semi-publieke goederen, zijn alternatieve ideeën betreffende de hiërarchisch gestructureerde organisatie van productie binnen bedrijven, de plaats van bestuur en cultuur in de samenleving en de studie van de dienstensector in de economie.¹

Zijn doel hierbij is om de geformuleerde economische vraagstukken te modelleren door middel van wiskundig economische theorieën. Voor Pieter vormt hierbij de gebruikte wiskunde een taal om exacte uitdrukking te geven aan de onderliggende ideeën: Wiskundige modellering is voor Pieter niet een doel op zich, maar vooral een middel om ideeën precies uit te drukken en volgens strikte logische deductie conclusies te formuleren.

Centraal in Pieter's denken staat zijn *tripolaire model*. Deze meta-theorie vormt

*Dit artikel is gebaseerd op vele discussies met Pieter Ruys, Dimitrios Diamantaras en mijn studenten in Blacksburg, in het bijzonder Kate Johnson, Sudipta Sarangi en Tim Chakrabarti. Ik bedank Edward Droste voor zijn commentaar op een eerdere versie van deze bijdrage.

[†]Adres: Department of Economics, Virginia Tech (0316), Blacksburg, VA 24061, USA. *Email:* rgilles@vt.edu.

¹Pieter's gedachtengoed is vrijwel uitsluitend ontwikkeld over een (zeer) lange periode van herformulering en in meerdere bijdragen tot de literatuur. Voor een recente formulering van het probleem van semi-publieke goederen verwijs ik naar van der Laan, Ruys en Talman (2000) en voor het probleem van bestuursstructuren in de hedendaagse diensteneconomie verwijs ik naar Ruys, van den Brink en Semenov (2001) en Ruys (2002). Het moet echter opgemerkt worden dat deze bijdragen slechts tussenstations zijn op de lange weg van groei en herformulering in Pieter's denken.

de basis van zijn denken over de geformuleerde vraagstukken. Pieter is ook hier trouw aan zijn methode en heeft een wiskundige formulering aan deze meta-theorie gegeven. Als denkkader is het tripolaire model zeer vruchtbaar en de basis van de modellen van Pieter en meerdere van zijn promovendi.

Persoonlijk heb ik mijn onderzoekstijd in Tilburg van 1986 tot 1991 als de meest inspirerende jaren van mijn leven ervaren. Achteraf blijkt dat Pieter en ik in deze jaren vraagstukken onderzochten die hun tijd ver vooruit lijken te zijn geweest. We hadden zeer diepe discussies over economische basisconcepten die door bijna alle economen als vanzelfsprekend aanvaard werden en worden. Dit leidde ons naar de vernieuwende modelering van netwerkeconomieën en economische instituties.

Eén van de meest memorabele momenten in mijn samenwerking met Pieter was toen wij ons terugtrokken in het Ruys familiehotel te Zeist om de onderliggende basisconcepten van de modellen in mijn proefschrift (Gilles, 1990) uitgebreid te bespreken. Deze specifieke discussies hebben tot op heden mijn werk geïnspireerd.²

Dit artikel is in menig opzicht een extrapolatie van de discussies over basisideeën met Pieter in Zeist. Recentelijk is er in de economische literatuur veel aandacht gegeven aan de plaats van sociaal kapitaal en cultuur in de economische theorie. Ik probeer in deze bijdrage deze recente ontwikkelingen te combineren met enkele onopgeloste economische vraagstukken en enige ideeën die wij reeds aan het eind van de jaren '80 formuleerden.

1 Onbehagen in de economische wetenschap

In de Engelstalige wereld staat de economische wetenschap ook bekend als de “dismal science”. Er is momenteel onder economen steeds meer een gevoel dat dit inderdaad het geval is. We lijken allen zoekende naar relevantie in onze theoretische exercities.³

Vreemd genoeg zijn economen veelal niet gedreven door de gebeurtenissen in de reële economie, maar eerder door de behoefte om hun artikelen in vooraanstaande academische tijdschriften te publiceren en de kritiek van de referees betreffende deze artikelen te bevredigen. Hierdoor lijkt de ontwikkeling van economische theorie louter nog gedreven door ontwikkelingen in de economische literatuur zelf en niet door belangrijke gebeurtenissen in de wereldeconomie.

Dit is het duidelijkst in een kritische beschouwing van de gevolgen van enkele belangrijke gebeurtenissen in de wereldeconomie gedurende de laatste 20 jaar op de economische wetenschap. De val van de socialistische en communistische economische systemen hebben slechts tot weinig belangwekkende en baanbrekende bijdragen in de economische theorie geleid. Evenmin heeft het falen van de topeconomen van

²Onze discussies betroffen voornamelijk de notie van een economische agent als een entiteit in een sociale structuur en de precieze betekenis van een economisch goed vanuit deze context. Dit is neergelegd in Gilles en Ruys (1994) alsmede Gilles (1996) en vormde de basis voor ons werk over markteconomieën met transactiekosten en thuisproductie van goederen. Hierbij verwijs ik naar Gilles, Diamantaras en Ruys (2000) en Gilles en Diamantaras (2003).

³In Frankrijk heeft dit geleid tot de oprichting van een stroming voor “Post-Autistic Economics”, die zeer kritisch is over de huidige stand van zaken in het neo-Walrasiaanse onderzoeksprogramma. Ik verwijs naar hun web site, <http://www.pae-news.com>, voor artikelen en discussies.

Harvard en andere Ivy League universiteiten om een kapitalistische markteconomie in Rusland op te zetten tot uitgebreid zelfkritisch onderzoek geleid.

Zeer recentelijk heeft de vertrouwenscrisis in het management van grote corporaties en het falen van de controlerende accountants tot vreemde situaties geleid. Economen hebben geen goede theorie om deze crisis te verklaren. De zelfzuchtige verrijking van managers staat namelijk niet in tegenspraak met de doelstelling van een economische agent in onze theorieën.

Het is eveneens onbevredigend dat moderne economische politiek steeds meer gebaseerd lijkt te worden op het antieke neo-klassieke gedachtengoed van het begin van de 20-ste eeuw. Dat was anders aan het eind van de tweede wereldoorlog toen het Keynesiaanse gedachtengoed vol overtuiging werd toegepast om de na-oorlogse economie op gang te brengen en in de jaren '70 en '80 van de 20-ste eeuw ten tijde van de toepassing van de monetaristische theorieën van Milton Friedman. Het lijkt alsof er weinig recente ontwikkelingen in de economische theorie relevant genoeg zijn voor de formulering van het economisch beleid met betrekking tot de 21-ste eeuwse informatie economie.

Ik denk dat de huidige crisis in de economische wetenschap gekarakteriseerd kan worden als een groeiende frictie van het huidige *neo-Walrasiaanse* onderzoeksprogramma⁴ met de harde realiteit van het geobserveerd gedrag in markten en andere economische organisaties.⁵ Om wederom relevant te worden moet de economische wetenschap zich richten op het verklaren van de economisch-maatschappelijke krachten die ten grondslag liggen aan de vorming van economische organisaties. We moeten komen tot theorieën welke kunnen verklaren hoe deze organisatiestructuren gevormd worden en welke krachten deze structuren determineren. Uiteindelijk moet ons doel zijn om succesvolle en structureel efficiënte organisatiestructuren te kunnen ontwerpen. Ik verwijs naar Davis (2002) voor een goede uiteenzetting van enkele aspecten van dit probleem.

In dit artikel bespreek ik eerst vier vraagstukken van een institutioneel-theoretische aard. Ik gebruik deze discussies om te illustreren dat instituties zeer verrassende rollen in de economie speelden en nog steeds spelen. Deze vraagstukken betreffen de functie van ambachtsgilden in de middeleeuwse economie, het fenomeen van de Gouden Eeuw in de Nederlandse economie, het probleem van het kleine muntgeld en, tenslotte, de raadselachtige rol van bedrijven in de kapitalistische markteconomie. De beschouwing van deze problemen leidt tot vragen over enkele fundamentele aspecten van onze hedendaagse neo-Walrasiaanse en speltheoretische modellen.

Vervolgens zal ik een vier-ledige structuur presenteren welke de basis kan vormen voor een alternatieve benadering voor de bestudering van de rol van instituties in de

⁴Het zogenaamde neo-Walrasiaanse onderzoeksprogramma is gebaseerd op drie fundamentele principes: methodologisch individualisme, het principe van individuele rationaliteit en de axiomatische methode. Dit heeft geleid tot de vorming van wiskundige theorieën van economisch gedrag, in het bijzonder algemene evenwichtstheorie en speltheorie. Het is mijn stelling dat vele hedendaagse problemen niet opgelost kunnen worden zonder enige van deze fundamentele principes te verlaten. In het bijzonder het principe van individuele rationaliteit lijkt rijp voor aanpassing.

⁵Voor een uitgebreide evaluatie van het Walrasiaanse gedachtengoed als een onderzoeksprogramma in de zin van Lakatos (1978) verwijs ik naar Weintraub (1985). Tevens verwijs ik voor een herformulering naar Deel I in Gilles (1990).

hedendaagse economie. Aan de basis ligt het individueel kapitaal zoals weergegeven door middel van individuele preferenties en productiemogelijkheden. (Dit is in de engelstalige literatuur aangegeven als “human capital”.) Een complexe structuur van gerelateerde sociale en economische netwerken, welke individuele economische agenten met elkaar verbinden, vormt de basis van hetgeen als “sociaal kapitaal” aangeduid wordt. Ik geef aan dat de vorming van netwerkrelaties — en aldus van sociaal kapitaal — niet begrepen kan worden zonder de toepassing van een “sociale rationaliteit”; individueel egoïstische motieven zijn niet voldoende om zulke sociale relaties te verklaren. Dit leidt tot de conclusie dat een derde structuur bestaande uit de toegepaste sociale gedragsnormen noodzakelijk is voor de fundering van iedere economische organisatiestructuur. Deze gedragsnormen zijn economische instituties in de zin van Ménard (1995).⁶ Tenslotte identificeer ik een vierde structuur bestaande uit hogere sociaal-economische en culturele instituties, welke gebaseerd zijn op de onderliggende netwerken en sociale gedragsnormen.

2 Vier interessante vraagstukken

In deze paragraaf presenteer ik vier vraagstukken uit de economische geschiedenis en de economische theorie om de onderliggende problemen betreffende het algemene vraagstuk van economische instituties te illustreren. Ik presenteer deze problemen vanuit een nauw gezichtspunt: Ik beperk mij tot een korte beschrijving van het probleem en de bespreking van de relevantie van iedere casus voor de probleemstelling van dit artikel. Voor een uitgebreide discussie van iedere casus verwijs ik naar bijdragen in de genoemde literatuur.

2.1 De rol van middeleeuwse ambachtsgilden

Smith (1776) stelde dat ambachtsgilden slechts een negatieve invloed hadden op de (middeleeuwse) proto-industriële economie. Volgens Smith waren deze gilden monopolistische organisaties, die vernieuwing in de kiem smoorden en grote economische inefficiënties veroorzaakten. Sindsdien hebben overheden er alles aan gedaan om gilden uit te bannen en hebben (economische) historici Smith’s hypothese vrijwel klakkeloos overgenomen.

Recentelijk wordt echter erkend dat Smith’s hypothese te negatief was en dat ambachtsgilden wel degelijk een essentiële, positieve rol in de proto-industriële economieën van de late middeleeuwen vervulden. Veel van de negatieve verhalen over machtsmisbruik van gilden blijken zwaar overtrokken. (Zie bijvoorbeeld Hicksen en Thompson, 1991, Epstein, 1998, en Prak, 2002.) Er is echter nog geen algemeen geaccepteerde theorie welke de rol van ambachtsgilden in de economie verklaart.

⁶Ménard (1995) ontwikkelt diverse theoretische concepten en beschouwt een economische institutie als “manifested in a long-standing historically determined set of stable, abstract and impersonal rules, crystallized in traditions, customs, or laws, so as to implement and enforce patterns of behavior governing the relationships between separate social constituencies”. (Ménard, 1995, p. 167.)

Ik benadruk dat ambachtsgilden onderscheiden dienen te worden van *handels-gilden*. Handelsgilden legden zich toe op de bescherming van de activiteiten van (internationale) handelaren. Dit resulteerde in organisaties zoals het Duitse Hanse systeem. Greif, Milgrom en Weingast (1994) en Greif (2000) ontwikkelen een theoretische verklaring voor het bestaan van zulke organisaties. Hun verklaring is dat handelsgilden (collectieve) onderhandelingsmacht aan de handelaren gaven zodat hun positie beter beschermd werd.⁷

Ons begrip van ambachtsgilden is relatief beperkt. Alhoewel er voldoende beschrijvend materiaal over ambachtsgilden beschikbaar is, zoals Epstein (1991), zijn er slechts weinig theoretische verklaringen beschikbaar. Het is verbijsterend te constateren dat er slechts een handvol bijdragen hierover in de historisch economische literatuur te vinden zijn. Ik bespreek twee mogelijke verklaringen.

Hickson en Thompson (1991) beargumenteren dat ambachtsgilden opkwamen als antwoord op de bedreiging van eigendom in de middeleeuwse stedelijke economieën. Hierbij ging het vooral om de bescherming van kennis van productiemethoden. De functie van een gilde was tweeledig, namelijk het bieden van interne en externe bescherming. Betreffende interne bescherming kan men constateren dat de overheid in de vorm van het stedelijk bestuur geen middelen had om bescherming voor investeringen in de lokale economie te bieden; Ambachtsgilden namen deze interne taak waar.

Ten tweede hadden ambachtsgilden een externe beschermingstaak. De lokale edelman bedreigde namelijk veelal de productiemiddelen van een stedelijke economie door behalve geldelijke middelen ook leden van de mannelijke bevolking op te eisen voor de oorlogsvoering. Deze inefficiënte vorm van belasting werd door de gilden omgevormd tot meer efficiënte vormen van belasting. Productieve leden van de stedelijke economie werden zodoende gevrijwaard van deze militaire dienst. Geldelijke bijdragen voor de oorlogsvoering werden op meer efficiënte wijze binnen de stedelijke economie geallokeerd.

Epstein (1998) stelt echter dat er geen bewijs in de data over middeleeuwse gilden is voor de verklaring van Hickson en Thompson (1991). Hij stelt als alternatief dat gilden verklaart moeten worden als organisaties voor de opleiding van ambachtsslieden en daarmee de verspreiding en ontwikkeling van nieuwe productietechnologieën. Ambachtsgilden vervulden een essentiële rol in de bescherming van de ambachtelijke economie bij de afwezigheid van goed functionerende arbeidsmarkten en algemeen toegankelijk opleidingen.

Alhoewel de gepresenteerde verklaringen niet volledig lijken, is het duidelijk dat instituties zoals ambachtsgilden een essentiële rol vervulden in de proto-industriële economie. Deze economische instituties waren een gevolg van de afwezigheid van enkele essentiële mechanismes zoals de bescherming van eigendom en goed func-

⁷Zijdelings hieraan gerelateerd is Greif (1994) die de praktijken van twee verschillende handelsculturen beschrijft. Hij concludeert dat individualistische handelssamenlevingen — zoals die in Genua — tot fundamenteel andere economische instituties leidde dan collectief georganiseerde handelssamenlevingen. In de individualistische samenlevingen was er meer specialisatie en werden meerdere economische beroepsklassen gevormd. In collectieve handelssamenlevingen — zoals die van de Maghribis — werden handelscontracten collectief uitgevoerd en op basis van familierelaties in plaats van contractuele relaties. De samenleving van de Maghribis bleef daardoor gesloten en meer homogeen.

tionerende arbeidsmarkten: Ambachtsgilden zijn een institutioneel antwoord op het ontbreken van andere economische instituties.

2.2 Waarom was er een Nederlandse “Gouden Eeuw”?

De *Gouden Eeuw* in de Nederlandse economische geschiedenis is een unieke periode in de proto-industriële wereld, namelijk een periode van significante economische groei. De Nederlandse Republiek van die tijd kan beschouwd worden als de eerste economische wereldmacht. De periode van de *Gouden Eeuw* is gewoonlijk gezien als de periode tussen de onafhankelijkheid in 1588 en het einde van de regering van Stadhouders Willem III in 1702. (Zie Israel, 1995.) Andere auteurs, zoals Hickson en Thompson (1991), plaatsen het einde van de *Gouden Eeuw* eerder tussen 1650 en 1672. Deze periode van wonderbaarlijke economische ontwikkeling roept de vraag op waarom de Nederlandse Republiek zo’n unieke positie had. Ondanks vele lijvige werken van historici en economisch historici is er geen algemeen geaccepteerde verklaring voor dit fenomeen.

Een aantal aspecten van de Nederlandse republiek kunnen echter geïdentificeerd worden als belangrijke factoren in deze ontwikkeling. Ik wil mij daar voornamelijk op richten in deze uiteenzetting. Voor uitgebreidere behandeling verwijs ik naar het standaardwerk van Israel (1995) en de studies van de Vries en van der Woude (1997) en Prak (2002).

Ten eerste was de Nederlandse rijkdom gebaseerd op het grote succes van de stapelmarkt, in het bijzonder de graanhandel met het Oostzeegebied.⁸ De opkomst van het mercantilisme in Europa was een belangrijke factor in de neergang van de Nederlandse economie in de tweede helft van de 17-de eeuw. Het mercantilisme propageerde de toepassing van hoge tariefstructuren om de binnenlandse economie te beschermen tegen invloeden van buiten. Dit had grote stagnatie in de handel met Nederland tot gevolg.

Ten tweede had de Nederlandse economie diverse “moderne” aspecten zoals beschreven in de Vries en van der Woude (1997): markten voor goederen alsmede productiefactoren waren relatief goed ontwikkeld; het eigendomsrecht was eveneens goed onderbouwd in een volledig gedecentraliseerde staatsvorm, waarin de regenten hun economisch profijtelijke activiteiten goed wisten te waarborgen; en de aanwezige productietechnologieën waren in staat om een voortgaande ontwikkeling en vernieuwing mogelijk te maken.⁹

Tenslotte was de Nederlandse Republiek een toevluchtsoord voor vele economisch succesvolle groeperingen die elders vervolgd werden. Dit bracht institutionele vernieuwing met zich mee, die binnen de gedecentraliseerde structuur van de Republiek tot zeer vruchtbare resultaten leidde. Hoogtepunten van deze vernieuwing was de creatie van de succesvolle Amsterdamse beurs en de VOC.

Het moge duidelijk zijn dat de institutionele structuur van de Nederlandse samenleving die zich tijdens deze periode ontwikkelde, een zeer belangrijke rol speelt

⁸Het aandeel in de vergaring van rijkdom via de VOC was relatief klein. Zie Prak (2002) en de Vries en van der Woude (1997).

⁹Hickson en Thompson (1991) merken op dat het Nederlandse gildensysteem hierin een grote rol speelde. De ondergraving van dit systeem door de Raadspensionaris Joan de Witt tussen 1650 en 1670 was volgens hen mede debet aan de verzwakking van de Nederlandse economie.

in de verklaring van de opkomst van de Nederlandse *Gouden Eeuw*. Deze casus toont tevens aan dat economische ontwikkeling zeer gevoelig is voor de institutionele structuur van de economie.

2.3 Het probleem van het kleine muntgeld

Een opmerkelijk voorbeeld van institutionele vernieuwing in de economische geschiedenis is de ontwikkeling van geld. Na het verval van het Romeinse Rijk werd pas in de 7-de eeuw weer muntgeld geïntroduceerd. De zilveren Denarius of “Penning” was de enige munt die in de vroege middeleeuwen in omloop was. (Spufford, 1988, Chapter 2.) Met de ontwikkeling van de Europese economie en de versnipperde structuur van de Europese samenleving werd er al snel een veelheid van nieuw (vooral lokaal) muntgeld gecreëerd.

De zilverwaarde van de Penning als de basiseenheid werd hierbij snel verlaagd en de behoefte aan grotere munten werd opgevangen door de introductie van munten met grotere zilverwaarde (deze munten stonden bekend als de “Grosso” of “Groot”) en gouden munten (bekend als de “Gulden”, het “Florijn” of het “Dukaat”). Deze munten stonden formeel in zekere verhoudingen tot elkaar.¹⁰ Al deze muntsoorten werden verhandeld op verschillende markten. In het bijzonder door de zilver- en goudwaarde van deze munten waren er waardingsverschillen die niet corresponderden met de officiële waarden van deze munten.

Gewoonlijk werden de kleinere muntsoorten geleidelijk minder waard ten opzichte van de gouden munt, welke nu het steunpunt van het monetaire systeem was. Dit ging hand in hand met een groeiend tekort van de kleinere munten zodat er regelmatig in de Europese economie een monetaire crisis het hoofd geboden moesten worden. Deze monetaire krachten veroorzaakten regelmatig koerscorrecties en er waren diverse grote monetaire herstructureringen noodzakelijk. Het is duidelijk dat dit de economische vooruitgang belemmerde.

Pas in de 19-de eeuw werd een oplossing gevonden in de vorm van de *Gouden Standaard*. Hierbij werd de intrinsieke waarde uit het kleingeld verwijderd door het te vervangen door koperen munten. Alleen de grootste munt behield zijn goud- of zilverwaarde. De waarde van het kleingeld werd daardoor volledig gebaseerd op vertrouwen in de economie en de overheden die de convertibiliteit van het geld garandeerden. De gevolgen waren reusachtig; vele nieuwe geldvormen zoals chartaal papiergeld en giraal geld werden gecreëerd en geaccepteerd. Dit ondersteunde de enorme economische groei in de westerse economieën vanaf het einde van de 19-de eeuw.

De les van deze ontwikkeling is dat een paradoxale aanpassing van de institutionele structuur van de economie grote positieve gevolgen kan hebben. Het ging hier in principe om de verzwakking van het marktsysteem, zelfs de verwijdering van markten en de vervanging van private goederen (de diverse muntsoorten) door een publiek goed (convertibel geld). Sargent en Velde (2002) presenteren een specifiek macroeconomisch model dat deze paradox verklaard.

¹⁰De Groot was gewoonlijk tussen 4d en 12d waard. De gouden Gulden was tussen 80d en 240d waard. Zie ook de Appendix in Spufford (1988) en Sargent en Velde (2002), Chapter 4.

2.4 De theoretische paradox van het bedrijf

Waarom vindt productie in een markteconomie vrijwel volledig binnen het verband van bedrijven plaats? Op het eerste gezicht lijkt dit een vreemde vraag, maar de existentie van het bedrijf als de belangrijkste organisatievorm voor productie is in tegenspraak met de geaccepteerde microeconomische fundering van de Walrasiaanse theorie van het marktmechanisme.

Coase (1937) was de eerste die zich deze vraag stelde. In een Walrasiaans systeem van volledige markten zoals beschreven is in Debreu (1959) zouden alle economische activiteiten via marktcontracten geregeld moeten worden. Coase stelde echter dat markten imperfect zijn en dat de aanwezigheid van transactiekosten zulke marktcontracten voor arbeidsrelaties te kostbaar maken. Aldus is het efficiënter om deze arbeidsrelaties via onvolledige arbeidscontracten binnen de organisatie van een hiërarchisch bedrijf te sturen.

Met de verklaring van Coase (1937) kwamen diverse onopgeloste problemen naar voren zoals beschreven in, bijvoorbeeld, Simon (1951 en 1991), Grossman en Hart (1986) en Hart en Moore (1990). De conclusie is dat alvorens te begrijpen waarom bedrijven überhaupt bestaan, wij moeten onderzoeken hoe ze functioneren en hoe transactiekosten een rol spelen binnen een markteconomie.

In de “New Institutional Economics” zijn er diverse theorieën ontwikkeld die beschrijven hoe bedrijven zouden functioneren. De theorie van de principal-agent relatie speelt daarbij een cruciale rol. Deze is verder uitgewerkt door Simon (1951), Grossman en Hart (1986) en Maskin en Tirole (1999). Momenteel zijn er nog grote onopgeloste vraagstukken binnen dit onderzoeksgebied. Het grootste probleem lijkt dat betreffende de analyse van gedrag binnen complexe hiërarchische organisaties te zijn. De eerste aanzetten hiertoe zijn gegeven door Ruys en van den Brink (1999), Rajan en Zingales (2001) en van den Brink en Gilles (2002). Al deze modellen hebben het nadeel dat ze slechts een enkel aspect van deze complexe materie belichten.

Transactiekosten binnen het marktmechanisme zijn nog minder begrepen dan het functioneren van hiërarchisch georganiseerde bedrijven. Onderzoek heeft diverse inzichten opgeleverd, maar het beeld blijft fragmentarisch. In het bijzonder noem ik het werk van mijzelf en Pieter Ruys betreffende dit onderwerp. In Gilles en Ruys (1994) werd een eerste aanzet gegeven om deze problemen te classificeren. In Gilles, Diamantaras en Ruys (2000) en Gilles en Diamantaras (2003) worden algemene evenwichtsmodellen ontwikkeld waarin transactiekosten centraal staan. In deze benadering zijn transactiekosten direct gerelateerd aan het ontwerp van markt-instituten. Onze analyse beperkt zich echter tot de vraag of volledige efficiëntie binnen zulke institutionele marktsystemen haalbaar is. Met genoeg ontwikkelde prijssystemen is dat inderdaad het geval. Verder onderzoek is noodzakelijk om deze materie beter te begrijpen en meer praktische conclusies te trekken.

3 Een economisch institutionele benadering

Mijn algemene hypothese, welke geïllustreerd wordt door de vier bovenstaande discussies, is dat economische instituten een centrale rol spelen in de economische

ontwikkeling. Deze hypothese wordt ook onderbouwd door Greif (2000) vanuit een alternatief perspectief.¹¹ Ik argumenteer, in tegenstelling tot Greif (2000), dat de moderne speltheoretische en neo-Walrasiaanse theorieën ons te weinig houvast bieden voor een goed begrip van de rol en de vorming van economische instituties. Dit wordt duidelijk door de eenvoudige speltheoretische modellen in Greif (1994) en Greif, Milgrom en Weingast (1994) nader te beschouwen. Deze modellen zijn geschikt om specifieke situaties te verklaren, maar missen de breedheid om algemene conclusies te trekken. In deze paragraaf zal ik proberen aan te geven waarom een fundamentele aanpassing van het onderzoeksprogramma in de economische wetenschap noodzakelijk is om de vorming van economische instituties te begrijpen.

Mijn uitgangspunt is dat onze traditionele modellen van een markt vanuit een te eenvoudig perspectief gestructureerd zijn. Gewoonlijk wordt een markt gemodelleerd als een simpele structuur waarin vraag en aanbod geconfronteerd worden. Mijn hypothese is echter: *Markten zijn handelsnetwerken*. In deze handelsnetwerken worden individuen samengebracht die onderdelen van hun individueel human capital met elkaar ruilen. Deze ruil leidt tot waardevermeerdering en vormt de basis van economische productie en economische groei.¹²

Eenvoudige ruilnetwerken verdichten zich al snel tot meer complexe netwerken waarin verschillende soorten van relaties door elkaar lopen. Bedrijven zijn bijvoorbeeld hiërarchische netwerken van onvolledige arbeidsrelaties binnen diverse ruilnetwerken. Het moet ook duidelijk zijn dat deze verschillende netwerken met elkaar verbonden zijn.¹³ In de sociologie wordt de term “embeddedness” gebruikt om de plaats van een individu in zo’n structuur aan te duiden.¹⁴ De gehele economie moet aldus beschouwd worden als een complexe structuur van gerelateerde netwerken waarin individueel gedrag gestuurd wordt door sociale gedragsnormen.

In de volgende deelparagrafen zal ik deze theoretische opzet nader uitwerken.

3.1 De netwerken om ons heen: sociaal kapitaal

Hoe zien de netwerken om ons heen er eigenlijk uit? Recent hebben enkele natuurkundigen zich hierover gebogen. Een overzicht van deze bevindingen wordt gegeven door Albert en Barabási (2002) en Dorogovtsev en Mendes (2002). Zij concluderen

¹¹Greif stelt het primitieve concept van tweezijdig profijtelijke handel centraal. Hij concludeert dat de bestudering van economische en sociale instituties een centrale rol moet krijgen binnen de economische wetenschap. Dit standpunt staat al veel langer centraal in de institutionele historische economie zoals verwoord in North (1990), maar schijnt tot nu toe weinig navolging te krijgen.

¹²In een iets andere vorm wordt dit proces ook gemodelleerd in Romer (1994) en Gilles en Diamantaras (2003). In deze bijdragen wordt namelijk het ontwikkelingsproces van een economie beschouwd als de marketing van nieuwe goederen. Zoals aangegeven in mijn discussie van het probleem van het kleine muntgeld zijn hierbij enige kanttekeningen te plaatsen.

¹³Een aanzet tot de beschouwing van een economie als een complexe structuur van gerelateerde netwerken en organisaties werd reeds gegeven door Simon (1991) en is nader beschouwd in Kirman (1997) en Loasby (2000).

¹⁴Zie Granovetter (1985) en Burt (1992). Ik verwijs ook naar Davis (2002) voor een soortgelijke discussie. Davis stelt dat de economische wetenschap pretendeert gebaseerd te zijn op methodologisch individualisme echter zonder een theorie van het individu te hebben. Daaruit leidt hij af dat de fundering eerder gezocht moet worden in sociale structuren, in het bijzonder op het concept van sociale “embeddedness”.

dat er in principe twee fundamentele statistische netwerk eigenschappen zijn: De “small world” en schaalvrije netwerk eigenschappen.¹⁵

Ik geef een kort overzicht van beide netwerk eigenschappen:

“Small world” netwerken. Een “small world” netwerk is gekarakteriseerd door een hoge “clustering” van agenten en simultaan de korte paden die agenten in de verschillende van het netwerk met elkaar verbinden. De hoge clustering duidt op hoge graad van organisatie en ordening. Daarnaast duiden de relatief korte verbindingen tussen alle agenten, waarvan deze soort netwerken zijn naam ontleent, op de aanwezigheid van vele brug-relaties in het netwerk.¹⁶ Deze twee eigenschappen staan op gespannen voet met elkaar. In een netwerk van hoge ordening zijn de paden tussen de verschillende agenten gewoonlijk zeer lang. De lage gemiddelde padlengte duidt eerder op een netwerk waarin alle relaties een gelijke kans hebben op aanwezigheid, een zogenaamd “random” netwerk. Vandaar de uitdrukking “small world networks are between complete order and complete disorder.”

Wiskundig-statistische modellen van “small world” netwerken zijn slechts van zeer recente datum. De aanzet is gegeven in het werk van Duncan Watts en is samengevat in zijn boek Watts (2003). Ik benadruk dat deze modellen *niet* gedragstheoretisch van aard zijn: De relaties tussen de verschillende agenten worden volgens een statistische verdeling aangebracht en worden niet bepaald door (rationeel) keuzegedrag van de individuen. In dat opzicht zijn deze modellen puur beschrijvend.

Een eerste poging om tot een gedragstheoretisch model van een “small world” netwerk te komen is gegeven in Johnson en Gilles (2000). Hierin baseren wij ons op het speltheoretisch concept van paarsgewijze stabiliteit van Jackson en Wolinsky (1996). In een ruimtelijk model waarin de kosten van het formeren van een relatie recht evenredig is aan de ruimtelijke afstand, kan bewezen worden dat de resulterende paarsgewijs stabiele netwerken inderdaad “small world” eigenschappen hebben. Ik benadruk dat de ruimtelijke structuur een sociaal-economische interpretatie heeft zoals eerder ontwikkelt in Gilles (1990) en nader uitgewerkt in Johnson en Gilles (2000). “Afstand” is hierbij een sociaal-economisch begrip en geeft aldus uitdrukking aan begrippen zoals inkomensverschillen, preferentieovereenkomsten alsmede fysieke afstand.

Schaalvrije netwerken. De andere belangrijke klasse van statistische netwerken is die van schaalvrije netwerken. Hierbij heeft de verdeling van de uitgraden van de punten in de onderliggende graaf een Poisson verdeling. Dit betekent dat er zeer weinig agenten met zeer veel relaties zijn, weinig agenten met veel relaties en veel agenten met weinig relaties. Agenten met veel verbindingen

¹⁵Een statistisch model voor een algemenere klasse van deze netwerken is ontwikkeld in Newman, Strogatz en Watts (2001).

¹⁶Ik noem hier de uitdrukking “six degrees of separation”, welke aangeeft dat iedere persoon verbonden is met ieder ander persoon in de wereld via een rij van gemiddeld zes gerelateerde personen. Albert en Barabási (2002) laten zien dat in het *World Wide Web* twee arbitraire web pagina's een gemiddelde afstand van 19 “clicks” hebben. Het World Wide Web is dus ook een “small world” netwerk. Voor een uitgebreide uiteenzetting verwijs ik naar Watts (2003).

worden aangeduid als netwerk “hubs”. Albert en Barabási (2002) geven een overzicht van de situaties waarin schaalvrije netwerken toepasbaar zijn. Hieronder vallen het Internet, het netwerk van relaties tussen acteurs, de biochemische reaktierelaties tussen stoffen in een cel en de netwerken bestaande uit treinverbindingen, wegen en vliegroutes tussen steden.

Voor de vorming van schaalvrije netwerken heeft men nog geen algemeen gedragstheoretische modellen ontworpen. Er is wat economisch onderzoek gedaan naar de vorming van hubs in specifieke situaties zoals de vliegindustrie (Hendricks, Piccione en Tan, 1999) en transportnetwerken (Konishi, 2000). Daarbij kwam naar voren dat transactiekosten een centrale rol spelen in de verklaring van het ontstaan van deze hubs.

Een evolutionaire beschouwing van verdichtende sociale netwerken heeft in de theoretisch sociologische literatuur geleid tot het begrip van sociaal kapitaal. *Sociaal kapitaal* is de sociaal-economische waarde die resulteert uit langdurige relaties tussen personen. Sociaal kapitaal vormt de basis van vele sociaal-economische instituties. Het fundeert zulke begrippen als vertrouwen, loyaliteit, verbondenheid en vriendschap. Deze afgeleide concepten vormen het fundament voor gedragsnormen en het goed functioneren van organisaties. Sociaal kapitaal is echter op zichzelf geen institutie in de zin van Ménard (1995).

In de sociologische literatuur is veel onderzoek gedaan naar de fundering van institutionele gedragsnormen in sociaal kapitaal. Ik verwijs hierbij naar, bijvoorbeeld, Burt (1992) en Buskens (1999). Eveneens is een verbinding gelegd tussen sociaal kapitaal en individueel human capital. Namelijk de aanwezigheid van sociaal kapitaal draagt bij tot de vorming van human capital. Ik verwijs naar Coleman (1988) voor een studie die hechte familiestructuren koppelt aan het succesvol doorlopen van scholingsprogramma's.

De opgang van het begrip sociaal kapitaal heeft tot een uitgebreide discussie in de economische wetenschap geleid. Het begrip gaat namelijk direct in tegen de fundamentele norm van het methodologisch individualisme en heeft als zodanig geen plaats in het neo-Walrasiaanse onderzoeksprogramma. Voor een inleiding betreffende de discussie verwijs ik naar Robison, Schmid en Siles (2002). Daar wordt ook een interessante alternatieve definitie van het begrip sociaal kapitaal geïntroduceerd.

3.2 Sociale rationaliteit

In deze paragraaf zal ik aantonen dat een gedragstheoretisch model over de formatie van sociale netwerken noodzakelijk gebaseerd dient te worden op een concept van sociale rationaliteit. Daartoe behandel ik de theorie van netwerkvorming welke in Gilles en Sarangi (2003) ontwikkeld is.

Laat $N = \{1, \dots, n\}$ een eindige verzameling economische agenten zijn. Een **binaire relatie** tussen agenten i en j in N is gewoonlijk aangegeven met $ij := \{i, j\}$. $g^N := \{ij \mid i, j \in N, i \neq j\}$ is nu de verzameling van alle mogelijke binaire relaties op N . Een **netwerk** is gedefinieerd als een verzameling relaties $g \subset g^N$.

Laat $g \subset g^N$ een arbitrair netwerk op N zijn. Voor agent $i \in N$ introduceer ik $N(i, g) := \{j \in N \mid ij \in g\}$ als de verzameling van direct gerelateerden. Voor ieder paar agenten $i, j \in N$ met $i \neq j$ definieer ik $g + ij := g \cup \{ij\}$ als het netwerk dat

resulteert door het toevoegen van de relatie ij en $g - ij := g \setminus \{ij\}$ als het netwerk dat resulteert door wegname van de relatie ij .

Voor iedere agent $i \in N$ definieer ik de functie $\varphi_i: \{g \mid g \subset g^N\} \rightarrow \mathbb{R}_+$ als haar **netwerk nutsfunctie** welke aan ieder netwerk $g \subset g^N$ een nutswaarde $\varphi_i(g) \geq 0$ toekent dat beschrijft hoe agent i deelname in netwerk g apprecieert. We nemen aan dat het lege netwerk geen nut heeft: $\varphi_i(\emptyset) = 0$.

Tenslotte introduceer ik een stabiliteitsconcept voor netwerken gebaseerd op deze netwerk nutsfuncties:

- A. Een netwerk $g \subset g^N$ is **relatieveel stabiel** indien voor iedere $i \in N$ en $j \in N(i, g)$ geldt dat $\varphi_i(g - ij) \leq \varphi_i(g)$.
- B. Een netwerk $g \subset g^N$ is **sterk relationeel stabiel** indien voor iedere $i \in N$ en $M \subset N(i, g)$ geldt dat $\varphi_i(g \setminus h_M) \leq \varphi_i(g)$ met $h_M = \{ij \in g \mid j \in M\} \subset g$.

Ik benadruk dat de definitie van relationele stabiliteit niet gedragstheoretisch is, maar slechts een beschrijving geeft van de (zwakke) eigenschap dat de verwijdering van een relatie *niet* profijtelijk is voor *beide* agenten die deze relatie vormen. Het netwerk is dus stabiel in de zin dat niemand één van hun relaties zou willen verwijderen. Sterke relationele stabiliteit breidt dit uit tot het verwijderen van één of meerdere relaties.

Vervolgens behandel ik een speltheoretisch model betreffende netwerkformatie. Dit model is bekend uit de speltheoretische literatuur¹⁷ waarbij kosten van relatievorming zijn toegevoegd.

Relatievorming is een binair proces, waarin twee agenten tot een vergelijk komen. Dit betekent dat beide agenten in moeten stemmen met de totstandkoming van de relatie. Hierbij moeten vormingskosten expliciet in beschouwing genomen worden. Laat voor agent $i \in N$ de individuele vormingskosten gegeven zijn door de vector $c_i = (c_{ij})_{j \neq i} \in \mathbb{R}_+^{N \setminus i}$. $c_{ij} \geq 0$ representeert aldus de kosten van agent i om een relatie ij met agent j aan te gaan.

Ik ben nu in staat om het strategisch spel te introduceren. Laat voor $i \in N$ de aktieruimte gegeven zijn door

$$A_i = \{(\ell_{ij})_{j \neq i} \mid \ell_{ij} \in \{0, 1\}\}.$$

Agent i maakt contact met agent j indien $\ell_{ij} = 1$. Een relatie $ij \in g^N$ is gevormd tussen i en j als beide agenten contact zoeken, ofwel indien $\ell_{ij} = \ell_{ji} = 1$.

Zij $A := A_1 \times \dots \times A_n$ en laat $\ell \in A$. Het netwerk resulterende uit de aktievector ℓ is nu gegeven als

$$g(\ell) = \{ij \in g^N \mid \ell_{ij} = \ell_{ji} = 1\}.$$

Tenslotte definiëren we voor iedere agent $i \in N$ een speltheoretische nutsfunctie gegeven door

$$\pi_i(\ell) = \varphi_i(g(\ell)) - \sum_{j \neq i} \ell_{ij} \cdot c_{ij},$$

¹⁷Voor een overzicht van de recente ontwikkelingen in de gedragstheoretische benadering van netwerkvorming verwijs ik naar Jackson (2001). Hij behandelt de meest bekende en geaccepteerde speltheoretische concepten in deze theorie en haar toepassingen.

waarbij $c = (c_{ij})_{i \neq j}$ de gegeven relatie vormingskosten zijn. Het paar (φ, c) genereert op deze wijze een spel in strategische of normale vorm gegeven door (A, π) . We beschouwen nu de netwerken die resulteren uit de Nash evenwichten in het spel (A, π) .¹⁸ Nu wordt een netwerk $g \subset g^N$ ondersteund door het Nash evenwicht $\ell \in A$ in het spel (A, π) indien $g \equiv g(\ell)$.

Stelling 1 *Laat $\varphi: \{g \mid g \subset g^N\} \rightarrow \mathbb{R}_+^N$ en $c \geq 0$ gegeven zijn.*

- (a) *Een netwerk $g \subset g^N$ wordt ondersteund door een Nash evenwicht in het spel (A, π) dan en slechts dan als g sterk relationeel stabiel is ten opzichte van de kosten-gecorrigeerde netwerk nutsfunctie ψ gegeven door*

$$\psi_i(g) = \varphi_i(g) - \sum_{j \in N, ij \in g} c_{ij}.$$

- (b) *Indien $c_{ij} > 0$ voor alle $i, j \in N$ met $i \neq j$, dan wordt het lege netwerk ondersteund door een strict Nash evenwicht in het spel (A, π) .*

Voor een bewijs van Stelling 1 verwijs ik naar Gilles en Sarangi (2003).

De gevolgen van Stelling 1 zijn verstrekkend. Een Nash evenwicht beschrijft de stabiele situaties die volgen uit de maximalisatie van individuele nutswaarden. Individueel rationeel gedrag leidt aldus tot een Nash evenwicht. Stelling 1 stelt nu dat in een situatie met relationele vormingskosten individueel rationeel gedrag in principe slechts resulteert in lege netwerken. Het vormen van een relatie onder positieve vormingskosten is namelijk niet profijtelijk indien de andere agent geen beslissing neemt om ook tot vorming van de relatie te besluiten.

Stelling 1 toont aldus aan dat individueel rationeel gedrag op zich niet voldoende is om binaire relatievorming te verklaren. Men moet de stap zetten naar hogere rationaliteitsvoorwaarden, namelijk naar *sociale rationaliteit* waarin agenten erop vertrouwen dat alle andere agenten ook de sociaal rationele beslissing zullen nemen om relaties aan te gaan. Deze sociale rationaliteitsvoorwaarden staan in de literatuur bekend als “farsightedness”. Een eerste aanzet tot de ontwikkeling van dit soort rationaliteitsbegrip werd gegeven door Chwe (1994). Sindsdien hebben meerdere theoretici hun licht laten schijnen over dit onderwerp.¹⁹

Voor een uitgebreide discussie en ontwikkeling van dit model van netwerkvorming met relatie vormingskosten verwijs ik naar Gilles en Sarangi (2003).

In de vorige alinea's heb ik vanuit een puur theoretisch perspectief proberen aan te tonen dat sociaal rationeel gedrag noodzakelijk in beschouwing genomen moet worden om de vorming van netwerken en sociaal kapitaal te verklaren. Er is echter ook overweldigend bewijsmateriaal beschikbaar uit de experimentele economie dat dit het geval is. Het zou te ver voeren om een uitgebreide discussie van de bevindingen in de experimentele economie te voeren. In plaats daarvan noem ik slechts het werk betreffende het zogenaamde “ultimatum bargaining” probleem.

¹⁸Formeel is $\ell \in A$ een *Nash evenwicht* indien $\pi_i(\ell) \geq \pi_i(L_i, \ell_{-i})$ voor iedere alternatieve strategie $L_i \in A_i$.

¹⁹“Farsightedness” staat in de sociologische literatuur bekend als “trust”, ofwel vertrouwen. Voor een overzicht van deze sociologische benadering verwijs ik naar Buskens (1999).

Güth, Schmittberger en Schwarze (1982) waren de eersten om twee individuen een eenvoudige onderhandelingsprocedure betreffende de verdeling van een som geld te laten uitvoeren. In deze procedure stelt de eerste persoon een verdeling voor. Vervolgens kan de tweede persoon dit voorstel afwijzen of aannemen. Indien het voorstel wordt afgewezen, krijgen beide individuen niets. In het andere geval wordt het voorstel uitgevoerd.

In het experimentele onderzoek dat resulteerde na Güth, Schmittberger en Schwarze (1982) wordt aangetoond dat individuen niet zozeer gedreven worden door egoïstische, individueel rationele motieven, maar veeleer door sociale gedragsnormen gebaseerd op mededogen en gelijke behandeling. Voorstellen waarin de eerste persoon teveel opeist (meer dan 70%) worden veelal afgewezen. Dit maakt duidelijk dat individuen inkomen op willen offeren om sociale normen te bevestigen tegenover de andere persoon. Een gemiddelde van 60% wordt wel acceptabel geacht; Een min of meer gelijke verdeling van de som geld staat dus centraal voor veel individuen.

Over het algemeen zijn ook in andere experimentele studies soortgelijke conclusies getrokken.²⁰ Sociale gedragsnormen lijken ons individueel gedrag meer te sturen dan het aloude economische principe van puur egoïstisch gedrag.

3.3 De invloed van culturele instituties

Zeer recentelijk hebben diverse theoretisch economen zich gericht op de bestudering van de invloed van sociale instituties en cultuur in het algemeen op het gedrag van economische agenten. Dit gebeurt vrijwel uitsluitend via een speltheoretische benadering.

Mijn discussie was tot op heden beperkt tot de interactie van human capital, sociaal kapitaal (in de vorm van complexe netwerken) en sociale gedragsnormen. Zoals Bowles (1998) en Okuno-Fujiwara (2002) argumenteren zijn er eveneens interacties tussen deze elementen en hogere gedragsnormen en culturele instituties. In economische zin zijn deze hogere culturele instituties een uitvloeisel van de elementen die ik reeds besproken heb: *Culturele instituties ontwikkelen zich historisch en komen voort uit de behoefte aan een verdere verdieping en ontwikkeling van het gevormde sociale kapitaal.*

Om tot verdieping van sociaal kapitaal te komen dienen hogere “culturele” instituties ontwikkeld te worden die het functioneren van de sociale netwerken vereenvoudigen en versterken. Ik doel hierbij op gedragsnormen en uitingen of signalen die een individu makkelijk doet identificeren met een bepaalde sociale groepering of netwerk. Hieronder vallen bijvoorbeeld tafelmanieren, kledingsstijl of mode en taal. Het moge duidelijk zijn dat het modelleren van cultuur als een signaal in een speltheoretisch model een voor de hand liggende oplossing is. Dit is inderdaad wat Okuno-Fujiwara (2002) en Mailath en Postlewaite (2002) voor ogen staan.

Bowles (1998) beschrijft de interactie tussen human capital, in het bijzonder preferenties, en culturele instituties. Deze culturele instituties worden geïnternaliseerd in het gedrag van de economische agent en aldus verankerd in haar preferenties. Ik

²⁰Voor een uitgebreid overzicht van verschillende economische experimenten verwijs ik naar Kagel en Roth (1995). In dit overzichtswerk behandelt Roth verschillende onderhandelingsexperimenten uit de literatuur.

plaats hierbij de kanttekening dat men hierbij moet oppassen dat de relatie tussen culturele instituties en preferenties geen tautologie wordt. Volledig geïnternaliseerde instituties kunnen namelijk niet meer geïdentificeerd worden en verliezen aldus een verklarend karakter in de daaruitvolgende theorie.

Tenslotte stel ik de vraag hoe culturele instituties verklaard kunnen worden. Bowles (1998), Okuno-Fujiwara (2002) en Mailath en Postlewaite (2002) stellen dat een evolutionaire benadering noodzakelijk is. Deze mening is eveneens gedeeld door Kirman (1997), Loasby (2000) en diverse andere economen. Op dit moment is het theoretisch kader voor zo'n benadering nog niet voldoende ontwikkeld. De evolutionaire speltheorie is tot op heden vrij beperkt in haar toepasbaarheid en er moet nog zeer veel werk verzet worden om tot een bruikbaar kader te komen.

Ten tweede is het nog niet duidelijk hoe evolutionaire theorieën betreffende de vorming van sociaal kapitaal en culturele instituties zich zullen ontwikkelen. De zeer eenvoudige structuren zoals voorgesteld in Okuno-Fujiwara (2002) en Mailath en Postlewaite (2002) zijn onvoldoende. Eventuele verklaringen zullen het nivo van zulke simpele signaalmodellen moeten overstijgen.

Een eerste aanzet tot de ontwikkeling van een evolutionaire theorie voor de vorming van netwerken is gegeven in Jackson en Watts (2002). Zij beschrijven hoe evolutionaire processen tot bepaalde netwerken leidt. Hierbij is het evolutionaire proces op de relaties gelegd in plaats van de individuele agenten. Het is duidelijk dat verdere ontwikkeling van zulke theorieën noodzakelijk is om tot diepere inzichten te komen.

4 Enige conclusies

In deze bijdrage heb ik geprobeerd om te schetsen welke factoren belangrijk zijn in de ontwikkeling van een dieper begrip van het functioneren van sociaal-economische instituties. In het bovenstaande identificeer ik een drietal gerelateerde sociale structuren: netwerken welke de basis van het sociaal kapitaal vormen, sociaal rationele gedragsnormen in de vorming van deze netwerken en, tenslotte, culturele instituties die de vorming van sociaal kapitaal vereenvoudigen en versterken.

Om tot verdere ontwikkeling van een institutioneel economische theorie te komen, moet de economische wetenschap het zichzelf opgelegde juk van het neo-Walrasiaanse onderzoeksprogramma van zich afwerpen. Een open visie op de onderliggende neo-Walrasiaanse assumpties en methoden is daarvoor noodzakelijk. Een nieuwe visie op een markteconomie als een complexe structuur van gerelateerde netwerken en economische instituties kan alleen onderbouwd worden indien de onderzoeksmethoden daarmee meegaan. Uit het voorgaande volgt dat de traditionele econometrie een nadeel heeft betreffende de empirische identificatie en validatie van deze instituties.

Een tweetal innovatieve methoden verdienen nadere beschouwing: experimentatie en simulatie. Experimentele economie is reeds ver ontwikkeld en dient verder ontwikkeld te worden. Dit moet in nauwe relatie met de theorie gebeuren. Vooruitgang van de economische wetenschap is alleen mogelijk indien empirische validatie van theoretische hypothesen door experimenten kan worden geboden. Anderzijds zijn experimenten welke ingegeven zijn door puur empirische vragen minder vrucht-

baar, omdat ze geen test vormen van theoretisch geformuleerde stellingen.

Simulatie is een techniek die in iedere andere wetenschap met een wiskundige basis in gebruik is. Het is een zeer belangrijke techniek in de natuurkunde en wordt effectief gebruikt in de ontwikkeling van de statistische netwerk theorieën zoals beschreven in Watts (2003), Albert en Barabási (2002) en Dorogovtsev en Mendes (2002). Het is daarom onbegrijpelijk dat deze techniek zo radicaal vermeden wordt in de economische wetenschap. Er is bijna geen onderzoek gedaan waarin complexe economische structuren verklaard worden door de simulatie van eenvoudige speltheoretische modellen. Zelfs in de evolutionaire speltheorie is dit niet het geval.

Het is duidelijk dat de besproken structuren en instituties niet begrepen kunnen worden zonder het simuleren van dynamische modellen. Daartoe moet de publicatiecultuur in de economische wetenschap behoorlijk aangepast worden. De belangrijke journals moeten open staan voor bijdragen welke gebaseerd zijn op deze innovatieve technieken.

Aangezien ik niet de enige ben die momenteel aan de bel trekt om de tekortkomingen van de economische wetenschap te beschrijven,²¹ lijkt het dat de tijd rijp is voor vernieuwing in de economische wetenschap. Op dit moment verbergt deze zich achter een vloed van zogenaamde toegepaste modellen en empirische studies. Al deze toegepaste modellen en studies zijn zeer nauw en dragen vrijwel niets bij tot de oplossing van de vraagstukken die ik in dit artikel beschreven heb. De doelstelling van veel economische opleidingsprogramma's lijkt beperkt te zijn tot het opleiden van praktisch georiënteerde toegepaste economen die het dogma van de volledig efficiënte markt verder kunnen propageren. Ik hoop dat ik duidelijk gemaakt heb dat dit een onproductieve benadering is. Indien de economische wetenschap wederom relevant wil worden dan moet het weer terugkeren tot haar kernvraagstukken en open staan voor nieuwe benaderingen om deze te bestuderen.

Referenties

- Albert, R., en A-L. Barabási (2002), "Statistical Mechanics of Complex Networks", *Review of Modern Physics*, 74, 47–97.
- Bowles, S. (1998), "Endogenous Preferences: The Cultural Consequences of Markets and Other Economic Institutions", *Journal of Economic Literature*, 36, 75–111.
- Brink, R. van den, and R.P. Gilles (2002), "Explicit and Latent Authority in Hierarchical Organizations", Mimeo, Department of Economics, Virginia Tech, Blacksburg.
- Burt, R.S. (1992), *Structural Holes: The Social Structure of Competition*, Harvard University Press, Cambridge.
- Buskens, V. (1999), *Social Networks and Trust*, Proefschrift, ICS, Utrecht.

²¹Okuno-Fujiwara (2002) en Davis (2002) voeren een vrijwel soortgelijke discussie. Dit zijn echter slechts twee voorbeelden van vele discussies over de status van de economische wetenschap in de literatuur.

- Chwe, M.S.-Y. (1994), “Farsighted Coalitional Stability”, *Journal of Economic Theory*, 63, 299–325.
- Coase, R.H. (1937), “The Nature of the Firm”, *Economica*, 4, 386–405.
- Coleman, J.S. (1988), “Social Capital in the Creation of Human Capital”, *American Journal of Sociology*, 94, S95–S120.
- Davis, J.B. (2002), “The Emperor’s Clothes”, *Journal of the History of Economic Thought*, 24, 141–154.
- Debreu, G. (1959), *Theory of Value*, J. Wiley & Sons, New York.
- Dorogovtsev, S.N., en J.F.F. Mendes (2002), “Evolution of Networks”, *Advances in Physics*, 51, 1079–1187.
- Epstein, S.A. (1991), *Wage Labor and Guilds in Medieval Europe*, University of North Carolina Press, Chapel Hill.
- Epstein, S.R. (1998), “Craft Guilds, Apprenticeship, and Technological Change in Preindustrial Europe”, *Journal of Economic History*, 58, 684–713.
- Gilles, R.P. (1990), *Core and Equilibria of Socially Structured Economies: The Modelling of Social Constraints in Economic Behaviour*, Proefschrift, Katholieke Universiteit Brabant, Tilburg.
- Gilles, R.P. (1996), *Economic Exchange and Social Organization: The Edgeworthian Foundations of General Equilibrium Theory*, Kluwer Academic Publishers, Boston.
- Gilles, R.P., en D. Diamantaras (2003), “To Trade or Not to Trade: Economies with a Variable Number of Tradeables”, *International Economic Review*, forthcoming.
- Gilles, R.P., D. Diamantaras en P.H.M. Ruys (2000), “Optimal Design of Trade Institutions”, *Working Paper E00-01*, Department of Economics, VPI&SU, Blacksburg.
- Gilles, R.P., en P.H.M. Ruys (1994), “Inherent Imperfection of Economic Organizations”, in R.P. Gilles en P.H.M. Ruys (eds.), *Imperfections and Behavior in Economic Organizations*, Kluwer Academic Publishers, Boston, 1–14.
- Gilles, R.P., en S. Sarangi (2003), “The Role of Trust in Costly Network Formation”, Working Paper, Department of Economics, Virginia Tech, Blacksburg.
- Granovetter, M.S. (1973), “The Strength of Weak Ties”, *American Journal of Sociology*, 78, 1360–1380.
- Granovetter, M.S. (1985), “Economic Action and Social Structure: The Problem of Embeddedness”, *American Journal of Sociology*, 91, 481–510.

- Greif, A. (1994), “Cultural Beliefs and Organization of Society: A Historical and Theoretical Reflection on Collectivist and Individualist Societies”, *Journal of Political Economy*, 102, 912–950.
- Greif, A. (2000), “The Fundamental Problem of Exchange: A Research Agenda in Historical Institutional Analysis”, *European Review of Economic History*, 4, 251–284.
- Greif, A., P. Milgrom en B.R. Weingast (1994), “Coordination, Commitment, and Enforcement: The Case of the Merchant Guild”, *Journal of Political Economy*, 102, 745–776.
- Grossman, S.J., en O.D. Hart (1986), “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 94, 691–719.
- Güth, W., R. Schmittberger en B. Schwarze (1982), “An Experimental Analysis of Ultimatum Bargaining”, *Journal of Economic Behavior and Organization*, 3, 367–388.
- Hart, O., en J. Moore (1990), “Property Rights and the Nature of the Firm,” *Journal of Political Economy*, 98, 1119–1158.
- Hendricks, K., M. Piccione en G. Tan (1999), “Equilibria in Networks”, *Econometrica*, 67, 1407–1434.
- Hickson, C.R., en E.A. Thompson (1991), “A New Theory of Guilds and European Economic Development”, *Explorations in Economic History*, 28, 127–168.
- Israel, J.I. (1995), *The Dutch Republic: Its Rise, Greatness, and Fall 1477-1806*, Oxford University Press, Oxford.
- Jackson, M.O. (2001), “The Stability and Efficiency of Economic and Social Networks,” forthcoming in M. Sertel (ed.), *Advances in Economic Design*, Springer Verlag, Heidelberg.
- Jackson, M.O., en A. Watts (2002), “The Evolution of Social and Economic Networks,” Mimeo, California Institute of Technology, Pasadena, CA 91124, USA.
- Jackson, M.O., en A. Wolinsky (1996), “A Strategic Model of Social and Economic Networks,” *Journal of Economic Theory*, 71, 44–74.
- Johnson, C., en R.P. Gilles (2000), “Spatial Social Networks,” *Review of Economic Design*, 5, 273–300.
- Kagel, J.H., en A.E. Roth (1995), *The Handbook of Experimental Economics*, Princeton University Press, Princeton.
- Kirman, A. (1997), “The Economy as an Evolving Network”, *Journal of Evolutionary Economics*, 7, 339–353.

- Konishi, H. (2000), "Formation of Hub Cities: Transportation Cost Advantage and Population Agglomeration", *Journal of Urban Economics*, 48, 1-28.
- Laan, G. van der, P.H.M. Ruys, en D. Talman (2000), "Optimal Provision of Infrastructure in a Public-Private Partnership Relation", Center for Economic Research, Katholieke Universiteit Brabant, Tilburg.
- Lakatos, I. (1978), *The Methodology of Scientific Research Programmes*, Cambridge University Press, Cambridge.
- Loasby, B.J. (2000), "Market Institutions and Economic Evolution," *Journal of Evolutionary Economics*, 10, 297-309.
- Mailath, G.J., en A. Postlewaite (2002), "Social Assets", Mimeo, Department of Economics, University of Pennsylvania, Philadelphia.
- Maskin, E., en J. Tirole (1999), "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies*, 66, 83-114.
- Ménard, C. (1995), "Markets as Institutions versus Organizations as Markets? Disentangling some Fundamental Concepts," *Journal of Economic Behavior and Organization*, 28, 161-182.
- Newman, M.E.J., S.H. Strogatz en D.J. Watts (2001), "Random Graphs with Arbitrary Degree Distributions and their Applications", cond-mat/0007235 v2.
- North, D.C. (1990), *Institutions, Institutional Change and Economic Performance*, Cambridge University Press, Cambridge.
- Okuno-Fujiwara, M. (2002), "Social Relations and Endogenous Culture", *Japanese Economic Review*, 53, 1-24.
- Prak, M. (2002), *Gouden Eeuw: Het raadsel van de Republiek*, Uitgeverij SUN, Nijmegen.
- Rajan, R.G., en L. Zingales (2001) "The Firm as a Dedicated Hierarchy", *Quarterly Journal of Economics*, 116, 805-851.
- Robison, L.J., A.A. Schmid and M.E. Siles (2002), "Is Social Capital Really Capital?", *Review of Social Economy*, 60, 1-21.
- Romer, P. (1994), "New Goods, Old Theory, and the Welfare Costs of Trade Restrictions", *Journal of Development Economics*, 43, 5-38.
- Ruys, P.H.M. (2002), "A General Equilibrium of Managed Services", in C.H. Hommes, R. Ramer and C.A. Withagen (editors), *Equilibrium, Markets and Dynamics: Essays in Honour of Claus Weddepohl*, Springer Verlag, Berlin.
- Ruys, P.H.M., en R. van den Brink (1999), "Positional Abilities and Rents on Equilibrium Wages and Profits", in P.J.J. Herings, G. van der Laan and A.J.J. Talman (eds.), *The Theory of Markets*, North Holland, 261-269.

- Ruys, P.H.M., R. van den Brink, en R. Semenov (2001), “Values and governance systems”, Chapter 27 in: C. Ménard (ed.), *Institutions, Contracts and Organizations*, Edward Elgar Publishing Co., 451–485.
- Sargent, T.J., en F.R. Velde (2002), *The Big Problem of Small Change*, Princeton University Press, Princeton.
- Simon, H.A. (1951), “A Formal Theory of the Employment Relationship,” *Econometrica*, 19, 293–305.
- Simon, H.A. (1991), “Markets and Organization”, *Journal of Economic Perspectives*, 5, 25–44.
- Smith, A. (1776), *An Inquiry into the Nature and Causes of the Wealth of Nations*, herdruk: University of Chicago Press, Chicago, 1976.
- Spufford, P. (1988), *Money and its Use in Medieval Europe*, Cambridge University Press, Cambridge.
- Vries, J. de, en A. van der Woude (1997), *The First Modern Economy: Success, Failure, and Perseverance of the Dutch Economy, 1500–1815*, Cambridge University Press, Cambridge.
- Watts, D.J. (2003), *Six Degrees: The Science of a Connected Age*, W.W. Norton & Company, New York.
- Weintraub, E.R. (1985), *General Equilibrium Analysis: Studies in Appraisal*, Cambridge University Press, Cambridge.

HIËRARCHISCH GESTRUKTUREERDE MARKTEN

Willy Spanjers*

University of Birmingham

Het eerste college dat Pieter in het voorjaar van 1983 gaf voor de cursus “Oriëntatie Econometrie”, staat me nog duidelijk voor de geest. Het ging over preferentie relaties en na anderhalf uur bleven wij, studenten, achter in een staat van totale verwarring. Enkele avonden werden gevuld met discussies, voordat we dachten begrepen te hebben waar het om ging.

Het is moeilijk te zeggen waarom ik me na twintig jaar nog zo goed dit college kan herinneren. Misschien omdat het de eerste keer was dat ik Pieter als leraar heb beleefd. Een andere verklaring is dat deze ervaring zich nog vaak zou herhalen, na de wekelijkse voortgangsbesprekingen met Pieter was ik doorgaans niet minder verward.

Tegenwoordig geldt het niet als wenselijk, dat studenten aan het eind van een college het overzicht over de stof verloren hebben. Met het beschreven college heeft Pieter mij en anderen er echter toe gebracht, bewust over het vertelde na te denken. Het was ons eerste contact met de wetenschappelijke houding. 's Avonds discussieren over dingen die overdag in colleges niet duidelijk waren, werd tot een goede gewoonte onder vrienden, waarbij in de loop der jaren veel droge stof tot leven is gekomen.

Toch is Pieter's vermogen mensen te inspireren niet het meest kenmerkende voor hoe ik hem heb leren kennen. Het meest kenmerkende is, dat hij me steeds heeft aangemoedigd mijn eigen weg te gaan en me daarbij altijd en onvoorwaardelijk ondersteund heeft, vaak tegen beter weten in. Of het er nu om ging een vrij doctoraal te organiseren (waar uiteindelijk niets van terecht kwam), om mijn pogingen een half jaar in de DDR te studeren, om het thema van mijn proefschrift, of om het tijdstip van mijn promotie. Ik vrees dat ik Pieter menig grijs haar heb gekost.

Een van zijn aansporingen was, dat ik als AIO het strakke koord op moest, en hij voor het vangnet zou zorgen als ik naar beneden zou tuimelen. Hoe treffend de vergelijking van een universiteit met een circus is, was me toen nog niet bewust. Belangrijke delen van het universitaire leven bestaan uit koorddansen, kunst- en vliegwerk, leeuwen temmen en soms de Paljas zijn.

Een ander advies was, “als je jong bent moet je je met wiskunde bezig houden, als je ouder wordt met economie”. Dit advies heb ik geprobeerd te volgen. Hoe

* *Adres:* Department of Economics, The University of Birmingham, Edgbaston, Birmingham B15 2TT, United Kingdom. *Email:* w.spanjers.1@bham.ac.uk

lang de reis van wiskundige economie naar economie is heb ik me echter in mijn stoutste dromen niet kunnen voorstellen.

Met bestuurlijke dingen was Pieter altijd in de weer. Of het er nu om ging de studierichting Econometrie voor Tilburg te behouden, het Erasmus programma op te zetten, het Tias op te bouwen, of zich voor het CentER in te zetten. Met zijn tomeloze inzet heeft Pieter er in belangrijke mate aan bijgedragen dat in Tilburg een omgeving ontstond waarin ik, evenals vele anderen, kon gedijen.

Ik beschouw het als een bijzonder voorrecht bij Pieter te hebben mogen promoveren. Voor de beschrijving van Pieter's rol is het nederlandse "promotor" even misplaatst als het anglosaksische "supervisor". Het juiste woord is het duitse "Doktorvater". Bedankt Pieter.

1 Inleiding

De leden van een maatschappij werken op vele verschillende manieren met en tegen elkaar. Een belangrijke deel hiervan is economische interactie. De economische interactie tussen de verschillende deelnemers vindt doorgaans niet plaats op het niveau van de gehele maatschappij, maar binnen organisatorische sub-eenheden. Enkele voorbeelden zijn markten, bedrijven en de overheid. Deze verschillende soorten organisatorische eenheden hebben ieder hun eigen sterktes en zwaktes.

1. **Markten** zijn geschikt voor relatief vrijblijvende interactie. In principe kunnen in een markt de verschillende handelspartners willekeurig uitgewisseld worden. Het is niet eens nodig de identiteit van de handelspartner te kennen. In een dergelijke omgeving kan het riskant zijn lange termijn investeringen door te voeren, die op een specifieke handelspartner, bijvoorbeeld een afnemer, gericht zijn. Zou deze afnemer immers besluiten een andere leverancier te zoeken, dan zou de investering sterk aan waarde verliezen. Dit geeft de afnemer de gelegenheid de leverancier onder druk te zetten en hiermee zijn economisch voordeel te doen. Daarom vinden in markten minder op specifieke handelspartners gerichte investeringen plaats dan wenselijk is.
2. **Bedrijven** zijn organisaties, die bijzonder geschikt zijn voor op de lange termijn gerichte samenwerking. Binnen een bedrijf is niet zo makkelijk een andere partner voor samenwerking te zoeken als binnen een markt. Dit maakt het minder riskant lange termijn investeringen door te voeren die op specifieke afdelingen binnen het bedrijf gericht zijn. Het risico dat de afdeling zich een andere partner zoekt, en de investering hierdoor aan waarde verliest, is kleiner dan in een markt.
3. De kracht van de **overheid** ligt daar in, dat ze een bepaalde mate van eenzijdige dwang op haar burgers kan uit oefenen. Op deze wijze kan ze samenwerking tussen burgers afdwingen en in stand houden. Hiervan wordt in het bijzonder dan gebruik gemaakt, wanneer het algemeen belang in strijd is met het eigenbelang van de direkt betroffenen. Een gevolg hiervan is, dat enkel de overheid in staat is om (direkt of indirekt) een passende verzorging met publieke en semi-publieke goederen te bereiken.

Niet het toeval bepaalt of een specifiek soort van economische interactie door markten, bedrijven of overheidssturing plaats vindt. Het zijn de specifieke eigenschappen van de interactie, die bepalen welk type organisatie het geschiktste is. Idealiter wordt voor iedere soort van economische interactie de geschiktste organisatievorm gekozen.

1.1 Markten

In deze bijdrage wordt de aandacht gericht op organisaties, die wezenlijke kenmerken van markten bezitten, maar waarin handelspartners niet willekeurig uitgewisseld kunnen worden. Zo worden situaties geanalyseerd, waarin zich stabiele handelsrelaties hebben gevormd. Anders dan bij bedrijfsorganisaties wordt echter aangenomen, dat zelfs binnen een stabiele handelsrelatie geen der handelspartners bindende opdrachten kan geven.

De gebruikelijke beschrijving van een markt ziet deze als een plaats waar vraag en aanbod van een identiek goed elkaar treffen. Het goed wordt door anonieme marktdeelnemers tegen eenzelfde prijs gekocht en verkocht. Bij dit begrip van een markt speelt het geen enkele rol, welke marktdeelnemer bij welke andere marktdeelnemers koopt of verkoopt. In principe kan iedereen met willekeurige en willekeurig veel verschillende anderen handelen. Of en hoe dit gebeurt heeft geen invloed op de prijs en de uiteindelijk resulterende verdeling van het goed.

Het is evident, dat dit beeld sterk afwijkt van de gang van zaken op alledaagse markten. De verschillen tussen alledaagse markten en hun abstracte weergave is zelfs zo groot, dat twijfels aan de bruikbaarheid van het theoretisch/abstracte concept op hun plaats zijn. De belangrijkste in het oog springende verschillen betreffen:

1. de aanname, dat op een markt identieke goederen worden gekocht en verkocht,
2. het voorhanden zijn van een rechtssysteem dat waarborgt dat
 - (a) de gehandelde goederen aan de voor deze markt geldende specificaties voldoen, dat
 - (b) gekochte goederen geleverd worden en dat
 - (c) geleverde goederen betaald worden.
3. het tot stand komen van een enkele prijs waartegen ieder marktdeelnemer naar believen kan kopen en verkopen,
4. het tot stand komen van een evenwichtsprijs die zodanig is, dat de vraag en het aanbod op de markt even groot zijn.

Voor iedere van deze aannames bestaat een uitgebreide wetenschappelijke literatuur, die analyseert hoe afwijkingen hiervan het functioneren van markten beïnvloeden.

Zo houdt de *nieuwe institutionele economie* zich bezig met de vraag, wat kan gebeuren als de tweede aanname afgezwakt wordt. Hierbij spelen de uitwerkingen van asymmetrieën in beschikbare informatie een bijzondere rol. Er wordt gevraagd, hoe door een adequate specificatie van handelscontracten de kwaliteit en andere karakteristieken van de geleverde goederen en diensten positief kunnen

worden beïnvloed. Hierbij staat de wens in het middelpunt, de neiging van beter geïnformeerde marktdeelnemers tot “opportunistisch gedrag” (d.w.z. niet bewijsbaar bedrog) te verminderen. Door middel van opportunistisch gedrag zijn deze immers in staat voordeel te slaan uit hun informatievoorsprong.

Het moge duidelijk zijn, dat de gevolgen van dergelijke handelscontracten sterk afhangen van de kennisstand, de doelstellingen en de intellectuele vaardigheden van de betrokken marktdeelnemers. Onder zulke omstandigheden kan er nauwelijks van uitgegaan worden, dat de interactie in de markt anoniem is. Het is wenselijk voor iedere handelsrelatie tussen twee marktdeelnemers een maatgepast contract te ontwerpen en af te sluiten. Hiervoor is het van belang, zowel de specifieke eigenschappen van de handelspartners, als ook de stand van hun informatie te kennen. Enkel dan kan een contract ontworpen worden, dat de mogelijkheden tot opportunistisch gedrag effectief inperkt.

In dergelijke “markten” kunnen verschillende handelscontracten naast elkaar worden aangetroffen. Deze contracten kunnen verschillen met betrekking tot prijs, hoeveelheid, kwaliteit, garantie en andere condities. Er is geen enkele reden om aan te nemen, dat alle handelsrelaties eenzelfde standaardcontract ten gronde ligt. En al helemaal niet, dat dit standaardcontract van het type is “de ene handelspartner kan voor de geldende prijs bij de andere willekeurige hoeveelheden kopen en verkopen”.

De situatie in een markt wordt nog onoverzichtelijker, als voor het tot stand komen van bepaalde transacties meerdere marktdeelnemers nodig zijn. In een dergelijke situatie kan het zinvol zijn een contract af te sluiten, dat meer dan twee partijen bindt. Hierbij moet afgesproken worden, wie waarvoor verantwoordelijk is. Daarenboven moet vastgelegd worden, wat gebeurd als dingen mislopen en niet één maar meerdere deelnemers hun verplichtingen niet zijn nagekomen. In concrete gevallen kan een dergelijk meerpartijencontract veel van de interne organisatiestructuur van een bedrijf weg hebben.¹

In principe is er ook een andere mogelijkheid transacties te organiseren waarvoor de samenwerking van meerdere partijen nodig is. Een aantal maatgepaste bilaterale contracten tussen de partijen kan eventueel ook de noodzakelijke mate van samenwerking tot stand brengen. In zulke gevallen lijkt de organisatiestructuur meer op die van een markt. Niettemin kan het wegvallen van een schijnbaar op zichzelf staand contract verrassend grote gevolgen hebben doordat een domino-effect in werking treedt.

Maar zelfs als dergelijke domino-effecten uitgesloten zijn, leidt dit niet automatisch tot een stabiele structuur van handelsrelaties. In het algemeen zal het voor sommige marktdeelnemers zinvol lijken informatie te verzamelen over de condities in andere contracten. Het is immers denkbaar, dat de condities in de contracten met hun vaste handelsrelaties niet zo gunstig zijn als die elders in de markt.

De idee, dat het zinvol kan zijn niet het eerste het beste aanbod aan te nemen, is de basis van de *zoektheorie*. In deze theorie wordt het tegendeel van vaste

¹Het doelgericht ontwerpen van instituties die meerdere personen omvatten valt onder het gebied van *mechanism design*. Hierbij worden regels voor interactie ontworpen, die tot doel hebben bepaalde voorgegeven resultaten te bereiken. De belangrijkste zwakte van dit soort modellen is dat de resultaten nog niet erg robuust zijn. Bij toepassingen kunnen onaangename verrassingen optreden.

handelsrelaties aangenomen. De marktdeelnemers proberen telkens weer of ze niet elders betere condities kunnen krijgen. Hoewel dergelijk zoekgedrag zinvol is en een belangrijke rol in het functioneren van markten speelt, wordt er in deze bijdrage van geabstraheerd.

De belangrijkste reden hiervoor is, dat zelfs in het tijdperk van het internet veel handel met vaste partners plaatsvindt. Vaak zorgt het vinden van een betere offerte er niet toe, dat een andere handelspartner gekozen wordt. Meestal worden in onderhandelingen met bestaande handelspartners enkel geldende condities verbeterd. Daarom lijkt het zinvol ook verder het oogmerk te richten op markten met een vaste structuur van (bilateral) handelsrelaties.

1.2 Hiërarchisch gestructureerde markten

Het model van successieve monopolies is een wat ouder model, dat als een model met vaste handelsrelaties geïnterpreteerd kan worden. Hierin komen drie verschillende groepen van marktdeelnemers voor. Een monopolist, tussenhandelaars en consumenten. De monopolist kan als enige het betreffende goed produceren. Bovendien legt hij de groothandelsprijs vast waarvoor de tussenhandelaars bij hem kunnen inkopen.

De tussenhandelaars beschouwen deze groothandelsprijs als onveranderbaar en proberen, door de keuze van de detailhandelsprijs voor de consumenten, een zo groot mogelijke winst te behalen. Hierbij mogen ze niet uit het oog verliezen, dat een hogere detailhandelsprijs weliswaar de winstmarges verhoogt, maar tegelijkertijd de afzet kan verminderen.

Het beslissingsprobleem van de consumenten is relatief eenvoudig. Gegeven de door de tussenhandelaars vastgelegde detailhandelsprijs beslissen ze hoeveel ze van het betreffende goed willen kopen.

Als we aannemen dat er naast de monopolist precies één tussenhandelaar en precies één consument is, dan hebben we een eenvoudig voorbeeld van een markt met bilaterale handelsrelaties. De kunst voor de monopolist is het hierbij, voor iedere mogelijke groothandelsprijs correct in te schatten, hoeveel hij uiteindelijk aan de tussenhandelaar kan verkopen. Deze hoeveelheid hangt mede af van de hoeveelheid die de consument bereid is te kopen voor de detailhandelsprijs die de tussenhandelaar vastlegt nadat hij de groothandelsprijs kent.

De grondgedachte achter hiërarchisch gestructureerde markten is, het model van successieve monopolies uit te breiden. Het aantal marktdeelnemers is niet tot drie beperkt maar kan willekeurig groot zijn. In plaats van verticale lijnverbindingen is een willekeurige structuur van handelsrelaties mogelijk. En in plaats van de contractvorm: “de ene zet de prijs en de andere beslist hoeveel hij koopt” zijn willekeurige bilaterale standaardcontracten toegelaten. Een zulk bilateraal standaardcontract wordt in het vervolg de institutionele karakteristiek van een handelsrelatie genoemd.

De enige aanname van het model van successieve monopolies die niet zo makkelijk kan worden afgezwakt betreft de (impliciete) informatie die de marktdeelnemers hebben over de mogelijke gevolgen van hun beslissingen.² De tussenhandelaar

²Zie Spanjers (1992, Chapter 5).

hoeft “slechts” te weten, hoe het koopgedrag van de consument op de door hem vastgelegde detailhandelsprijs reageert. De monopolist moet niet alleen implicite informatie over dit koopgedrag hebben. Hij moet tevens kunnen inschatten, hoe de groothandelsprijs de door de tussenhandelaar gekozen detailhandelsprijs beïnvloedt.

Op een zelfde manier moet in een hiërarchisch gestructureerde markt aangenomen worden, dat er een strikte informatiehiërarchie bestaat. Een marktdeelnemer, die hoger in deze hiërarchie staat beschikt over meer (implicite) informatie over de reacties van de anderen, dan iemand die onder hem staat.

De rest van deze bijdrage is als volgt georganiseerd. Paragraaf 2 beschrijft het model van een hiërarchisch gestructureerde markt. De Existentie Stelling in Paragraaf 3 geeft de voorwaarden die de existentie van een evenwicht in een hiërarchisch gestructureerde markt garanderen. De Equivalentie Stelling maakt duidelijk, dat monopolistische markten, en zodoende markten met volledige mededinging, als bijzondere gevallen van het model opgevat kunnen worden. Paragraaf 4 trekt conclusies.

2 Het Model

2.1 Beschrijving

Net als in het model van successieve monopolies, beschrijven we de markt voor een enkel goed.³ Verder nemen we aan, dat in deze markt enkel consumenten actief zijn. Productie wordt buiten beschouwing gelaten.⁴

Het basismodel van een hiërarchisch gestructureerde markt beschrijft zowel de consumenten als ook de structuur van handelsrelaties. Hiermee wijkt de beschrijving op een belangrijke wijze af die van een gebruikelijke markt. De beschrijving betreft:

1. de consumenten en hun individuele karakteristieken,
2. de structuur van handelsrelaties tussen deze consumenten en de institutionele karakteristieken van de handelsrelaties, en
3. een informatiehiërarchie over de consumenten die vast legt, welke consumenten weten hoe welke andere consumenten op hun acties reageren.

Laten we beginnen met de beschrijving van de informatiehiërarchie over de consumenten. Deze legt een strikte volgorde over de consumenten vast. De eerste consument in de hiërarchie weet hoe de rest van de markt op zijn acties zal reageren. De tweede consument weet hoe degenen die in de hiërarchie onder hem staan op zijn acties reageren, gegeven de acties van de eerste consument, enzovoorts. Naar de consumenten in de markt wordt verwezen met hun rang in de hiërarchie. De verzameling consumenten is zodoende $I := \{1, 2, \dots, n\}$, waarbij n het aantal consumenten is. In het voorbeeld van successieve monopolies staat de monopolist

³Deze aanname is niet noodzakelijk. In Spanjers (1997) wordt de analyse uitgevoerd voor een systeem van markten voor een willekeurig aantal goederen.

⁴Het is mogelijk ook productie in het model op te nemen. Zie Spanjers (1997, blz. 19).

bovenaan in de informatiehiërarchie. De tussenhandelaar staat op plaats twee en de consument is laatste.

De individuele karakteristieken van consument i worden beschreven door zijn nutsfunctie $U_i : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ en zijn beginvoorraad aan het goed, \bar{x}_i , en geld, \bar{m}_i . De nutsfunctie legt vast, welke waardering hij voor een combinatie van x eenheden van het goed en m eenheden geld heeft. Een hogere waardering betekent dat de betreffende combinatie geprefereerd wordt. De beginvoorraad beschrijft over welke hoeveelheid van het goed en over hoeveel geld de consument beschikt voordat handel plaats vindt.

Als er tussen de consumenten i en j een handelsrelatie bestaat, wordt deze met (i, j) aangegeven indien $i < j$ en met (j, i) indien $i > j$. De consument die hoger in de hiërarchie staat, de leider, wordt dus als eerst genoemd. De consument die lager in de hiërarchie staat, de volger, wordt als laatste genoemd. De verzameling leiders van consument i wordt met $\mathcal{L}(i)$ aangegeven, de verzameling volgers met $\mathcal{F}(i)$. De verzameling van alle handelsrelaties in de markt is R . In het voorbeeld van successieve monopolies bevat de verzameling van handelsrelaties twee elementen, namelijk (*monopolist, tussenhandelaar*) en (*tussenhandelaar, consument*).

Enkel de wetenschap welke handelsrelaties tussen de consumenten in de markt voorhanden zijn, maakt het nog niet mogelijk vast te stellen hoeveel over welke handelsrelatie gehandeld zal worden. De reden hiervoor is dat nog niet bepaald is welke regels voor de interactie binnen een handelsrelatie (i, j) gelden. Deze regels worden beschreven door de institutionele karakteristiek $\mathcal{T}_{(i,j)}$ van de handelsrelatie $(i, j) \in R$.

Hierbij kiest de leider, i , een signaal s_{ij} en bepaalt de volger, j , hoeveel van het goed over deze relatie gehandeld wordt, t_{ij} , en tegen welke betaling, b_{ij} . Deze keuze van de volger is echter niet willekeurig. Hij kan enkel combinaties (t_{ij}, b_{ij}) kiezen die in de verzameling $\mathcal{T}_{(i,j)}(s_{ij})$ liggen. De institutionele karakteristiek is zodoende een correspondentie $\mathcal{T}_{(i,j)} : S_{ij} \rightrightarrows \mathbb{R}^2$, waarbij S_{ij} de verzameling van beschikbare signalen beschrijft.

De institutionele karakteristiek legt vast hoe de leider door zijn keuze van het signaal s de keuzevrijheid van de volger kan beperken met betrekking tot hun handelsrelatie. In het voorbeeld van successieve monopolies is de institutionele karakteristiek de volgende. De leider kiest een prijs p (het signaal), waarna de volger een willekeurige combinatie van handel en betaling (t, b) kan kiezen, zolang de betaling overeenstemt met de kosten van de gehandelde hoeveelheid, d.w.z. $b = p \cdot t$. Met andere woorden, de volger kan een willekeurige combinatie (t, b) kiezen uit de verzameling

$$\mathcal{T}^{mon}(p) := \{ (t, b) \in \mathbb{R}^2 \mid b = p \cdot t \}.$$

Indien $t > 0$ koopt de volger bij de leider en betaalt hiervoor de aangekondigde prijs. Is $t < 0$, dan verkoopt de volger aan de leider en ontvangt hiervoor de aangekondigde prijs. De karakteristiek wordt beschreven door de correspondentie $\mathcal{T}^{mon} : \mathbb{R}_+ \rightrightarrows \mathbb{R}^2$.

Samengevat leidt dit tot de volgende definitie van een hiërarchisch gestructureerde markt:

Definitie 1 Een hiërarchisch gestructureerde markt is gegeven door een combinatie $\mathbf{M} := (\{U_i, (\bar{x}_i, \bar{m}_i)\}_{i \in I}, \xi, \{\mathcal{T}_{(i,j)}\}_{(i,j) \in R})$ waarbij:

- (a) I de verzameling van consumenten is. Ieder consument $i \in I$ heeft een nutsfunctie $U_i : \mathbb{R}_+^2 \rightarrow \mathbb{R}$ en beginvoorraden $(\bar{x}_i, \bar{m}_i) \in \mathbb{R}_+^2$.
- (b) ξ een partitie is van de verzameling consumenten in hiërarchische niveaus, waarbij ieder niveau precies een consument bevat.
- (c) R een verzameling van bilaterale handelsrelaties tussen verschillende consumenten is. Iedere handelsrelatie $(i, j) \in R$ heeft een institutionele karakteristiek $\mathcal{T}_{(i,j)} : S_{ij} \rightrightarrows \mathbb{R}^2$.

Een combinatie van acties in \mathbf{M} wordt aangegeven met

$$((x_i, m_i)_{i \in I}, (s_{ij}, t_{ij}, b_{ij})_{(i,j) \in R}) \in \mathbb{R}_+^{2|I|} \times \prod_{(i,j) \in R} (S_{ij} \times \mathbb{R}^2).$$

2.2 Evenwicht

Nu we met Definitie 1 een wiskundig precieze beschrijving van een hiërarchisch gestructureerde markt hebben, kunnen we het volgende probleem aanpakken. Dat betreft de vraag, hoe de consumenten in deze markt zich gedragen. Dit is een probleem, omdat het gedrag van de consumenten niet alleen van hun doelstelling en de omgeving waarin ze handelen afhangt, maar ook van hun beslissingsproces.

Bij een dergelijk proces spelen de intellectuele vaardigheden van de consumenten evenzeer een rol als hun inschattingen over de gevolgen van hun acties. Het eerste probleem wordt “opgelost” door aan te nemen dat consumenten over de intellectuele vaardigheid beschikken hun eigenbelang te herkennen en overeenkomstig te handelen. De nutsfunctie beschrijft de doelstellingen (d.w.z. het eigenbelang) van de consument. Uit een gegeven verzameling bereikbare alternatieven kiest de consument de beste.

Het tweede probleem, het inschatten van de gevolgen van de eigen acties, is moeilijker op te lossen. Een actie kan vele verschillende denkbare reacties van anderen tot gevolg hebben. Van welke van deze reacties moet de consument bij het nemen van zijn beslissing uitgaan? Dit probleem geldt voor ieder consument in de markt. Een mogelijke oplossing is, de individuele consumenten uitvoeriger te beschrijven. Een consument zou dan niet alleen door zijn nutsfunctie en zijn beginvoorraden beschreven worden, maar tevens door zijn inschattingen over de gevolgen van zijn acties.⁵

Deze oplossing is echter om verschillende redenen niet bevredigend. Ten eerste wordt op deze wijze geen herkenbare verbinding tot stand gebracht tussen de inschattingen van de consument en de omgeving waarin hij zijn beslissing neemt. Het lijkt echter zinvol aan te nemen, dat de inschattingen tot stand komen volgens een vastgelegde regel. Als deze regel op verschillende omgevingen wordt toegepast, leidt dit tot verschillende maar niet tot willekeurige inschattingen over de gevolgen van de acties van de consument.⁶

⁵Deze idee wordt o.a. vervolgt in Hahn (1978) en, in een iets andere vorm, in Vind (1983). Zie ook Spanjers (1992, Chapter 2).

⁶Dergelijke regels ter bepaling van de inschattingen voor verschillende omgevingen worden onder andere gebruikt in Vind (1983), Grodal en Vind (1989) en Spanjers (1992, Hoofdstuk 5 en Hoofdstuk 9).

Ten tweede bestaat er geen herkenbaar verband tussen de inschattingen van de consument en de ervaringen die hij eventueel in zijn omgeving kan opdoen. Indien een consument gedurende langere tijd in dezelfde omgeving actief is, kunnen eventuele foutieve inschattingen gecorrigeerd worden.

In de *speltheorie* worden beide problemen tegelijk opgelost door aan te nemen, dat iedere consument de plannen van de andere consumenten kent. Deze plannen, die “strategieën” genoemd worden, leggen voor iedere denkbare beslissingssituatie waarin de consument kan komen vast welke actie hij kiest. Een combinatie van dergelijk plannen is consistent als geen consument zijn plannen wil veranderen, gegeven de plannen van de anderen. In een zulke combinatie heet een Nash-evenwicht.⁷

Het Nash-evenwicht wordt in deze vorm vooral gebruikt voor situaties, waarin alle consumenten hun plannen tegelijk vastleggen en eenmaal gemaakte plannen tijdens het verloop van de interactie niet meer veranderd kunnen worden.

Voor een hiërarchisch gestructureerde markt is deze oplossing niet bijzonder geschikt.⁸ Consument 1 beslist immers als eerste over de signalen die hij zet voor zijn handelsrelaties. Daarna pas beslist consument 2 over de signalen die hij voor zijn volgers zet. Indien er een handelsrelatie tussen 1 en 2 bestaat, beslist consument 2 tevens over de hoeveelheid van het goed die hij met consument 1 wil handelen. Dan beslist consument 3, enzovoorts.

Daarom is het in de context van dit model zinvol, de inschattingen van de consumenten op een andere manier te bepalen, opdat aan beide genoemde bezwaren tegemoet gekomen wordt.

Consument n , die op de laatste plaats in de informatiehiërarchie staat, beslist als laatste over zijn acties. Als hij zijn acties kiest, zijn op al zijn handelsrelaties de signalen door de leiders gezet. Zodoende hoeft hij enkel te beslissen, welke toegestane combinatie van handel en betaling over deze handelsrelaties voor hem de beste is. Inschattingen over de reacties van anderen hoeft hij niet te maken, de anderen hebben hun beslissingen immers al getroffen.

Vervolgens bepalen we de inschattingen van degene, die direkt voor n in de hiërarchie staat, consument $n - 1$. Indien consument $n - 1$ geen handelsrelatie met n heeft, denkt hij dat n precies datgene doet wat in het voorgaande beschreven werd. Deze beslissing is voor consument $n - 1$ echter zonder betekenis.

Indien er een handelsrelatie tussen de consumenten $n - 1$ en n bestaat, kan $n - 1$ vermoeden hoe consument n op een door hem gezet signaal reageert. Voor ieder signaal zal consument n proberen er voor zich het beste van te maken. Consument $n - 1$ zal inschatten dat over hun handelsrelatie de bijbehorende hoeveelheid handel zal plaatsvinden.

Op soortgelijke wijze, door middel van “backward induction”, kunnen de inschattingen van de overige consumenten bepaald worden. Zij $(s_h)_{h < i}$ de combinatie van signalen, die de consumenten die hoger in de hiërarchie staan dan i zetten voor hun volgers. Zij $s_i := (s_{ij})_{j \in \mathcal{F}(i)}$ de combinatie van signalen van consument i voor zijn volgers. Dan worden de inschattingen van consument i voor de combinatie

⁷Zie Nash (1950).

⁸Het hier gebruikte oplossingconcept is verwant met het “deelspel perfecte” evenwicht, welk op zijn beurt weer een verfijning van het Nash-evenwicht is.

$((s_h)_{h<i}, s_i)$ beschreven als $r_i((s_h)_{h<i}, s_i) := (r_{ij}((s_h)_{h<i}, s_i))_{j \in \mathcal{F}(i)}$. De correspondentie r_i beschrijft de gecombineerde directe en indirecte reacties van de volgers van consument i over hun handelsrelaties met i , d.w.z. over de handelsrelaties (i, j) met $j \in \mathcal{F}(i)$.

In het voorbeeld van successieve monopolies werkt deze procedure als volgt. De enige consument denkt niet na over de uitwerkingen van zijn beslissing welke hoeveelheid van het goed hij tegen de detailhandelsprijs koopt bij de tussenhandelaar.

De tussenhandelaar weet voor iedere detailhandelsprijs welke hoeveelheid de consument koopt. Dit komt er op neer, dat hij de vraagfunctie van de consument naar het goed kent. Over de gevolgen van zijn beslissing hoeveel hij voor de vastgelegde groothandelsprijs bij de monopolist koopt, maakt de tussenhandelaar zich geen zorgen.

De monopolist weet voor iedere groothandelsprijs die hij voor de tussenhandelaar zet hoeveel deze bij hem zal kopen. Deze hoeveelheid hangt zowel af van de hoeveelheid die de tussenhandelaar voor eigen gebruik afneemt, als van de hoeveelheid die de consument tegen de detailhandelsprijs bij de tussenhandelaar koopt. Het totale inkomen van de tussenhandelaar bestaat uit zijn aanvankelijk inkomen plus zijn handelswinst. Deze handelswinst kan ontstaan als de tussenhandelaar een detailhandelsprijs kiest die hoger is dan de groothandelsprijs.

Het doel van een model is het, explicite en implicite aannames over de interactie van beslissers te maken, die een abstractie van de werkelijkheid zijn. Een belangrijke eigenschap is daarom, dat het geen logische tegenspraken bevat. Het model moet een combinatie van beslissingen toelaten die consistent is. Een dergelijke combinatie van beslissingen wordt een evenwicht genoemd.

In een hiërarchisch gestructureerde markt is een combinatie van acties van de consumenten een evenwicht indien:

1. iedere consument in zijn eigenbelang handelt en
2. iedere consument in staat is zijn plannen om te zetten.

Een willekeurige consument i handelt in zijn eigenbelang, indien hij zich conform zijn beslissingsprobleem gedraagt ⁹

$$\max_{(x_i, m_i, (t_{ki}, b_{ki})_{k \in \mathcal{L}(i)}, (s_{ij})_{j \in \mathcal{F}(i)}) \in cl C_i((s_h)_{h<i})} U_i(x_i, m_i)$$

⁹Om een wiskundige voetangel te omzeilen wordt in het beslissingsprobleem van consument i de afsluiting van $C_i((s_h)_{h<i})$ als de keuzeverzameling genomen. Een op deze wijze tot stand komend evenwicht is een "Approximate Equilibrium in Very Nice Plays". Zie Spanjers (1997, Appendix B, Section 3).

waarbij $cl C_i((s_h)_{h<i})$ de afsluiting is van de verzameling

$$C_i((s_h)_{h<i}) := \left\{ \begin{array}{l} ((x_i, m_i), (t_{ki}, b_{ki})_{k \in \mathcal{L}(i)}, (s_{ij})_{j \in \mathcal{F}(i)}) \text{ zodanig dat:} \\ (x_i, m_i) \in \mathbb{R}_+^2, (t_{ki}, b_{ki})_{k \in \mathcal{L}(i)} \in \mathbb{R}^{2|\mathcal{L}(i)|}, \\ (s_{ij})_{j \in \mathcal{F}(i)} \in \prod_{j \in \mathcal{F}(i)} S_{ij}, \\ x_i = \bar{x}_i + \sum_{k \in \mathcal{L}(i)} t_{ki} - \sum_{j \in \mathcal{F}(i)} t_{ij}, \text{ en} \\ m_i = \bar{m}_i - \sum_{k \in \mathcal{L}(i)} b_{ki} + \sum_{j \in \mathcal{F}(i)} b_{ij} \\ \text{waarbij geldt dat } (t_{ij}, b_{ij})_{j \in \mathcal{F}(i)} \in r_i((s_h)_{h<i}, s_i) \end{array} \right\}.$$

Dit leidt tot de volgende definitie van een evenwicht.

Definitie 2 Een evenwicht in een hiërarchisch gestructureerde markt \mathbf{M} is een combinatie van acties

$$((x_i, m_i)_{i \in I}, (s_{ij}, t_{ij}, b_{ij})_{(i,j) \in R}) \in \mathbb{R}_+^{2|I|} \times \prod_{(i,j) \in R} (S_{ij} \times \mathbb{R}^2)$$

zodanig dat:

1. voor iedere consument i de combinatie $(x_i, m_i, (t_{ki}, b_{ki})_{k \in \mathcal{L}(i)}, (s_{ij})_{j \in \mathcal{F}(i)})$ een oplossing voor zijn beslissingsprobleem is, en dat
2. voor iedere consument i geldt

$$x_i = \bar{x}_i + \sum_{k \in \mathcal{L}(i)} t_{ki} - \sum_{j \in \mathcal{F}(i)} t_{ij}$$

en

$$m_i = \bar{m}_i - \sum_{k \in \mathcal{L}(i)} b_{ki} + \sum_{j \in \mathcal{F}(i)} b_{ij}.$$

3 Enkele Resultaten

3.1 Existentie

De eerste vraag die zich voor het model van een hiërarchisch gestructureerde markt stelt, is onder welke voorwaarden een evenwicht bestaat.¹⁰ D.w.z. onder welke specificaties het model vrij van tegenspraken is. Zulke specificaties worden in de volgende Existentie Stelling beschreven. Om te beginnen maken we evenwel een Technische (wiskundige) Aanname over de nutsfuncties en de institutionele karakteristieken.

Technische Aanname

Zij \mathbf{M} een hiërarchisch gestructureerde markt. Dan is de nutsfunctie van iedere

¹⁰Bij modellen van successieve monopolies is de existentie van een evenwicht niet gegarandeerd. Zelfs bij de directe interpretatie van het model van successieve monopolies als een hiërarchisch gestructureerde markt verdwijnt dit probleem niet. Zie bijvoorbeeld Spanjers, Gilles en Ruys (1990) of Spanjers (1997, Example 3.3.1).

consument $i \in I$, U_i , een continue neo-klassieke nutsfunctie.¹¹ Daarenboven geldt voor de institutionele karakteristiek van iedere handelsrelatie $(i, j) \in R$, $\mathcal{T}_{ij} : S_{ij} \rightrightarrows Y_{ij} \subset \mathbb{R}^2$, dat er een $a_{ij} \in \mathbb{N}$ bestaat, zodanig dat $S_{ij} \subset \mathbb{R}^{a_{ij}}$.

Existentie Stelling¹²

Zij \mathbf{M} een hiërarchisch gestructureerde markt waarvoor de Technische Aanname geldt en voor iedere handelsrelatie $(i, j) \in R$:

Begrensdheid S_{ij} en Y_{ij} begrensde verzamelingen zijn, en

Geen gedwongen handel voor ieder signaal $s \in S_{ij}$ geldt $(0, 0) \in \mathcal{T}_{ij}(s)$ en er bovendien een signaal $\hat{s} \in S_{ij}$ bestaat zodanig dat $\mathcal{T}_{ij}(\hat{s}) = \{(0, 0)\}$.

Dan bestaat er een evenwicht in \mathbf{M} .

De eerste voorwaarde van de Existentie Stelling, begrensdheid, zegt het volgende. In iedere institutionele karakteristiek mogen de hoeveelheden die gekocht en verkocht kunnen worden, evenals de bijbehorende betalingen, een van te voren vastgelegd maximum niet overschrijden. Het is echter mogelijk, dat deze bovengrens bijzonder hoog is, bijvoorbeeld hoger dan de totale hoeveelheden geld en goed die in de markt beschikbaar zijn.

De tweede voorwaarde verlangt dat noch de volger, noch de leider in een handelsrelatie tot handel gedwongen kan worden. Deze voorwaarde benadrukt het verschil tussen een markt en een bedrijf of de overheid. Binnen een markt hebben beide handelspartners de mogelijkheid niet te handelen, zonder dat dit het einde van de handelsrelatie betekent.

Binnen een bedrijf kan een medewerker weliswaar weigeren opdrachten uit te voeren, maar in sommige gevallen enkel op straffe van ontslag. Het niet nakomen van door de overheid opgelegde verplichtingen kan consequenties hebben die ernstiger zijn dan ontslag. Zo kan niet naleven van wetten tot langdurige gevangenisstraffen leiden.

De Existentie Stelling is voor het model van een hiërarchisch gestructureerde markt van grote betekenis. Ze zegt immers, dat het model consistent is indien aan twee plausibel klinkende economische voorwaarden voldaan is. Wat dit betreft is het model buitengewoon robuust en bijzonder geschikt als basis voor verdergaand onderzoek.

3.2 Equivalentie

Nu de vraag naar de consistentie van het model beantwoord is, stelt zich een tweede vraag, namelijk of het model van een hiërarchisch gestructureerde markt “normale” markten als speciale gevallen omvat. Als dit niet zo is, is dat een groot nadeel. Het

¹¹Een nutsfunctie is **neo-klassiek** als ze een preferentie relatie weergeeft, welke danwel: (1) strikt monotoon en strikt convex is, of (2) strikt monotoon en strikt convex is op het inwendige van de consumptieverzameling, hier \mathbb{R}_{++}^2 , en bovendien iedere bundel in het inwendige strikt beter is dan iedere bundel op de rand. Zie ook Aliprantis, Brown en Burkinshaw (1990, Def. 1.3.4).

¹²Deze Existentie Stelling volgt direkt uit Spanjers (1997, Theorem 6.2.1, blz. 130).

model heeft immers niet tot doel een alternatief voor het gebruikelijke marktmodel te zijn, dat meer dan honderd jaar trouwe dienst heeft gedaan en de basis van de algemene economie vormt. Het doel is de rol van handelsrelaties in een markt te verduidelijken.

De Equivalentie Stelling laat zien dat monopolistische markten een speciaal geval van hiërarchisch gestructureerde markten zijn.¹³ Hieruit volgt dat markten met volledige mededinging eveneens een speciaal geval zijn. Voordat we echter tot de formulering van de Equivalentie Stelling over kunnen gaan, moeten we de in deze markten gebruikte institutionele karakteristiek beschrijven.

Zij \mathbf{M} een hiërarchisch gestructureerde markt. Zij

$$\bar{x} := \sum_{i=1}^n \bar{x}_i \quad \text{en} \quad \bar{m} := \sum_{i=1}^n \bar{m}_i$$

respektievelijk de hoeveelheden goed en geld die in de markt beschikbaar zijn. Nu wordt de verzameling A voor een willekeurige waarde $a > 0$ gedefinieerd als

$$A := \{(t, b) \in \mathbb{R}^2 \mid -\bar{x} - a \leq t \leq \bar{x} + a \quad \text{en} \quad -\bar{m} - a \leq b \leq \bar{m} + a\}.$$

De verzameling A is een verzameling van combinaties van handel en betaling waarvoor de gehandelde hoeveelheid goederen groter kan zijn dan de hoeveelheid goederen die in de markt beschikbaar is, maar niet willekeurig groot. Hetzelfde geldt voor de bijbehorende betaling.

De karakteristiek $\widehat{\mathcal{T}}^{mon} : \mathbb{R}_+ \rightrightarrows A$ is een aangepaste versie van “normale prijszetting”, \mathcal{T}^{mon} , die de Technische Aanname vervult. Voor alle $s \geq 0$ is het signaal s de prijs voor het goed en geldt:

$$\widehat{\mathcal{T}}^{mon}(s) := \mathcal{T}^{mon}(s) \cap A,$$

waarbij

$$\mathcal{T}^{mon}(s) := \{(t, b) \in \mathbb{R}^2 \mid b = s \cdot t\}.$$

Equivalentie Stelling.¹⁴

Zij \mathbf{M} een hiërarchisch gestructureerde markt waarvoor de Technische Aanname geldt en bovendien:

1. *iedere handelrelatie de institutionele karakteristiek $\widehat{\mathcal{T}}^{mon}$ heeft.*
2. *er een keten $(i_1, i_2, \dots, i_n, i_1)$ van consumenten bestaat*
 - (a) *waarin alle consumenten in de markt voorkomen en*
 - (b) *waarbij iedere twee opeenvolgende consumenten door een handelsrelatie verbonden zijn.*

¹³Dit resultaat is eenvoudig uit te breiden naar monopolistische prijsdifferentiatie van de derde soort. Dat markten met monopolistische prijsdifferentiatie van de derde soort een speciaal geval van hiërarchisch gestructureerde markten zijn, volgt direct uit Spanjers (1997, Theorem 6.2.3, blz. 133).

¹⁴Deze stelling volgt direct uit Spanjers (1997, Theorem 6.2.3, blz. 133).

Dan is \mathbf{M} een monopolistische markt waarbij de consument die het hoogste in de informatiehierarchie staat, consument 1, de monopolist is.

Gevolg

Zij \mathbf{M} een hiërarchisch gestructureerde markt waarvoor de aannames van de Equivalentie Stelling gelden en bovendien consument 1 niets heeft, d.w.z., $\bar{x}_1 = 0$ en $\bar{m}_1 = 0$. Dan is \mathbf{M} een markt met volledige mededinging.

De structuur van handelsrelaties in de Equivalentie Stelling zorgt ervoor, dat op alle handelsrelaties binnen de markt dezelfde prijs geldt. Indien dit niet zo is, is er minstens één consument, die winst zou kunnen maken door arbitrage. D.w.z. hij kan het goed voor de lage prijs kopen (of door een volger laten leveren) om het tegen een hoge prijs verkopen (of door een volger laten kopen). Zijn winst gaat ten laste van de consument die het prijsverschil heeft veroorzaakt. Het veroorzaken van een dergelijk prijsverschil is dus niet in het eigenbelang van deze consument. Daarom komen dergelijke prijsverschillen in het evenwicht niet voor.

Het resultaat dat volledige mededinging een speciaal geval van een hiërarchisch gestructureerde markt is volgt nu direkt. Vanwege de Equivalentie Stelling geldt voor alle handelsrelaties binnen de markt dezelfde prijs. Indien de monopolist niets heeft, blijft hem niets anders over dan een prijs te zetten waarvoor de rest van de markt niets van hem wil kopen en niets aan hem wil verkopen. Maar dit is precies de prijs waarvoor vraag en aanbod in deze markt gelijk zijn, de prijs bij volledige mededinging.

Bij dit laatste resultaat dringt zich de vraag op, of een combinatie van de Existentie Stelling en de Equivalentie Stelling een nieuw bewijs van de existentie van een evenwicht onder volledige mededinging mogelijk maakt.

Om deze vraag te beantwoorden moet de institutionele karakteristiek $\hat{\mathcal{T}}^{mon}$ zodanig veranderd worden dat aan de voorwaarden van de Existentie Stelling voldaan wordt. Dit betekent, dat de leider in de handelsrelatie de mogelijkheid moet krijgen niet te handelen.

Hiervoor brengen we een kleine verandering aan in de institutionele karakteristiek $\hat{\mathcal{T}}^{mon} : \mathbb{R}_+ \rightrightarrows A$ en beschouwen we het signaal $s = 0$ niet langer als een prijs, maar als de wens niet over deze handelsrelatie te handelen. Zo ontstaat de institutionele karakteristiek $\mathcal{T}^{mon*} : \mathbb{R}_+ \rightrightarrows A$ waarbij voor alle $s > 0$ geldt

$$\mathcal{T}^{mon*}(s) := \hat{\mathcal{T}}^{mon}(s) = \{(t, b) \in \mathbb{R}^2 \mid b = s \cdot t\} \cap A$$

en voor $s = 0$ geldt

$$\mathcal{T}^{mon*}(0) := \{(0, 0)\}.$$

Deze institutionele karakteristiek voldoet aan de voorwaarden van de Existentie Stelling. Hieruit volgt, dat ook voor een hiërarchisch gestructureerde markt als in het Gevolg met de institutionele karakteristiek \mathcal{T}^{mon*} in plaats van $\hat{\mathcal{T}}^{mon}$ een evenwicht bestaat. Maar is dit evenwicht altijd een evenwicht met volledige mededinging? Het antwoord is nee!

Bij \mathcal{T}^{mon*} heeft consument 1, anders dan bij $\hat{\mathcal{T}}^{mon}$, de mogelijkheid van handel met zijn volgers af te zien. Hiertoe kan hij voor iedere handelsrelatie $(1, j)$ het

signaal $s_{1j} = 0$ kiezen. In dit geval trekt consument 1 zich feitelijk uit de markt terug ten gunste van consument 2. In de zo overgebleven markt hoeven de voorwaarden voor het Gevolg of de Equivalentie Stelling echter niet meer te gelden. Daarom is het, zelfs als een evenwicht in de overgebleven hiërarchisch gestructureerde markt bestaat, niet meer zeker, dat voor alle handelsrelaties dezelfde prijs geldt. Laat staan dat een dergelijke prijs de prijs bij volledige mededinging is.

4 Conclusies

In deze bijdrage werd een model van een hiërarchisch gestructureerde markt besproken, dat de gevolgen van stabiele handelsrelaties in markten verduidelijkt. Om de existentie van een evenwicht, en daarmee de consistentie van het model, te garanderen, waren zwakke en plausible aannames voldoende. Bovendien bleken monopolistische markten en markten met volledige mededinging op een natuurlijke wijze speciale gevallen te zijn van hiërarchisch gestructureerde markten.

Interessant zijn ook de inzichten die uit een nadere beschouwing van de basisstructuur van het model volgen. Hierbij speelt de informatiehiërarchie, die nodig is om het principe van “backward induction” toepasbaar te maken, een belangrijke rol. De rol die informatie speelt bij de bepaling van het evenwicht wordt bijzonder duidelijk. De marktdeelnemers hoeven niet alle individuele en institutionele karakteristieken te kennen voor het deel van de markt dat lager in de informatiehiërarchie staat. Het is voldoende over *geaggregeerde* informatie te beschikken, welke dan door het zetten van signalen stukje voor beetje gedesaggregeerd wordt.¹⁵ Dit benadrukt de rol van markten als informatie verwerkende instituties, die voor von Hayek (1945) van wezenlijke betekenis is.

Een belangrijke uitdaging ligt in het verklaren van de institutionele karakteristieken van de verschillende handelsrelaties, van de structuur van handelsrelaties, als mede van de specifieke informatiehiërarchie.¹⁶ Indien prijszetting de enige toegestane institutionele karakteristiek is, lijkt het plausibel dat een structuur van handelsrelaties ontstaat, die tot een uniforme prijs in de markt leidt. Voor twee consumenten die ieder met een andere prijs geconfronteerd worden, zou het zinvol kunnen zijn een handelsrelatie op te bouwen om van dit prijsverschil te profiteren.

Indien ook andere institutionele karakteristieken toegestaan worden lijkt het evenwel onwaarschijnlijk, dat de optimale institutionele karakteristiek voor deze handelsrelatie die van prijszetting is. Vermoedelijk zullen beide consumenten een maximaal handelbare hoeveelheid willen afspreken om de prijsverschillen niet te laten verdwijnen. Anders zou de bron van hun winsten immers snel opdrogen.

De eigenlijke kracht van hiërarchisch gestructureerde markten ligt echter elders. Het model is buitengewoon flexibel en robuust en is niet tot markten beperkt. Het kan in principe ook gebruikt worden om andere organisaties te modelleren waarin beslissingen gecoördineerd worden. Dit geldt in het bijzonder voor bedrijven. De organisatie van de productie tussen verschillende medewerkers of verschillende afdelin-

¹⁵Spanjers (1992, Section 8.3) en Spanjers (1997, Section 4.4).

¹⁶Voor een voorbeeld van het endogeniseren van de informatiehiërarchie zie Spanjers (1992, Chapter 7). Voor het endogeniseren van de structuur van handelsrelaties zie Spanjers (1997, Section 4.5 en Section 4.6). Zie ook Tirole (1988, Chapter 4).

gen wordt dan gemodelleerd op basis van bilaterale samenwerking. Een belangrijk verschil met een hiërarchisch gestructureerde markt is dat de samenwerking binnen een bedrijf niet zo makkelijk kan worden beëindigd. Op deze wijze kan een model ontstaan dat de grenzen tussen markten en bedrijven bepaalt en zo de resultaten van Coase (1937) en Williamson (1975) verdiept.

Referenties

- Aliprantis, C., D. Brown en O. Burkinshaw (1990), *Existence and Optimality of Competitive Equilibria*, Springer Verlag, Berlin.
- Coase, R.H. (1937), “The Nature of the Firm”, *Economica*, Vol. 4, 386–405.
- Grodal, B., en K. Vind (1989), “Equilibrium with Arbitrary Market Structure”, *Mimeo*, University of Copenhagen.
- Hahn, F. (1978), “On Non-Walrasian Equilibrium”, *Review of Economic Studies*, Vol. 45, 1–17.
- Hayek, F.A. (1945), “The Use of Knowledge in Society”, *American Economic Review*, Vol. 35, 519–530.
- Nash, J. (1950), “Equilibrium Points in n -Person Games”, *Proceedings of the National Academy of Sciences*, Vol. 36, 48–49.
- Spanjers, W. (1992), *Price Setting Behavior in Hierarchically Structured Economies*, Proefschrift, Katholieke Universiteit Brabant, Tilburg.
- Spanjers, W. (1997), *Hierarchically Structured Economies: Models with Bilateral Exchange Institutions*, Theory and Decision Library, Series C, Vol. 15, Kluwer Academic Publishers, Dordrecht.
- Spanjers, W., R.P. Gilles en P.H.M. Ruys (1990), “Hierarchical Trade and Downstream Information”, *FEW Research Memorandum 508*, Katholieke Universiteit Brabant, Tilburg.
- Tirole, J. (1988), *The Theory of Industrial Organization*, MIT Press, Cambridge Mass.
- Vind, K. (1983), “Equilibrium with Coordination”, *Journal of Mathematical Economics*, Vol. 12, 275–285.
- Williamson, O.E. (1975), *Markets and Hierarchies: Analysis and Antitrust Implications*, Free Press, New York.

POSITIEGERELATEERDE LONEN, CULTUUR EN BEDRIJFSSTRUCTUUR

René van den Brink*

Vrije Universiteit

Ik heb Pieter Ruys in 1983 leren kennen toen ik als eerstejaars student Econometrie aan de Katholieke Universiteit Brabant (toen nog Katholieke Hogeschool Tilburg) het vak Oriëntatie Econometrie volgde. Daarna heb ik nog colleges Inleiding Wetenschappelijke Economie en Bedrijfstakorganisatie bij Pieter gevolgd. Zijn colleges waren zo inspirerend dat ik op een gegeven moment bij hem aanklopte met het verzoek mijn derdejaars scriptie over ‘Investeringen in Research en Development’ te begeleiden. Naar aanleiding van deze scriptie bracht Pieter mij in contact met Rob Gilles die op dat moment zijn promotieonderzoek aan de KUB uitvoerde. Dit leidde achtereenvolgens tot mijn afstudeerscriptie (‘Analysis of Social Positions of Agents in Hierarchically Structured Organizations’) en aansluitend promotieonderzoek (‘Relational Power in Hierarchical Organizations’). Tijdens mijn promotieonderzoek hebben Pieter, Rob en ik veel van gedachten gewisseld. Als promotiebegeleider kon Pieter uitstekend zijn waardering voor het gedane onderzoek laten blijken, maar tegelijkertijd ook zijn kritiek daarop uiten.

Een half jaar na mijn promotie in 1994 hebben Pieter en ik, tijdens zijn verblijf op het NIAS, aan ons eerste gezamenlijke artikel gewerkt. Door onze vele gesprekken in de daaraan voorafgaande jaren kwamen we al snel op ideeën met betrekking tot het endogeen verklaren van de omvang en interne organisatie van bedrijven, waarin niet alleen rekening gehouden wordt met strategisch handelen, maar ook ruimte is voor sociale en maatschappelijke normen die de bedrijfsstructuur mede vormgeven. Deze gesprekken en het daaropvolgende NWO project ‘Oneerlijke Concurrentie en Overheidsbelang’ binnen het ESR onderzoeksprogramma ‘Competitie en Regulering’, hebben geleid tot een reeks artikelen over dit onderwerp. In deze bijdrage wil ik een overzicht geven van enkele van de daarin verkregen inzichten.

1 Inleiding

In de neo-klassieke theorie is een bedrijf een black-box die wordt weergegeven via een productiefunctie. Deze productiefunctie geeft aan hoe het bedrijf bepaalde pakketten van inputgoederen omzet in een pakket outputgoederen, maar er wordt niets gezegd over het proces volgens welke deze omzetting plaatsvindt. Een van de

* *Adres:* Vakgroep Econometrie, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam.
Email: jrbrink@feweb.vu.nl

eerste pogingen om een betere beschrijving van bedrijven te geven is de transactiekosten literatuur beginnend met Coase (1937). In meer recente modellen wordt het productieproces binnen bedrijven meer expliciet beschreven. Voorbeelden van deze modellen zijn de principal-agent modellen (zie, o.a. Holmström (1979), en Grossman en Hart (1983)), modellen van verticale integratie (zie, o.a. Klein, Crawford en Alchian (1978) en Grossman en Hart (1986)), incomplete contract modellen (zie, o.a. Grossman en Hart (1986), Hart en Moore (1990, 1999) en Maskin en Tirole (1999)), en optimal control modellen (zie o.a. Qian (1994)). Deze modellen maken gebruik van speltheoretische concepten om het gedrag van agenten binnen bedrijven te beschrijven. Deze speltheoretische concepten zijn meestal niet-coöperatieve concepten, en de problemen die bestudeerd worden zijn meestal 'incentive' problemen die samengaan met asymmetrische informatie en onzekerheid.

Hoewel deze studies interessante inzichten hebben opgeleverd, is een belangrijke kritiek op deze modellen het feit dat ze zich volledig concentreren op het strategisch gedrag van agenten. Een bedrijf is dan niet meer dan een verzameling individuele agenten die via samenwerking hun eigen individuele uitbetaling (nut) proberen te maximaliseren. De organisatie wordt volledig gevormd door de individuele eigenschappen van de agenten. Dit past in de traditie van het methodologisch individualisme, een belangrijk fundament van de micro economische theorie.

De gebruikelijke tegenpool van het methodologisch individualisme is het methodologisch holisme waarin individuele eigenschappen nauwelijks een rol spelen, en alles verklaard wordt uit globale eigenschappen. Ons onderzoek zit hier tussen in. Aan de ene kant zijn individuele eigenschappen van belang, maar zij zijn niet voldoende om economische verschijnselen volledig te verklaren. Derhalve kan ons onderzoek geplaatst worden onder het methodologisch relationisme. Ritzer en Gindoff (1992) stellen dat: "Methodological relationism takes the position that individuals *are* the basic components of social wholes such as groups and society. Relationism, however, also accepts the idea of emergence and therefore acknowledges that social wholes *are* more than the sum of the individual parts". Hieraan gerelateerd is ook de (hoofdzakelijk) sociologische literatuur op het gebied van 'Embeddedness'. Granovetter (1992) begint met: "Institutional economics has moved from a position, earlier in the twentieth century, of drawing electically on several other disciplines, to a stance of building its arguments almost entirely out of neoclassical materials." Hij beargumenteerd dat "such a stance cannot provide a persuasive account of economic institutions, and suggests a broader foundation based on classical sociological arguments about the embeddedness of economic goals and activities in socially oriented goals and structures". Granovetter (1985) stelt, dat "Actors do not behave or decide as atoms outside a social context nor do they adhere slavishly to a script written for them by the particular intersection of social categories that they happen to occupy. Their attempts to purposive action are instead embedded in concrete, ongoing systems of social relations". Soortgelijke ideeën zijn te vinden in de 'Theory of Action', zie o.a. Coleman (1986). In de economische literatuur zijn deze ideeën terug te vinden in de New Institutional Economics, zie Menard (2000) voor een overzicht.

Ons uitgangspunt bij het bestuderen van economische organisaties zijn de economische agenten, maar het is noodzakelijk om die in een sociale (organisatorische)

context te plaatsen. Laten we daarom eerst enkele definities en aannames met betrekking tot economische agenten en organisaties geven. Een *economische agent* is een persoon die, gegeven een verzameling alternatieve acties waaruit hij kan kiezen, beslissingen neemt met betrekking tot welke acties te ondernemen. Een *economische organisatie* bestaat uit een verzameling van participerende agenten, een verzameling relaties tussen deze agenten, en een aantal normen en regels volgens welke men in de organisatie beslissingen neemt. Voorbeelden van zulke economische organisaties zijn markten en bedrijven. Kijken we bijvoorbeeld naar de markt van een bepaald goed, dan zijn de agenten op de markt de verschillende vragers naar en aanbieders van dat goed. In het algemeen kunnen niet alle vragers en aanbieders direct met elkaar ruilen, maar heeft iedere agent een beperkt aantal 'relaties' waarmee direct geruild kan worden. (Uiteraard kan een agent via indirecte ruil veel meer andere agenten bereiken.) De agenten in een bedrijf zijn de verschillende werknemers (managers en uitvoerders) en eigenaren van het bedrijf. Bedrijven hebben een (meestal hiërarchische) organisatiestructuur waarin diverse verantwoordelijkheden worden vastgelegd.

Normen en regels van een markt kunnen per land of cultuur sterk verschillen. In de ene cultuur is het ondenkbaar dat een mondelinge afspraak (bijvoorbeeld tot het verrichten van een bepaalde transactie) geschonden wordt, terwijl in een andere cultuur een schriftelijke bevestiging noodzakelijk is. Zo ook is het in de ene cultuur gebruikelijk om eerlijk de sterke en zwakke punten van een te verkopen product te noemen, terwijl het in een andere cultuur zeer ongebruikelijk is dat een verkoper vrijwillig de zwakke punten van een door hem aangeboden product noemt. Ook voor de relaties binnen bedrijven gelden normen en regels die per cultuur verschillen, bijvoorbeeld met betrekking tot toegestane inkomensverschillen. In de ene cultuur zullen grote inkomensverschillen meer geaccepteerd, en derhalve gemakkelijker te realiseren, zijn dan in een andere cultuur.

In de modellen die ik in deze bijdrage bespreek wordt er vanuit gegaan dat de normen en regels in een economische organisatie vastliggen, en mogelijk cultureel bepaald zijn. Agenten die een organisatie binnentreden onderwerpen zich vrijwillig aan de normen en regels van die organisatie. Agenten die binnen een organisatie opereren identificeren zich met deze organisatie en zijn loyaal tegenover de normen en regels van de organisatie. Aan de andere kant stimuleren organisaties deze loyaliteit (zie o.a. Coleman (1980), Rosen (1982) en Simon (1991)).

In de komende paragrafen zal ik eerst het (coöperatief speltheoretische) model van een bedrijf met een exogeen gegeven omvang bespreken. Met behulp van dit model is te zien hoe de scheefheid van de inkomensverdeling binnen een bedrijf afhangt van de substitueerbaarheid van arbeid. Vervolgens gebruiken we dit model om de omvang van een bedrijf endogeen te verklaren. In eerste instantie bespreken we een partieel evenwichtsmodel waarin de reserveringslonen van werknemers exogeen gegeven zijn. Vervolgens plaatsen we dit bedrijfsmodel in een algemeen evenwichtscontext waarin de reserveringslonen endogeen bepaald worden als evenwichtslonen op de arbeidsmarkt. Tot slot geven we aan hoe culturele aspecten á la Hofstede (1980) van invloed kunnen zijn op de optimale bedrijfsomvang. Daarmee is de optimale bedrijfsomvang cultuur afhankelijk. Voor bepaalde waarden van de cultuurparameters

zal er geen evenwicht bestaan. Als er wel een evenwicht bestaat kunnen we spreken van een cultureel ondersteund evenwicht.

2 De interne organisatie van een bedrijf met vaste omvang

In van den Brink (1996) wordt een bedrijf gemodelleerd als een tripel (N, f, S) waarin N de verzameling medewerkers, f de produktiewaardefunctie en S de hiërarchische bedrijfs- of coördinatie structuur voorstelt¹. De *bedrijfsstructuur* S kan weergegeven worden als een boom met een unieke toppositie. Dat wil zeggen dat alle posities, behalve de toppositie, precies één direct superieure positie boven zich hebben en allen (direct of indirect) ondergeschikt zijn aan de toppositie (zie figuur 1).

De toppositie wordt ingenomen door de topmanager die aan het hoofd van het bedrijf staat. In een boomstructuur zijn er ook altijd posities die geen ondergeschikte posities hebben. Deze posities vormen het onderste niveau van de bedrijfsstructuur. Zoals in o.a. Simon (1957), Williamson (1967), Keren en Levhari (1979, 1983), en Radner (1992) nemen ook wij aan dat de toegevoegde waarde gegenereerd wordt door de werknemers op het onderste niveau van de bedrijfsstructuur. We noemen deze werknemers ook wel *uitvoerders* en hun positie de uitvoerderposities in het bedrijf. Deze posities vormen het domein van de produktiewaardefunctie. De *produktiewaardefunctie* f wijst aan ieder pakket van arbeidsinputs de waarde van de geproduceerde output toe. Uitgaand van volledige mededinging op de outputmarkt wordt de produktiewaarde verkregen door de outputhoeveelheid te vermenigvuldigen met de gegeven positieve outputprijs p .

De andere werknemers zijn *managers* die verantwoordelijk zijn voor de coördinatie van het productieproces². We gaan uit van complementariteit van arbeid, ofwel de produktiewaardefunctie is *supermodulair*. Zoals beargumenteerd in Milgrom en Roberts (1994), is met name voor zulke productieprocessen het coördineren van taken van groot belang³.

In het vervolg concentreren wij ons op bedrijven met een vaste spanwijdte s . Dat wil zeggen dat iedere manager hetzelfde aantal, s , direct ondergeschikten heeft. Als gevolg hiervan wordt de omvang van een bedrijf volledig bepaald door de diepte n . (In het voorbeeld in figuur 1 is zowel de spanwijdte als de diepte gelijk aan 2.) Dus liggen voor een gegeven diepte n ook het aantal posities in het bedrijf vast, in het bijzonder het aantal manager- en uitvoerderposities. Gegeven een bedrijf van omvang n zijn er $\sum_{l=0}^n s^l = \frac{s^{n+1}-1}{s-1}$ posities in het bedrijf, waarvan s^n uitvoerderposities, en $\frac{s^n-1}{s-1}$ managerposities. De coördinatie-intensiteit, dat wil zeggen de ratio tussen

¹Dit model is gebaseerd op *coöperatieve spelen met een toestemmingsstructuur* zoals bestudeerd in Gilles, Owen en van den Brink (1992), van den Brink en Gilles (1996), Gilles en Owen (1999) en van den Brink (1994, 1997).

²Voor een overzicht van de coördinatietaken van managers in bedrijven verwijzen we naar Radner (1992).

³Wij concentreren ons op de inzet van arbeid in het productieproces en negeren daarmee materiële inputkosten.

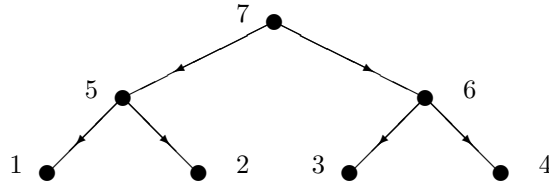


Figure 1: Bedrijfsstructuur met spanwijdte $s = 2$ en diepte $n = 2$

het aantal managerposities en het aantal uitvoerderposities (zie Gordon (1994)), is ongeveer gelijk aan $\frac{1}{s-1}$ als n groot genoeg is.

Uiteraard kan een bedrijf niet eindeloos doorgroeien zonder effect op de productiviteit van de uitvoerders op het onderste bedrijfsniveau. We nemen aan dat de effectiviteit van uitvoerders afneemt als de afstand tot de top toeneemt. Oorzaken van deze dalende effectiviteit zijn o.a. het informatieverlies dat samengaat met het doorgeven van informatie via meerdere intermediairs. De effectiviteit wordt vastgelegd via een *effectiviteitsparameter* $\alpha \in [0, 1]$, waarbij $1 - \alpha$ het procentuele verlies per niveau is⁴. Ofwel, de effectiviteit van een bedrijf met n niveaus is gelijk aan α^n . Vermenigvuldigen we de produktiewaarde van een bedrijf met α^n dan krijgen we de (netto) *toegevoegde waarde* van het bedrijf met n niveaus.

Het volgende aspect van een bedrijfsorganisatie is het beloningssysteem dat gebruikt wordt om de verschillende posities en taken in het bedrijf te waarderen. Alvorens een algemene klasse van beloningssystemen te introduceren bespreken we een specifiek beloningssysteem uit deze klasse. Gegeven dat alle werknemersposities actief zijn is de marginale bijdrage van een werknemer aan een bedrijf gelijk aan de vermindering in toegevoegde waarde als alle aan deze werknemer ondergeschikte uitvoerders op non-actief gesteld worden. Meer algemeen spreken we van de *marginale bijdrage* van werknemer i aan een (deel)verzameling werknemers F in het bedrijf als zijnde de vermindering in toegevoegde waarde als werknemer i stopt om actief te zijn, gegeven dat precies de werknemersposities in F actief zijn. Als een werknemer niet actief is dan staan alle aan deze werknemer ondergeschikte uitvoerders op non-actief. Het *gemiddelde marginale bijdrage*-, ofwel *GMB-beloningssysteem* verdeelt de toegevoegde waarde van een bedrijf proportioneel aan de gemiddelde marginale bijdragen van de werknemers, waarbij het gemiddelde over alle mogelijke verzamelingen actieve werknemers wordt genomen⁵.

⁴De effectiviteitsparameter kan afhankelijk gemaakt worden van de spanwijdte van het bedrijf.

⁵Merk op dat de som van de gemiddelde marginale bijdragen over alle werknemers niet gelijk hoeft te zijn aan de totale toegevoegde waarde, vandaar de normering. Dit beloningssysteem is gerelateerd aan de *Banzhaf waarde* voor coöperatieve TU-spelen (zie Banzhaf (1965)), en de multiplicatieve normalisatie daarvan zoals gekarakteriseerd in van den Brink en van der Laan

In plaats van alle mogelijke verzamelingen actieve werknemers gelijk te wegen, kunnen we de werknemers ook waarderen op basis van een gewogen gemiddelde van de marginale bijdragen. De klasse van beloningssystemen die we dan krijgen noemen we de klasse van marginale bijdrage, ofwel MB-beloningssystemen⁶. Aangezien de lonen volgens deze beloningssystemen, behalve door de taak in het productieproces, ook bepaald worden door de positie in de bedrijfsstructuur, noemen we dit *positiegerelateerde lonen*.

Zoals in de inleiding gezegd, heersen er in verschillende culturen andere normen met betrekking tot acceptabele inkomensverschillen. Zoals we in paragraaf 6 zullen zien kan dit tot uitdrukking gebracht worden in de keuze van het beloningssysteem. Met betrekking tot de bovenbeschreven MB-beloningssystemen kunnen we de volgende boven- en ondergrenzen voor verticale inkomensverschillen⁷ binnen bedrijven aangeven. Voor een bedrijf met een constante spanwijdte s ondersteunen o.a. Simon (1957), Williamson (1967), Calvo en Wellisz (1978, 1979), Carlson (1982), en Radner (1992), het verschijnsel dat de *loonratio*, dat wil zeggen het loon van een manager gedeeld door het gemiddelde loon van zijn direct ondergeschikten, tussen 1 en de spanwijdte s ligt. Van den Brink (1996) toont aan dat de beloningen in het bovenbeschreven model hieraan voldoen, en geeft ook een verklaring voor deze grenzen. Het blijkt dat deze ratio stijgt als de arbeid die door de uitvoerders op het laagste niveau geleverd wordt meer substitueerbaar wordt. Bovendien zijn deze grenzen scherp, in die zin dat er bedrijven zijn waarvoor deze grenzen bereikt worden. Als werknemer j *onmisbaar* is voor zijn directe superior i (dat wil zeggen als iedere uitvoerder die wel ondergeschikt is aan i maar niet aan j , aan iedere verzameling uitvoerders die niet ondergeschikt zijn aan j een marginale bijdrage gelijk aan nul heeft) dan zijn de beloningen van i en j aan elkaar gelijk, en dus is hun loonratio gelijk aan 1. Als alle ondergeschikte uitvoerders van manager i *dummy* uitvoerders zijn (dat wil zeggen, zij dragen aan iedere verzameling uitvoerders hetzelfde bij, namelijk datgene dat zij ook individueel kunnen genereren), dan is de beloning van i gelijk aan de som van de beloningen van al zijn direct ondergeschikten, en dus is de loonratio gelijk aan de spanwijdte s .

Voor produktiewaardefuncties met een constante substitutie-elasticiteit, de zogenaamde CES produktiewaardefuncties, betekent dit dat de maximale waarde, s , voor de loonratio bereikt wordt bij een lineaire produktiewaardefunctie (met substitueerbare arbeidsinputs). Het minimum, 1, wordt bereikt voor een Cobb-Douglas produktiewaardefunctie (met onmisbare arbeidsinputs).

(1998).

⁶Behalve het bovenbeschreven GMB-beloningssysteem bevat deze klasse o.a. ook het beloningssysteem waarin er vanuit gegaan wordt dat de werknemers een voor een het bedrijf binnenkomen, en iedere volgorde van binnenkomst gelijke kans van optreden heeft. Dit geeft een beloningssysteem dat is gerelateerd aan de *Shapley waarde* voor coöperatieve TU-spelen (Shapley (1953)).

⁷Vertikale inkomensverschillen betreffen inkomensverschillen tussen verschillende bedrijfsniveaus. Horizontale inkomensverschillen betreffen inkomensverschillen die afhangen van de uit te voeren taken en kunnen binnen hetzelfde bedrijfsniveau tot verschillende inkomens leiden.

3 Endogene bepaling van optimale bedrijfsomvang

In de vorige paragraaf hebben we de interne organisatie van een bedrijf met vaste bedrijfsomvang beschreven. We hebben gezien hoe de substitueerbaarheid van arbeid bepalend is voor de scheefheid van de interne inkomensverdeling. In van den Brink en Ruys (1996) wordt de bedrijfsomvang endogeen bepaald. Gegeven een vaste spanwijdte laten zij zien dat ook de optimale bedrijfsomvang afhankelijk is van de substitueerbaarheid van arbeidsinputs.

Gegeven een vaste spanwijdte s wordt de omvang van een bedrijf vastgelegd door de diepte n , dat wil zeggen het aantal hiërarchische niveaus in de bedrijfsstructuur. Nu veronderstellen we dat de topmanager van een bedrijf zelf kan kiezen hoe diep het bedrijf zal zijn⁸. Dat betekent dat we voor iedere mogelijke bedrijfsdiepte n een beschrijving van het bedrijf zoals in de vorige paragraaf moeten hebben. Een bedrijf met diepte n (zoals beschreven in de vorige paragraaf) zullen we daarom een *n-bedrijf* noemen. Een *bedrijf* definiëren we als een afbeelding die aan ieder natuurlijk getal n , een *n-bedrijf* toewijst.

De vraag is nu hoe de optimale bedrijfsomvang bepaald wordt. We nemen aan dat de topmanager die bedrijfsomvang kiest welke zijn eigen beloning maximaliseert. Echter, een noodzakelijke voorwaarde om de toegevoegde waarde te realiseren is dat de posities in het bedrijf daadwerkelijk bezet worden. In deze paragraaf nemen we aan dat de werknemers een exogeen gegeven positief *reserveringsloon* w hebben. Dit is het loon dat zij zichzelf op de arbeidsmarkt kunnen garanderen. Uiteraard zal de topmanager van een bedrijf de werknemers minstens dit loon moeten bieden. Dit heet de *participatie restrictie*. Het blijkt dat het loon van de uitvoerders op het laagste niveau daalt als het aantal niveaus van het bedrijf toeneemt⁹. Dus de optimale bedrijfsomvang is die waarbij de beloning van de topmanager maximaal is onder de voorwaarde dat de werknemers allen minstens hun reserveringsloon ontvangen. Gegeven reserveringsloon w en gegeven outputprijs p geven we de optimale bedrijfsomvang aan met $n(w, p)$. We hebben in de vorige paragraaf al gezegd dat een manager altijd minstens zoveel verdient als ieder van zijn ondergeschikten. Aannemende dat alle werknemers hetzelfde reserveringsloon hebben, betekent dit dat het voldoende is dat de uitvoerders op het laagste niveau minstens het reserveringsloon aangeboden krijgen¹⁰.

Merk op dat de enige manier om een bedrijf te laten groeien is om het aantal niveaus te laten stijgen. Dit leidt tot een grotere schaal (met meer uitvoerders actief), en dus een hogere (bruto) produktiewaarde. Er is echter ook een negatief effect op de toegevoegde waarde via de effectiviteitskosten. Vraag is dan of de optimale bedrijfsomvang eindig is. Van den Brink en Ruys (1996) tonen aan dat dat het geval is als de gemiddelde arbeidsproductiviteit per uitvoerder niet stijgt bij een toenemend aantal bedrijfsniveaus.

De participatie restrictie kan, maar hoeft niet beperkend te zijn voor de top-

⁸Als alternatief kunnen we de topositie ook aan de eigenaar van een bedrijf toewijzen, zie van den Brink en Ruys (1996).

⁹Merk op dat dit niet impliceert dat het loon in een gegeven niveau $k \leq n$ daalt als we een bedrijf vergroten van diepte n naar diepte $n + 1$.

¹⁰Het model is te generaliseren tot een gesegmenteerde arbeidsmarkt met verschillende reserveringslonen.

manager. Het blijkt dat voor een bedrijf dat voor iedere omvang volgens een Cobb-Douglas technologie met een constante gemiddelde arbeidsproductiviteit (op het laagste niveau) produceert, de beloning voor de topmanager ook dalend is in het aantal bedrijfsniveaus. Als gevolg is de optimale bedrijfsdiepte n^* gelijk aan 1. In het geval het bedrijf voor iedere omvang volgens een lineaire technologie met een constante gemiddelde arbeidsproductiviteit produceert, is de beloning voor de topmanager (voor redelijke waarden van de bedrijfsparameters¹¹) stijgend in het aantal bedrijfsniveaus. De optimale bedrijfsdiepte is dan de maximale diepte waarvoor het uitvoerdersloon groter of gelijk is aan het reserveringsloon. In dit geval wordt de optimale bedrijfsomvang dus volledig bepaald door het reserveringsloon van de werknemers. In andere gevallen zal de beloning voor de topmanager eerst dalen, en vanaf een bepaalde omvang stijgen. Als de beloning voor de topmanager bij het *reserveringsniveau* $n(w)$ (dat wil zeggen $n(w)$ is die diepte waarvoor het uitvoerdersloon gelijk is aan het reserveringsloon w) lager is dan de topuitbetaling in een 1-niveau bedrijf, zal de optimale bedrijfsomvang gelijk aan 1 zijn. In de andere gevallen zal de optimale bedrijfsomvang, zoals bij een lineair bedrijf, bepaald worden door het reserveringsloon (zie Figuur 1). Dit geeft een discontinuïteit in de optimale bedrijfsomvang als functie van het reserveringsloon, wat kan leiden tot een discontinuïteit in het aanbod van consumptiegoederen en de vraag naar arbeid van het bedrijf. We komen hier in de volgende paragraaf op terug.

4 Een institutioneel marktevenwicht: extern stabiel en intern optimaal

In de voorgaande paragrafen hebben we een partieel model van een hiërarchisch bedrijf besproken, en hebben we de optimale bedrijfsomvang endogeen bepaald via het maximaliseren van de beloning voor de topmanager, gegeven een participatie restrictie die zegt dat de werknemers minstens hun reserveringsloon aangeboden moeten krijgen. Dit reserveringsloon is exogeen gegeven, en is een primitieve weergave van de externe organisatie van het bedrijf. In deze paragraaf bespreken we het algemeen evenwichtsmodel van Ruys en van den Brink (1999), waarin het reserveringsloon endogeen bepaald wordt als het evenwichtsloon op de arbeidsmarkt. Daartoe dienen we eerst deze arbeidsmarkt, evenals de goederenmarkt voor het consumptiegoed dat door het bedrijf geproduceerd wordt, te definiëren.

In de vorige paragraaf hebben we, voor gegeven reserveringsloon w en gegeven outputprijs p de optimale bedrijfsomvang $n(w, p)$ bepaald. We gaan er nu vanuit dat het reserveringsloon w het evenwichtsloon op de arbeidsmarkt is, en de outputprijs p de evenwichtsprijs op de markt van het door het bedrijf geproduceerde consumptiegoed. In paragraaf 3 hebben we het aantal posities in een bedrijf van omvang n bepaald. Hieruit volgt dat, gegeven prijzen w en p , de vraag naar arbeid van het bedrijf gelijk is aan $d_l(w, p) = \frac{s^{n(w,p)+1}-1}{s-1}$. Er vanuit gaande dat de gemiddelde arbeidsproductiviteit van de uitvoerders op het onderste niveau gelijk (en genormeerd

¹¹Dat wil zeggen $\alpha s \geq e^{1/2}$, zie Williamson (1967).

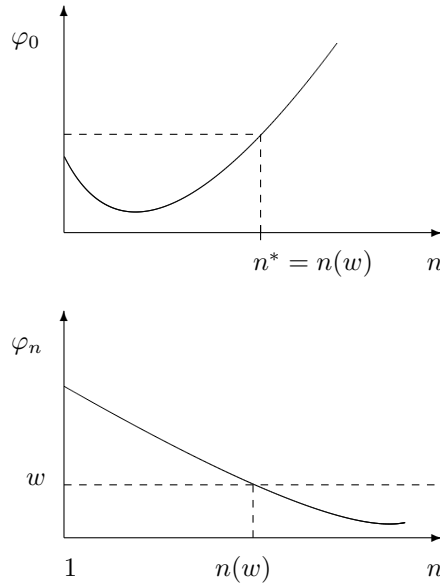


Figure 2: Toploon φ_0 en uitvoerdersloon φ_n in een n -bedrijf: optimale bedrijfsdiepte is n^*

op 1) blijft, is het aanbod van het consumptiegoed gelijk aan $s_c(w, p) = (\alpha s)^{n(w, p)}$. We nemen aan dat alle bedrijven op de markt identiek zijn, en het beschreven bedrijf dus als representatief voor de markt gezien kan worden. Uitgaande van een vast¹² aantal bedrijven m is de marktvrage naar arbeid en het marktaanbod van het consumptiegoed dus gegeven door $D_l(w, p) = m \bullet d_l(w, p)$, respectievelijk, $S_c(w, p) = m \bullet s_c(w, p)$.

De consumenten vragen consumptiegoederen en bieden arbeid aan. Iedere consument i is gekarakteriseerd door een initieel bezit b_i van het consumptiegoed, een hoeveelheid l_i van beschikbare tijd, en preferenties weergegeven met een reguliere nutsfunctie u^i , waarin $u^i(x, a)$ het nut is bij de consumptie van x eenheden van het consumptiegoed en $l_i - a$ eenheden vrije tijd (a is dus het individuele aanbod van arbeid). Maximaliseren van de individuele nutsfuncties geeft de individuele vrage naar het consumptiegoed en het individuele aanbod van arbeid. Sommeren over alle consumenten geeft de marktvrage $D_c(w, p)$ en het marktaanbod $S_l(w, p)$.

Voor dit institutioneel marktmodel definiëren Ruys en van den Brink (1999) een *Institutioneel Marktevenwicht* als een tripel (w, p, n) , zodanig dat n de optimale bedrijfsomvang gegeven de prijzen w en p is, en (w, p) de evenwichtsprijzen zijn gegeven dat de bedrijven van omvang n zijn. Zij tonen ook het bestaan van zo'n evenwicht aan als de optimale bedrijfsomvang een continue functie van de prijzen

¹²In Ruys en van den Brink (1999) is deze aanname gemotiveerd via een inelastische kapitaalmarkt.

w en p is¹³.

Echter, in de vorige paragraaf hebben we gezien dat dit in het algemeen niet het geval hoeft te zijn. We hebben daar ook gezien dat dit wel het geval is voor de twee extreme gevallen: een Cobb-Douglas- en een lineair bedrijf. In het geval van een bedrijf met een Cobb-Douglas produktie technologie hebben we gezien dat de optimale bedrijfsdiepte altijd kleiner of gelijk is aan 1. Als het reserveringsloon groter is dan het geboden loon bij een bedrijfsdiepte 1, dan is 1 de optimale bedrijfsdiepte en zijn de bijbehorende evenwichtsprijzen ook institutioneel markt-evenwichtsprijzen. Anders is de optimale bedrijfsdiepte gelijk aan 0 met institutioneel markt-evenwichtsprijzen zodanig gekozen dat werknemers niet in een bedrijf van omvang 1 willen werken. Voor een lineair bedrijf bestaat (voor redelijke waarden van de bedrijfsparameters) altijd een Institutioneel Marktevenwicht, omdat n dan continu is in de prijzen w en p .

5 Culturele dimensies

In het begin van deze bijdrage heb ik gemeld dat met het hier beschreven bedrijfsmodel niet alleen rekening gehouden wordt met strategisch handelen, maar ook ruimte is voor sociale en maatschappelijke normen die per cultuur of samenleving kunnen verschillen. Dit samenspel van strategisch handelen enerzijds en handelen volgens maatschappelijke normen anderzijds komt onder meer tot uitdrukking in de beloningen voor het werk dat geleverd wordt op de verschillende posities in een bedrijf. In een model waarin alleen strategisch handelen een rol speelt zullen werknemers nooit meer dan hun reserveringsloon aangeboden krijgen, hetgeen in ons model (net als in de modellen volgens de *efficiency wage theory*, zie o.a. Stiglitz (1976) en Akerlof (1984)) wel mogelijk is. Hoeveel de positiegerelateerde lonen boven de reserveringslonen liggen hangt onder andere af van het gebruikte beloningssysteem, wat weer afhangt van de geldende normen. Gegeven deze normen is er ruimte voor strategisch handelen van de werknemers door te kiezen voor deelname aan het bedrijf of het ontvangen van het reserveringsloon (al dan niet bepaald via een institutioneel marktmodel), en voor de topmanager die de bedrijfsomvang kiest welke, gegeven alle restricties, zijn eigen uitbetaling maximaliseert. In van den Brink, Ruys en Semenov (1999) en Ruys, van den Brink en Semenov (2000) zijn de maatschappelijke normen met betrekking tot het loon en de bedrijfsstructuur verklaard uit culturele dimensies zoals beschreven in Hofstede (1980), waaronder Power Distance, Uncertainty Avoidance en Collectivism-Individualism.

Power Distance geeft aan in hoeverre de minder invloedrijke leden van een maatschappij accepteren dat invloed ongelijk verdeeld is. Via empirische studies is aangetoond dat een grote Power Distance samengaat met diepe hiërarchische structuren. Het boven beschreven model kan hier een theoretische verklaring voor

¹³In de vorige paragraaf hebben we gezien dat een relatief hoog reserveringsloon alleen een (relatief) kleine bedrijfsomvang ondersteunt. In het extreme geval waarin de ratio $\frac{w}{p}$ naar oneindig gaat, gaat de optimale bedrijfsomvang $n(w, p)$ naar 0. Omgekeerd ondersteunt een (relatief) laag reserveringsloon een grote bedrijfsomvang. In het extreme geval waarin de ratio $\frac{w}{p}$ naar 0 gaat, gaat de optimale bedrijfsomvang $n(w, p)$ naar oneindig.

geven. Het GMB-beloningssysteem kan tot een grote ongelijkheid van beloning binnen bedrijven leiden. Dit volgt onder andere uit het feit dat iedere verandering in marginale bijdrage een verandering in het loon (in dezelfde richting) teweeg brengt, en het feit dat managers altijd minstens evenveel verdienen als hun ondergeschikten. Het beloningssysteem dat de laagste Power Distance uitdrukt is het *egalitaire* beloningssysteem volgens welke alle werknemers in een bedrijf hetzelfde loon ontvangen (ofwel alle werknemers krijgen eenzelfde aandeel in de toegevoegde waarde). Geven we met φ^ω het MB-beloningssysteem met een bepaald stelsel van gewichten $\omega = (\omega_F)_{FCN}$ voor de marginale bijdragen, en met γ het egalitaire beloningssysteem aan, dan is voor iedere $\beta \in [0, 1]$ het loon $\theta^{\beta,\omega} = \beta\gamma + (1 - \beta)\varphi^\omega$, de convexe combinatie waarin een gewicht β aan het egalitaire beloningssysteem gegeven wordt. Als we aannemen dat een lage Power Distance samengaat met een meer gelijke verdeling van verantwoordelijkheden, en dus meer gelijke beloningen binnen een bedrijf, dan kan het gewicht β als maat voor Power Distance gezien worden. Een lagere Power Distance geeft een hogere waarde voor β , en dus een groter gewicht op het egalitaire loon. Net zoals bij een lagere substitueerbaarheid van arbeid in CES produktiewaardfuncties, leidt een meer egalitair beloningssysteem tot een plattere bedrijfsstructuur.

Uncertainty Avoidance geeft aan hoe individuen in een samenleving zich bedreigd voelen door onzekerheid, en derhalve onzekerheid proberen te vermijden. In bedrijven komt dit tot uitdrukking door het geven van een vast inkomen in plaats van een inkomen dat afhankelijk is van de behaalde toegevoegde waarde¹⁴. Het meest extreme geval is dat waarin de onzekerheid volledig gedragen wordt door het topmanagement (of de eigenaar-ondernemer) van het bedrijf. Net als in Williamson (1967) ontvangen de werknemers dan een vast loon, en de top ontvangt het verschil tussen toegevoegde waarde en het uitbetaalde loon. Alle beloningssystemen die in de vorige alinea besproken zijn (dat wil zeggen, alle convexe combinaties van een MB-beloningssysteem en het egalitaire beloningssysteem) zijn afhankelijk van de behaalde toegevoegde waarde, en dus dragen alle werknemers mee in de onzekerheid. Dit is dus het andere extreme geval. Convexe combinaties van het vaste beloningssysteem en enig eerder beschouwd beloningssysteem $\theta^{\beta,\omega}$ kunnen weergegeven worden met $\theta^{\delta,\beta,\omega} = \delta w^* + (1 - \delta)\theta^{\beta,\omega}$, waarin $\delta \in [0, 1]$ het gewicht dat aan het vaste loonschema w^* gegeven wordt. Het gewicht δ is dan een maat voor Uncertainty Avoidance.

Tot slot betreft *Collectivism-Individualism* het belang van groepsgrootte in de samenleving. Des te groter het belang van groepsgrootte, des te meer collectief is de samenleving. Dit zegt nog niets over de geprefereerde groepsgrootte, maar over het belang van groepsgrootte. Zowel een samenleving waarin erg gehecht wordt aan produceren in grote bedrijven, als een samenleving waarin erg gehecht wordt aan produceren in familiebedrijven zijn collectieve samenlevingen. In het MB-beloningssysteem kan het belang van groepsgrootte uitgedrukt worden in het

¹⁴Merk op dat onzekerheid in ons deterministische model geen rol speelt. Formeel kunnen we alleen van Uncertainty Avoidance spreken als de toegevoegde waarde een stochastische variabele is. In het gepresenteerde model vertalen we Uncertainty Avoidance als het vermijden van beloningen die afhankelijk zijn van de toegevoegde waarde.

gewicht op de geprefereerde bedrijfsomvang in verhouding tot het gemiddelde gewicht. Het (ongewogen) GMB-beloningssysteem is een extreem geval. Alle gewichten zijn gelijk, ofwel groepsomvang doet er niet toe. Het GMB-beloningssysteem wijst dus op een individualistische samenleving. Er zijn meerdere beloningssystemen die een collectieve samenleving uitdrukken. De meest collectieve samenlevingen geven het volle gewicht aan een bepaalde groepsomvang. Dit kan groepsomvang 1 zijn, maar ook iedere andere omvang tussen 1 en de totale bevolkingsomvang.

6 Conclusies en uitbreidingen

We hebben laten zien hoe verschillende culturele dimensies vertaald kunnen worden in verschillende parameters van het bedrijfsmodel. Zoals aangegeven in de eerdere paragrafen bepalen deze parameters de interne organisatie van bedrijven, in het bijzonder de waardering- of beloningssystemen van de verschillende werknemersposities. Voor al deze beloningssystemen gelden de in paragraaf 2 t/m 4 beschreven resultaten. Hoewel voor al deze beloningssystemen de eindigheid van de optimale bedrijfsomvang dus is gegarandeerd, hoeft het Institutioneel Marktevenwicht, zoals gedefinieerd in paragraaf 5, in het algemeen niet te bestaan. Als de waarden van de cultuurparameters ω , β en δ zo'n evenwicht toestaan dan spreken we van een Cultureel Ondersteund Institutioneel Marktevenwicht. Verder onderzoek naar deze evenwichten en de waarden van de cultuurparameters waarvoor zij bestaan is een plan voor toekomstig onderzoek.

In het model zoals hier beschreven is alleen de aanbodzijde van de markt specifiek gemodelleerd. De consumenten in het institutioneel marktmodel van paragraaf 5 gedragen zich als individuele agenten. In van den Brink, Ruys en Semenov (1999) en Ruys, van den Brink en Semenov (2000) worden in een model van een diensteneconomie, behalve de effecten van cultuurparameters op bedrijven die diensten aanbieden, ook de effecten van cultuurparameters op clubs die diensten verkrijgen, besproken. Een meer gecompliceerd evenwichtconcept, waarin zowel de aanbodzijde (via bedrijven) als de vraagzijde (via clubs) is georganiseerd dient verder ontwikkeld en geanalyseerd te worden.

Referenties

- Akerlof, G. (1984), "Gift Exchange and Efficiency Wage Theory: Four Views", *American Economic Review Proceedings*, 74, 79-83.
- Banzhaf, J.F. (1965), "Weighted Voting Doesn't Work: A Mathematical Analysis", *Rutgers Law Review* 19, 317-343.
- Brink, R. van den, (1994), *Relational Power in Hierarchical Organizations*, Dissertation, Tilburg University, Tilburg.
- Brink, R. van den (1996), "Skewness of the Income Distribution in a Firm and the Substitutability of Labor Inputs", Research Memorandum FEW 739, Tilburg University.

- Brink, R. van den (1997), “An Axiomatization of the Disjunctive Permission Value for Games with a Permission Structure”, *International Journal of Game Theory* 26, 27–43.
- Brink, R. van den, en R.P. Gilles (1996) “Axiomatizations of the Conjunctive Permission Value for Games with Permission Structures”, *Games and Economic Behavior* 12, 113–126.
- Brink R van den, en G. van der Laan (1998) “Axiomatizations of the Normalized Banzhaf Value and the Shapley Value”, *Social Choice and Welfare* 15, 567-582.
- Brink, R. van den, en P.H.M. Ruys (1996), *The Internal Organization of the Firm and its External Environment*, CentER Discussion Paper 9643, Tilburg University, Tilburg.
- Brink, R. van den, P.H.M. Ruys, en R. Semenov (1999), Governance of Clubs and Firms with Cultural Dimensions, CentER Discussion Paper 99101, Tilburg University.
- Calvo, G.A., en S. Wellisz, (1978), “Supervision, Loss of Control, and the Optimum Size of the Firm”, *Journal of Political Economy*, 86, 943-952.
- Calvo, G.A., en S. Wellisz, (1979), “Hierarchy, Ability and Income Distribution”, *Journal of Political Economy*, 87, 991-1010.
- Carlson, N.W. (1982), “Time Rates Tighten their Grip on Manufacturing Industries”, *Monthly Labor Review*, 15-22.
- Coase, R.H. (1937), “The Nature of the Firm,” *Economica* 4, 386–405.
- Coleman, J.S. (1980), “Authority systems”, *Public Opinion Quarterly*, 44, 143-163.
- Coleman, J.S. (1986), “Social Theory, Social Research, and a Theory of Action”, *American Journal of Sociology*, 91, 1309-1335.
- Gilles, R.P., en G. Owen (1999), “Cooperative Games and Disjunctive Permission Structures,” *CentER Discussion Paper* 9920, CentER for Economic Research, Tilburg University, Tilburg.
- Gilles, R.P., G. Owen en R. van den Brink (1992), “Games with Permission Structures: the Conjunctive Approach”, *International Journal of Game Theory* 20, 277–293.
- Gordon, D.M. (1994), “Bosses of Different Stripes: a Cross-National Perspective on Monitoring and Supervision”, *American Economic Review*, AEA Papers and Proceedings, 375-379.
- Granovetter, M. (1985), “Economic Action and Social Structure: The Problem of Embeddedness”, *American Journal of Sociology*, 91, 481-510.

- Granovetter, M. (1992), "Economic Institutions as Social Constructs: A Framework for Analysis", *Acta Sociologica*, 35, 3-11.
- Grossman, S.J., en O.D. Hart (1983), "An Analysis of the Principal-Agent Problem," *Econometrica* 51, 7-45.
- Grossman, S.J., en O.D. Hart (1986), "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy* 94, 691-719.
- Hart, O., en J. Moore (1990), "Property Rights and the Nature of the Firm," *Journal of Political Economy* 98, 1119-1158.
- Hart, O., en J. Moore (1999), "Foundations of Incomplete Contracts," *Review of Economic Studies* 66, 115-138.
- Hofstede, G (1980), *Culture's Consequences. International Differences in Work-Related Values*, Sage Publications.
- Holmström, B. (1979), "Moral Hazard and Observability", *Bell Journal of Economics*, 10, 74-91.
- Keren, M., en D. Levhari (1979), "The Optimum Span of Control in a Pure Hierarchy", *Management Science*, 25, 1162-1172.
- Keren, M., en D. Levhari (1983), "The Internal Organization of the Firm and the Shape of Average Costs", *Bell Journal of Economics*, 14, 474-486.
- Klein, B., R.G. Crawford, en A. Alchian (1978), "Vertical Integration, Appropriate Rents, and the Competitive Contracting Process," *Journal of Law and Economics* 21, 297-326.
- Maskin, E., en J. Tirole (1999), "Unforeseen Contingencies and Incomplete Contracts," *Review of Economic Studies* 66, 83-114.
- Menard, C. (2000), *Institutions, Contracts and Organizations: Perspectives from New Institutional Economics*. Edward Elgar Publishing.
- Milgrom P., en J. Roberts (1994), "The Economics of Modern Manufacturing: Technology, Strategy and Organization", *Journal of Law and Economics* 19, 223-261.
- Qian, Y. (1994), "Incentives and loss of control in an optimal hierarchy", *The Review of Economic Studies*, 61, 527-544.
- Radner, R. (1992), "Hierarchy: The Economics of Managing", *Journal of Economic Literature*, 30, 1382-1415.
- Ritzer, G., en P. Gindoff (1992), "Methodological Relationism: Lessons for and from Social Psychology", *Social Psychology Quarterly*, 55, 128-140.

- Rosen, S. (1982), "Authority, Control, and the Distribution of Earnings", *Bell Journal of Economics*, 13, 311-323.
- Ruys, P.H.M., en R. van den Brink (1999), "Positional Abilities and Rents on Equilibrium Wages and Profits", in: *The Theory of Markets*, eds. J.J Herings, G. van der Laan, en A.J.J Talman, North-Holland, pp. 261-279.
- Ruys, P.H.M., R. van den Brink, en R. Semenov (2000), "Values and Governance Systems", in: *Institutions, Contracts, Organizations: Perspectives from New Institutional Economics*, (ed. C. Menard), Edward Elgar Publishing Co., pp 422-446.
- Shapley, L.S. (1953) "A Value for n -Person Games," in H.W. Kuhn en A.W. Tucker (editors), *Annals of Mathematics Studies 28: Contributions to the Theory of Games*, Volume 2, Princeton University Press, Princeton, NJ.
- Simon, H.A. (1957), "The Compensation of Executives", *Sociometry*, 20, 32-35.
- Simon, H.A. (1991), "Organizations and markets", *Journal of Economic Perspectives*, 5, 25-44.
- Stiglitz, J. (1976), "The Efficiency Wage Hypothesis, Surplus Labour, and the Distribution of Income in L.D.C.'s", *Oxford Economic Papers*, 28, 185-207.
- Williamson, O.E. (1967), "Hierarchical Control and Optimum Firm Size," *Journal of Political Economy* 75, 123-138.

OVER SAMENHANG

Chris van Raalte*

ICTRO – Strafrecht

De moderne westerse mens ziet zichzelf in eerste instantie als een individu dat met andere individuen op zijn eigen voorwaarden interageert. Het vroegere, min of meer vanzelfsprekende, besef van samenhang waarin het individu was ingebed en waarin zijn acties betekenis kregen, is goeddeels verdwenen. Steeds meer wordt de noodzaak gevoeld die samenhang te creëren vanuit het individuele. Vanuit de middeleeuwse situatie waarin individu en collectief in feite samenvielen, waarin de mens zich niet zag als individu, maar slechts in de rol die hij in het collectief speelde, begon hij via meer ‘verlichte’ percepties zichzelf los te zien van het collectief. In het kielzog daarvan veranderde ook de maatschappelijke structuur. Steeds meer kreeg de mens de mogelijkheid zich in verschillende rollen te bewegen, en deel uit te maken van verschillende maatschappelijke sectoren, en die ook zelf te creëren.

Vanuit deze perceptie hebben zich nieuwe maatschappelijke structuren ontwikkeld die veel meer individuele vrijheid bieden, zoals democratie en vrije markt. Het wegvallen van de oorspronkelijke structuren heeft echter ook een onderstroom van onzekerheid tot gevolg gehad, die, als het ware om het evenwicht te herstellen, tot verscheidene vormen van conformisme¹ heeft geleid. En waar men ‘flexibel’ dient te zijn, specifiek op het terrein van menselijke relaties, worden diezelfde relaties, in al hun veronderstelde vrijblijvendheid, tegelijkertijd overgewaardeerd. De nadruk op het individu heeft geleid tot het verheerlijken van de ongestructureerde persoonlijke expressie, resulterend in bijvoorbeeld een stroom van ‘candlelight’-gedichten die uitsluitend voor de maker betekenis hebben, of in ‘bekentenis’-programma’s waarin de deelnemers hun persoonlijke probleem tot nationale *issue* maken. Het onvoorwaardelijk serieus nemen van het persoonlijke is ook terug te vinden in de politiek correcte neiging om geen inhoudelijke oordelen meer te willen vellen. Net zoals een candlelight-gedicht gezien wordt als een ‘oprechte’ persoonlijke uiting, waarop als zodanig geen kwaliteitscriteria losgelaten kunnen worden, zo bestaat ook de neiging om een mening onlosmakelijk te koppelen aan degene die hem uit. Vanuit dit standpunt wordt iedere inhoudelijke aanval op een mening automatisch een persoonlijke aanval; ieder argument wordt per definitie *ad hominem*.²

Waar vanuit een gemis aan besef van samenhang de nadruk wordt gelegd op ‘gemeenschappelijkheid’, hebben de onderlinge verschillen de neiging om uit het

* *Adres:* ICTRO – Strafrecht, Ministerie van Justitie, Postbus 75, 3700 AB Zeist. *Email:* C.van.Raalte@ictro.drp.minjus.nl

¹Zie voor een uitwerking van dit thema Fromm (1952).

²De maatschappelijke gevolgen van het benadrukken van het persoonlijke domein en de daarmee samenhangende verheerlijking van het ‘slachtofferschap’ worden beschreven in Hughes (1994).

gezichtsveld te verdwijnen. Hierdoor wordt het werkelijke, het overstijgende, gemeenschappelijke kader niet meer waargenomen, waar de verschillen zich des te dwingender gaan opdringen. In de Griekse mythologie wordt Eros (god van de liefde; het bindende, overstijgende principe) onlosmakelijk gekoppeld aan Thanatos (god van de dood; het ontbindende, 'losmakende' principe). Eros zonder Thanatos leidt tot een illusionaire gemeenschappelijkheid, of symbiose; Thanatos zonder Eros leidt tot een illusionaire zelfvoorzienigheid. Ook de relatie van de godin Aphrodite (verbondenheid) en de god Ares (strijd), waaruit hun dochter Harmonia (harmonie) wordt geboren kan hiermee in verband worden gebracht.³

Niet alleen de maatschappelijke samenhang is grotendeels uit het gezichtsveld verdwenen. Dit geldt ook voor die niet concreet aanwijsbare, irrationele, 'transcendente' samenhang waarin ieders individuele bestaan is ingebed, traditioneel het domein van de religie. Geheel verdwenen is de perceptie van het transcendente echter niet; de existentialistische ontkenning ervan zou immers eenvoudigweg ondraaglijk zijn. Op dit punt bevindt de westerling zich echter in een netelige situatie. Rationeel moet hij besluiten tot de opvatting dat er geen samenhang, geen transcendent domein is, of liever gezegd, dat hij die zelf moet maken. Tegelijkertijd kan hij zich niet onttrekken aan een zekere perceptie van het transcendente.

In het sprookje "de Sneeuwkonigin"⁴ van Hans Christian Andersen wordt een beeld geschetst van deze situatie. Andersen beschrijft hoe de jongen Kay een splinter van een 'duivelsspiegel' in zijn oog krijgt en terecht komt in het paleis van de Sneeuwkonigin. Daar zit hij bij een bevroren meer dat in duizend stukken is gebroken.

"Hij sleepte rond met scherpe, platte stukken ijs, terwijl hij ze aan elkaar probeerde te passen. Elk stuk afzonderlijk leek hem het mooist, maar dat kwam door de glassplinter in zijn oog. Hij voegde ze samen tot een groot aantal vormen, echter, hij wilde er het woord 'Liefde' mee spellen. De Sneeuwkonigin had tegen hem gezegd: "Als je erin slaagt dat woord te spellen zul je meester over jezelf zijn. Ik zal je de hele wereld geven en een paar nieuwe schaatsen."

Maar hij slaagde niet."

Kay tracht de fragmenten van de werkelijkheid die hij waarneemt aan elkaar te smeden tot een 'allesomvattende theorie' die in zijn ogen equivalent is aan de samenhang ('liefde') waarvan hij slechts nog een vaag vermoeden heeft. Bovendien vindt hij ieder afzonderlijk stuk ijs het mooist. Een dergelijke benadering zag ik naar voren komen in een interview met een geneticus die meewarig alle vragen naar de zin van het bestaan van tafel veegde door te verklaren dat gewoon alle organismen genetisch in kaart dienden te worden gebracht, en "dat is het dan".

De westerling tracht op een rationele, noodzakelijkerwijs gefragmenteerde, manier de betekenis van zijn bestaan, op zijn eigen persoonlijke voorwaarden, vast te stellen: "Als het bestaan geen zin heeft, dan mkt het maar zin." In de sympathieke

³De wrijvingen die kunnen optreden tussen individualiteit en gezamenlijkheid in menselijke relaties worden besproken in Schellenbaum (1985).

⁴Gebaseerd op een Engelse vertaling gepubliceerd op de website <http://www.rickwalton.com/folktale/pink13.htm>.

film ‘The Accidental Tourist’ uit 1989 zie ik dit thema terugkomen. William Hurt speelt in deze film een licht-autistische schrijver. Zijn autisme manifesteert zich in de eerste plaats in zijn obsessie voor orde en voorspelbaarheid. Als schrijver van reisboeken informeert hij zijn lezers tot in het kleinste detail hoe zij zich tegen risico en verlies kunnen indekken. In de tweede plaats komt zijn autisme tot uiting in de relatie tot zijn vrouw Kathleen Turner, en later ook tot zijn vriendin Geena Davis. De relatie van de schrijver tot (veel meer dan met) de vrouwen is tweeslachtig. Hun aanwezigheid geeft hem wel een kader waarbinnen hij zijn eigen bestaan ‘op orde’ kan houden, maar aan de andere kant kan hij ze zelf niets terug geven, en zich niet daadwerkelijk met ze verbinden. De onmacht van de schrijver wordt overtuigend in beeld gebracht. Zijn afhankelijkheid van het kader dat ze hem verschaft blijkt als zijn vrouw hem verlaat; op dat moment helpen zijn obsessieve ordeningsrituelen hem ook niet meer.

Anderzijds is men zich nog wel vaag bewust van het transcendente dat zijn bestaan betekenis schenkt. Dit vage besef is uitgekristalliseerd in het onontkoombare cliché “ik geloof wel dat er iets is.” In zijn veel- en gelijk nietszeggendheid biedt dit cliché tevens een uitweg uit de double bind die ontstaat uit het ontkennen en gelijktijdig een perceptie hebben van het transcendente. Of, uit de discrepantie die bestaat tussen wat we beredeneren en wat we intuïtief ‘zien’, waarvan de volgende regels uit T.S. Eliot’s gedicht ‘The Waste Land’ (bijvoorbeeld in Eliot, 2000) getuigen:

“...
Who is the third who walks always beside?
When I count, there are only you and I together
But when I look ahead up the white road
There is always another one walking beside you
...”

Onder anderen uit de behoefte aan het ervaren van samenhang is het *ritueel* geboren. Rituelen maken het van oudsher mogelijk het transcendente domein te ervaren, zich opgenomen te weten in het maatschappelijke domein, en om het persoonlijke domein te structureren. Op welk niveau ook, rituelen geven de deelnemers het gevoel deel uit te maken van de ‘samenhang’; zij worden ingewijd in iets dat groter is dan henzelf. Met de geleidelijke concretisering en vermaatschappelijking van het transcendente door de kerk zijn ook de daarin ontwikkelde rituelen verdwenen of van hun betekenis ontdaan. Zoals de eucharistie in de traditionele rooms-katholieke liturgie nog een feitelijke verandering van brood en wijn in lichaam en bloed van Christus behelst, zo werd deze verandering na de Reformatie in het protestantse ‘avondmaal’ uitsluitend nog overdrachtelijk opgevat. Daarmee werd het mysterie generationaliseerd, en daardoor het ritueel van zijn (ervarings-)betekenis ontdaan. Hetzelfde kan gezegd worden over de modernisering van de liturgieën. Hoewel de ‘begrijpelijkheid’ van het geschrevene daardoor wellicht toeneemt, neemt tegelijkertijd zijn rituele betekenis af. De kracht van een ritueel is voor een belangrijk deel gelegen in zijn mysterieusheid; een Latijnse mis heeft ontgezeggelijk meer rituele kracht dan een Nederlandse. Het transcendente laat zich nu eenmaal niet begrijpelijk maken, of analyseren.

Op het gebied van de maatschappelijke rituelen is de kaalslag wat minder groot dan in de kerken. Het (burgerlijke) huwelijksritueel bijvoorbeeld heeft standgehouden, ondanks de luidruchtige aanwezigheid van de ‘alternatieve samenlevingsvormen’. De invulling van de plichtigheid zelf heeft echter meestal weer een sterk persoonlijk karakter, mede door het wegvallen van de religieuze context. De (strikt) persoonlijke invulling komt ook sterk naar voren in de moderne begrafenis-rituelen, waarbij net zo goed de religieuze context is verdwenen. De televisiespotjes waarmee uitvaartondernemingen zichzelf pleegden te profileren spelen hier op in. In een van die spotjes werd een man opgevoerd die glimlachend vertelt dat hij op zijn begrafenis ‘een feestje’ wil. Een nogal halfslachtige poging de dood tot hanteerbare proporties terug te brengen, om die van het transcendente in het persoonlijke domein ‘in te passen’. Halfslachtig, doordat het onherroepelijke, en vanuit een louter persoonlijk standpunt gezien, onrechtvaardige, aspect van de dood (*memento mori*) niet verdwijnt, maar hooguit toegedekt wordt onder een deken van vrijblijvend *carpe diem*. Slechts een transcendent ritueel maakt het mogelijk de tegenstrijdigheden van de gebeurtenis te verdragen, geeft zowel *memento mori* als *carpe diem* een plaats.

Er is recentelijk een nieuw ritueel ontstaan, dat de dood uit het strikt persoonlijke domein haalt en hem (opnieuw) in het maatschappelijke domein plaatst. Dit ritueel is de ‘stille tocht’, waarin collectief een, in feite willekeurig gekozen, overledene wordt herdacht. Het fenomeen kan niet los gezien worden van het gelijktijdig opgekomen cliché ‘zinloos geweld’. Dit cliché werd geïntroduceerd na een uit de hand gelopen vechtpartij met dodelijke afloop die ontstaan was naar aanleiding van opmerkingen van het slachtoffer naar de dader over diens gedrag. De afloop van de vechtpartij bleek, mede uit gesprekken met de zwaar aangeslagen dader, feitelijk een noodlottige samenloop van omstandigheden te zijn geweest. Vanuit dat standpunt gezien was het geweld inderdaad ‘zinloos’. De maatschappelijke perceptie van de vechtpartij was echter ongenueanceerd, en gaf het geweld een oneigenlijke betekenis: het slachtoffer kreeg de status van martelaar die de misstanden in de maatschappij aan de kaak had gesteld, de dader werd een gewetenloze moordenaar. Mocht dit de oorspronkelijke gedachte zijn geweest die ten grondslag ligt aan het begrip, later werd de term voor willekeurige gewelddadige situaties met dodelijke afloop gebruikt. De ongenueanceerde, gespleten perceptie van de maatschappij die opgeslagen ligt in het cliché is echter niet veranderd. Dit blijkt expliciet uit de nogal drammerige manier waarop men wordt opgeroepen zich uit te spreken ‘tegen zinloos geweld’. Een dergelijke stellingname moge triviaal zijn (niemand zal zich expliciet voor zinloos geweld uitspreken), zij heeft wel het vermogen tot het aanscherpen van de gespleten perceptie die ten grondslag lag aan het cliché.

De ‘stille tocht’ werd een ritueel waarmee de dood een nieuwe plaats werd gegeven. Hoewel zij ongetwijfeld voor de deelnemers betekenis heeft, is het wel zo dat men zich in wezen de dood van een onbekende ‘toe-eigent’ om ter wille van het ritueel. En ook al hebben de nabestaanden hier geen problemen mee, of weten zij zich er zelfs door gesteund, toch wringt dit. Nu gebeurt dit bij een ritueel als de staatsbegrafenis, waar de stille tocht zich op lijkt te baseren, op het eerste gezicht ook. Het grote verschil is echter dat de herdachte persoon bij de staatsbegrafenis bij leven reeds deel uitmaakte van het publieke domein; er kan dan geen sprake meer zijn van toe-eigening van zaken die feitelijk tot het persoonlijke domein behoren.

Met het eerder genoemde cliché “ik geloof wel dat er iets is” geeft men openlijk aan aan zichzelf genoeg te hebben, maar heimelijk verwacht men van het ‘iets’ dat het zijn bestaan betekenis geeft; echter wel binnen het beperkte gezichtsveld van het persoonlijke. Openlijk neemt men de ‘verantwoordelijkheid’ (aan zichzelf) voor zijn eigen leven op zich; heimelijk wordt het ‘iets’ ter verantwoording geroepen als het misloopt. Dit lijkt op de redenering waarmee dikwijls het atheïsme plausibel wordt gemaakt: “als God bestaat, waarom laat Hij dan lijden toe?” In zijn tweeslachtige houding is de westerling “een atheïst, maar hij doet er al een hele tijd niets meer aan.”⁵

Met de formulering van het ‘iets’ is een poging gedaan om het transcendente, het persoonlijke overstijgende, ‘in te passen’ in het persoonlijke. Het samenkomen van het transcendente en het persoonlijke domein op zich is niet iets wat zich nu voor het eerst manifesteert. In iedere religie (bijvoorbeeld) is het onvermijdelijk dat het transcendente op een bepaald niveau wordt geïnterpreteerd. Hoe persoonlijk ook, deze interpretatie past echter wel binnen het gemeenschappelijke kader, het ‘verhaal’. Met het ‘iets’ echter is van een dergelijk kader geen sprake meer. Of liever gezegd, het kader is een strikt persoonlijk kader geworden. Net zoals een candlelight-gedicht uitsluitend betekenis heeft binnen het persoonlijke domein, zo is ieders ‘verhaal’ een persoonlijk verhaal geworden. Het enige gemeenschappelijke is uiteindelijk het ‘iets’. Ook binnen de traditionele kerken is steeds minder overeenstemming over het ‘verhaal’; steeds meer wordt de individuele interpretatie als leidraad genomen. Een interpretatie die meestal de kant opgaat van het maatschappelijke en concrete; alle ‘historische bewijzen’ die daarbij worden opgevoerd om het ‘verhaal’ te ondersteunen, ondergraven dit juist nog verder. De kerk is niet langer intermediair tussen het persoonlijke en het transcendente domein, maar gaat in feite geruisloos ten onder in het maatschappelijke domein. Wat overblijft, is een onsamenhangend ‘braakland’ waar ‘niets met niets’ moet worden verbonden. Zoals T.S. Eliot in ‘The Waste Land’ (zie bijvoorbeeld Eliot, 2000) schrijft

“...
On Margate Sands.
I can connect
Nothing with nothing.
...”

De ‘inpassing’ van het transcendente in het persoonlijke heeft een wijdverbreide behoefte aan ‘zingeving’ doen ontstaan; een woord dat de paradoxale combinatie van de twee domeinen in zich draagt. Om in deze behoefte te voorzien is de ‘zingevingsmarkt’ ontstaan. Via deze markt is een keur aan boeken, cd’s, stenen, sieraden, workshops, seminars en ‘healings’ beschikbaar gekomen, waarmee het ‘iets’ aan de man wordt gebracht.

De zingevingsmarkt maakt het mensen mogelijk hun sterk persoonlijke wensen om hun bestaan ‘zin te geven’ adequaat in te vullen. Toch wordt deze markt soms met ‘oplichterij’ in verband gebracht. Een moeilijk verifieerbare beschuldiging. Men doelt waarschijnlijk op de notie dat geld gevraagd wordt voor eigenschappen

⁵Vrij naar Herman Finkers in zijn theaterprogramma *EHBO is mijn lust en leven*.

die niet getoetst kunnen worden. Nu is dit het geval voor het leeuwendeel van de consumptiegoederen. Consumenten laten zich immers voor een belangrijk deel leiden door het irrationele, niet toetsbare beeld dat zij van een product hebben. Misschien is het feit dat de zingevingsmarkt expliciet verwijst naar dit irrationele gebied, dat we zowel afwijzen als opzoeken, de reden dat de imagoverkopers van sigaretten en auto's niet als oplichters worden beschouwd, maar Jomanda wel.

De controverse die wordt ervaren tussen marktmechanisme en zingeving valt wellicht terug te voeren op de diepliggende perceptie van de tegenstelling tussen materiële en geestelijke zaken waar het christendom traditioneel van doortrokken is geweest. Op de een of andere manier schijnt het 'lage', materiële aspect van een markt niet te rijmen met het 'hoge', niet-materiële aspect van de zingeving. Zo moest Jomanda zich verdedigen tegen de aantijging dat zij winst zou maken met haar 'healings'. Zij verzekerde de interviewer dat zij slechts de huur van de hal die zij voor de healings gebruikte terugverdiende, maar het legde wel de vinger op het taboe. Ook de ophef om de financiële verwickelingen rond het *Oibibio* van Ronald-Jan Heijn was op zijn zachtst gezegd weinig gefundeerd. De indruk ontstond dat men het Heijn niet kon vergeven dat hij zich ingelaten had met de zingevingsmarkt. Of in ieder geval niet kon begrijpen, zoals (vermoed ik) Martin Simek. In een televisie-interview bleef Simek Ronald-Jan Heijn *ad nauseam* doorzagen over 'dat ene moment' dat hij terug wilde zien, hierbij doelend op een briljante, instinctieve actie van Heijn in een van zijn vroegere hockeywedstrijden. Dit naar aanleiding van het betoog dat Heijn had gehouden over het 'evenwicht' dat hij had gevonden. Het leek wel of er bij Simek een trekker werd overgehaald om Heijn zijn evenwicht te laten verliezen.

De (christelijke) houding tegenover materieel bezit is door de eeuwen heen sterk afwijzend geweest. Aan de andere kant was zij, juist door die afwijzing, ook ronduit tweeslachtig. Nergens is zoveel bezit opgebouwd als in de hogere regionen van de Katholieke kerk. Nooit werd er zo intensief gehandeld als tijdens de Calvinistische Gouden Eeuw. De hippiebeweging, die het 'spirituele paradijs' India verheerlijkte, kwam op in een periode van grote materiële bloei. Peter de Wit heeft in een van de afleveringen van zijn strip 'Sigmund' in de Volkskrant de dubbelheid van dat antimaterialisme op een prachtige manier in beeld gebracht. Hij voert een manager ten tonele, die naar India is gegaan om nieuwe inzichten te verwerven in zijn bedrijfsvoering. Op de vraag welk inzicht hij heeft meegenomen uit India om zijn zaken anders aan te pakken, antwoordt de manager: "Lage lonen, geen vakbonden en kinderarbeid." De alom gehoorde bewering dat we tegenwoordig te materialistisch zijn, zou daarom wel eens omgekeerd kunnen worden; het zou heel goed kunnen dat we juist niet materialistisch genoeg zijn. Dat wil zeggen, te weinig besef hebben van de invloed van materieel bezit op ons welzijn. Alle romantische beschouwingen over de voordelen van bezitloosheid, en alle krampachtige verzekeringen dat "geld niet gelukkig maakt" ten spijt, het materiële domein vormt wel de basis van ons bestaan, en geeft wellicht juist vanuit deze basis toegang tot het transcendente.

De zingevingsmarkt biedt een caleidoscopisch aanbod, 'voor elk wat wils', maar heeft tegelijkertijd het 'ets' als gemeenschappelijke achtergrond, een beeld dat ook wordt neergezet in de slotregels van het eerder aangehaalde 'The Waste Land':

“...
 London Bridge is falling down falling down falling down
 Poi s’ascose nel foco che gli affina
 Quando fiam uti chelidon - O swallow swallow
 Le Prince d’Aquitaine la tour abolie
 There fragments I have shored against my ruins
 Why then Ile fit you. Hieronymo’s mad againe.
 Datta, Dayadhvam. Damyata.
 Shantih shantih shantih
 ...”

een beschrijving van het uit ruïnes bestaande braakland, omspoeld door de fragmenten van een Babylonische spraakverwarring, die samen komen in de hindoeïstische gebedsregel ‘Shantih shantih shantih’.

Waar T.S. Eliot eindigt met een beeld van fragmentatie, biedt Andersen in ‘De Sneeuwkoningin’ een uitweg uit de fragmentatie. Net zo indringend als het eerdere beeld dat Andersen schetst van Kay die tevergeefs de stukken ijs ineen tracht te passen, is het beeld waarin zijn dilemma wordt opgelost. Waar Kay er niet in slaagde om de samenhang, het transcendente, ‘bij elkaar’ te redeneren, nu valt die hem vanzelf ten deel wanneer zijn intuïtie ontdooit, wanneer hij ‘ziet’, namelijk zijn vriendin Gerda, die hem gevolgd is tot in het paleis van de Sneeuwkoningin.

“... Gerda hilde hete tranen die op zijn hals vielen en zijn hart ontdooiden en de glassplinter van de duivelsspiegel wegveegde. Hij keek haar aan en barstte toen in tranen uit. Hij hilde zo hard dat de glassplinter uit zijn oog spoelde; toen herkende hij haar, en riep uit: “Gerda! lieve kleine Gerda! Waar ben je zo lang geweest? En waar ben ik zelf geweest?”

En hij keek rond.

“Wat is het hier koud! Wat uitgestrekt en leeg!” en hij stortte zich op Gerda, en zij lachte en hilde van van vreugde. Het was zo’n gelukkig moment dat zelfs de stukken ijs om hen heen dansten van vreugde, en toen ze moe werden en weer gingen liggen vormden ze zichzelf tot de letters waarvan de Sneeuwkoningin had gezegd dat hij ze moest spellen om zichzelf meester te worden en de hele wereld te bezitten en een nieuw paar schaatsen.

En Gerda kuste zijn wangen en zij werden rood als rozen; ze kuste zijn ogen en ze fonkelden als de hare; ze kuste zijn handen en voeten en hij werd warm, en gloeide. De Sneeuwkoningin kon nu thuis komen; zijn bevrijding: het woord ‘Liefde’ stond geschreven in flonkerend ijs. ...”

Referenties

- Eliot, T.S. (2000), *The Waste Land*, W. W. Norton & Company, New York.
 Fromm, E. (1952), *De angst voor vrijheid*, Bijleveld, Utrecht, oorspronkelijke titel *Escape from Freedom*, Rinehart & Co., New York, 1941.

Hughes, R. (1994), *De klaagcultuur*, Balans, Amsterdam, oorspronkelijke titel *The Culture of Complaint*, Oxford University Press, Oxford, 1993.

Schellenbaum, P. (1985), *Nee in de liefde*, Lemniscaat B.V., Rotterdam, oorspronkelijke titel *Das Nein in der Liebe*, Kreuz Verlag, Stuttgart, 1984.

NETWORK EXPERIMENTS

Michael Kosfeld*

University of Zürich

Networks play an important role in the research agenda of Pieter Ruys. When I arrived in Tilburg as a PhD student, the first paper Pieter gave me to read was his paper on public utilities and network industries (Ruys 1997). Later, together with René van den Brink, he developed models to analyze the internal mechanisms underlying the functioning of firms and markets (Ruys and van den Brink 1999, 2000). For Pieter the notion of a network provides a useful tool for opening the economists' "black boxes" of market allocation and firm organization, a fundamental aim he has always been interested in. This paper is a contribution towards reaching this aim: to open the black box a little bit further.

In recent years, economic research on networks has increased tremendously. By now there exists substantial evidence emphasizing the important role of networks on social and economic outcomes. Famous examples have pointed out network effects on job search (Holzer 1987, Montgomery 1991), trade (Lazerson 1993, Nishiguchi 1994), the granting of credit (McMillan and Woodruff 1999), mutual insurance (Fafchamps and Lund 2001) and welfare participation (Bertrand, Luttmer, and Mullainathan 2000).¹ While theoretical research on economic networks has received a lot of interest as well (see references in the subsequent sections), until very recently no experimental work on networks in economics existed. The number of network experiments is still very small but the literature is starting to grow. The aim of this paper is to provide an overview of existing experimental work and to suggest paths for interesting future research in this area.

Laboratory experiments present a useful and powerful technique to analyze economic questions. The main advantage of experiments lies in the ability to control variables (such as, e.g., costs, benefits, information, and timing) that possibly influence individual and aggregate behavior, something which is very hard and sometimes even impossible to achieve in the field. Together with theoretical models, in particular models based on game theory that provide the language for an exact formulation of hypotheses, controlled experiments are a key element for economics to become an empirical science. And, as one of the major contributors to economic theory writes in a recent paper, "moving from arm-chair theorizing to controlled laboratory experiments may be as important a step in the development of eco-

**Address:* Institute for Empirical Research in Economics, University of Zürich, Blümlisalpstrasse 10, CH-8006 Zürich, Switzerland; *Email:* kosfeld@iew.unizh.ch

¹Outside economics the role of networks and social structure has earlier been emphasized by anthropologists (e.g., Lévi-Strauss 1963) and sociologists (e.g., Granovetter 1974).

nomics as it once was for the natural sciences to move from Aristotelian scholastic speculation to modern empirical science” (Weibull 2001).

Before experimental economists discovered networks, other social scientists already started to investigate network effects in various experimental settings. Perhaps the earliest network experiments are the “MIT experiments” of the social psychologist Alex Bavelas and his colleagues in the early 50’s (Bavelas 1950, Leavitt 1951). In these experiments a group of individuals is assigned a problem to be solved. Typically, for this problem each individual receives a card showing different symbols. Individuals in a group have only one symbol in common and the problem is to discover that common symbol. Individuals can communicate by passing written messages to each other. Communication can only flow along a exogenously imposed network. Bavelas and his colleagues consider four different communication networks: the chain, the circle, the star, and the “Y”. They find that groups who communicate via the star or the “Y” are fastest in solving the given problem. Moreover, they use the least number of messages and also make the fewest errors. On the other hand, individuals in the circle and the chain report highest average satisfaction during the experiment. Only the individual that acts as the center in the star reports higher satisfaction. As a consequence of these experiments Bavelas *et al.* call attention to the role of structural centrality for group efficiency. In subsequent years further studies have been conducted to investigate the effect of centrality on communication and organization networks. See Shaw (1964) for a critical review. Freeman (1979, 1980) clarifies the terminology by defining and testing the role of three different concepts of structural centrality.

Clearly, sociologists have long been interested in the role of networks, as well. Early experiments were conducted by Cook and Emerson (1978) and Cook, Emerson, Gilmore, and Yamagishi (1983), who are interested in the relation of power and social structure in exchange networks. See also the related theoretical work of Markovsky, Willer and Patton (1988). Recent articles by Bienenstock and Bonacich (1993, 1997) incorporate game theoretic concepts into the discussion.

In the subsequent sections I give an overview of present experimental work in economics focusing on networks. The literature is still very small. I first discuss experiments considering coordination in networks (Keser, Ehrhart, and Berninghaus 1998, Berninghaus, Ehrhart, and Keser 2002, Corbae and Duffy 2002, Cassar 2002). Subsequent sections address experimental work on cooperation (Kirchkamp and Nagel 2001, Cassar 2002, Riedl and Ule 2002), buyer-seller networks (Charness, Corominas-Bosch, and Frechette 2001) and network formation (Deck and Johnson 2002, Falk and Kosfeld 2003, Callander and Plott 2003, Vanin 2002).

1 Coordination networks

Theory

Seminal work by Kandori, Mailath, and Rob (1993) and Young (1993) has triggered a lot of interest in the question of equilibrium selection in coordination games. Important research concerns the impact of different network structures on equilibrium selection and, if players can choose their network partners themselves, whether

players will form networks that lead to play of the efficient Nash equilibrium in the coordination game.

Ellison (1993) and Morris (2000) analyze the role of local interaction networks for the spread of particular strategies in 2×2 coordination games, showing, e.g., how play converges to the risk-dominant equilibrium if players are located on a circle and interact with their two nearest neighbors. Similarly, convergence to the risk-dominant equilibrium is proven by Blume (1993) and Kosfeld (2002) for a population of players located on a d -dimensional lattice.

In contrast, Ely (2002) and Bhaskar and Vega-Redondo (2002) show that once players are allowed to choose their partners themselves the situation looks very different. They introduce a number of locations where players can meet and play the coordination game with each other. Thus, at any time, players choose both a location and a strategy in the game. Under these conditions it is shown that risk dominance loses its selection force and that the population is most likely to coordinate on the efficient equilibrium. The reason is intuitive. Since players can freely choose their interaction partners, they are able to find partners that play the efficient equilibrium strategy and at the same time can avoid players that play the inefficient strategy. That the latter condition, i.e., the ability to avoid bad matches, is crucial, is emphasized by Mailath, Samuelson, and Shaked (2001), as well.

Since migration is cost free in these models, an interesting and important question is what equilibrium will be selected when migration or the formation of links is costly. Goyal and Vega-Redondo (2000) present a first theoretical approach in this direction. They find that the cost of a link to another player plays a decisive — and somewhat counterintuitive — role for the selection of equilibrium. Only if costs are high enough, players coordinate on the efficient Nash equilibrium. If, however, costs are low the risk-dominant equilibrium is selected. Droste, Gilles, and Johnson (2000) consider a similar stochastic learning model. In their set-up players are located on a circle where they can form links to other players. Links to more distant players are assumed to be more costly. It is shown that, while in the medium run coexistence of both equilibrium strategies is possible, in the long run the risk-dominant equilibrium is selected.

Experiments

There exists a large experimental literature on coordination games. See, for example, Ochs (1995) for a review. The first experiment that considers the role of networks for coordination games, along the lines discussed above, is presented by Keser, Ehrhart, and Berninghaus (1998). In this experiment, the authors study the impact of local interaction as analyzed by Ellison (1993) and Morris (2000). They implement two different treatments. In the local-interaction treatment groups of eight players each interact around a circle. Players that are located next to each other play a coordination game for 20 periods. In this treatment Keser *et al.* find that play converges to the risk-dominant Nash equilibrium, as it is predicted by the theory. In contrast, once players interact in isolated groups of size three, which equals the neighborhood size in the local-interaction treatment, they find that play converges to the efficient Nash equilibrium. This finding is compatible with a sim-

ilar experimental result of Van Huyck, Battalio, and Beil (1990) on coordination in small groups. In their second paper, Berninghaus, Ehrhart, and Keser (2002) put their result in a more general framework. In particular, they (i) modify the payoff function in the network-coordination-game, reducing the riskiness of the efficient Nash equilibrium, and (ii) vary the neighborhood structure that determines the way in which players locally interact. With regard to (i) they find that if the efficient Nash equilibrium becomes less risky, also populations that locally interact on a circle converge to efficient play in most of the cases. Varying the structure of the neighborhood, however, generates a converse effect. Precisely, in order to address point (ii) Berninghaus *et al.* compare two treatments, where in both treatments each player locally interacts with his four nearest neighbors. In one treatment players are located on a circle while in the second treatment players are located on a two-dimensional lattice.² Thus, the size of a player's neighborhood is kept constant across treatments and only the neighborhood structure differs. Given this experimental design, the authors find that play is more likely to converge to the risk-dominant equilibrium if players interact on the lattice than if they interact on the circle. This result is particularly interesting since subjects had exactly the same instructions in both treatments, i.e., they were not informed about the precise neighborhood structure of the population. One possible explanation the authors offer is the following: in the lattice treatment individual play is observed to be more changing than in the circle treatment. In consequence, risk dominance as an individual motive has more power in the lattice treatment than in the circle treatment. These findings offer an exciting starting point for further experimental (and theoretical) studies.

The local-interaction hypothesis is also tested in a recent preliminary experiment by Corbae and Duffy (2002). They look at groups of four players that are endowed with a network structure of either global interaction (the complete network), local interaction (the circle), or "marriage interaction". In the latter case the population is split into two isolated pairs of players, who interact with each other. Independent of the network, in each group subjects play ten periods of a coordination game, where the efficient Nash equilibrium is also risk-dominant. With the exception of one group, play is observed to converge to the efficient equilibrium. After these ten periods subjects face a new game that differs from the previous game only with respect to the off-equilibrium payoff of the efficient Nash equilibrium strategy, which renders the inefficient Nash equilibrium risk-dominant. This game is played for ten periods. The hypothesis is that if no player is forced to play the inefficient equilibrium strategy, players will keep coordinating on the efficient equilibrium irrespective of the underlying interaction structure. Given the observations in the experiment (two groups for each network), this hypothesis is confirmed. Yet, if a single player in the group is forced to play the inefficient strategy, the hypothesis is that convergence to the inefficient equilibrium will be observed in the local- and in the marriage-interaction treatment but less so in the global-interaction treatment. Again, this hypothesis is confirmed by the data.

Corbae and Duffy address also the question what network will be formed if players can choose the network themselves. They again consider groups of four

²In fact, players interact on a torus, which, similar to the circle, has no boundaries.

subjects who first interact on an exogenously imposed network structure for an initial phase of five periods. The network structure is either a global-, a local-, or a marriage-interaction network. Subjects play the same coordination game as in the second part of the exogenous-network treatment described before. After this initial phase, subjects can freely decide with whom of the other three subjects they want to interact in the subsequent periods. If two subjects mutually agree to interact, a link is formed between the two subjects.³ For the next five periods all subjects that are directly linked play the coordination game. This procedure is repeated four times. So far, the analysis in the paper is still preliminary. Basically, no stable network structure seems to emerge. Perhaps the main finding is that groups starting with a network of marriage interaction show a high tendency to form the same network also in later periods. Clearly, further investigation of the data but also more experimental evidence is needed.

Cassar (2002) compares convergence to equilibrium across three different network structures: a local interaction network, a random network and a “small-world” network.⁴ Small-world networks are obtained by starting from a circle and rewiring each link with probability p . A random network is a small-world network that is obtained by rewiring each link with probability 1. Intuitively, for intermediate p , small-world networks possess the nice feature of having a substantial degree of clustering, i.e. a large overlap of neighborhoods, and yet only short paths connecting any two individuals in the network. In contrast, random networks have short connecting paths and low clustering, whereas the circle has high clustering but also long connecting paths. Generally, high clustering implies that interaction in the network resembles interaction in a closed group. On the other hand, short connecting paths suggest that contagious behavior can spread more easily.

In the experiment, Cassar considers groups of size 18, each group being connected according to one of the three network structures. In every group, subjects who are connected with each other play around 80 periods of a coordination game, where risk dominance and efficiency of equilibria conflict. Results show that in the small-world network subjects almost always converge to the efficient Nash equilibrium, while convergence is less likely (although still above 60 percent) in the other networks. Moreover, convergence to the efficient equilibrium is fastest in the small-world network, as well. These findings are consistent with the network effects suggested above.

There are still many opportunities for experimental research on coordination networks. For example, in order to test and better understand the importance of endogenous network formation on equilibrium selection it would be interesting to take the models of Ely (2002) and Bhaskar and Vega-Redondo (2002) into the laboratory. Similarly, the predictions of Goyal and Vega-Redondo (2000) and Droste, Gilles, and Johnson (2000) offer valuable possibilities for empirical research. As the focus has been on theoretical modeling so far, there exists a wonderful collection of game theoretic predictions, but only few of them have been tested.

³This assumption is based on the notion of pairwise stability defined by Jackson and Wolinsky (1996). See the section on network formation in this paper.

⁴Small-world networks are introduced by Watts and Strogatz (1998), see also Watts (1999).

2 Cooperation networks

Theory

Eshel, Samuelson and Shaked (1998) show that cooperation in the prisoners' dilemma game can survive if players in a population locally interact with each other and adaptation is driven by imitation of successful behavior. The crucial effect of local interaction in this model is that it allows cooperative players to cluster together. The idea is the following. Since the positive externalities from cooperation are locally restricted, the interaction network reduces the possibility for other (more distant) players to exploit cooperation. In consequence, cooperators being surrounded by other cooperators can earn higher payoffs than defectors who are primarily surrounded by other defectors. Together with imitation this gives cooperation a chance to survive. Related work include Nowak and May (1992) and Kirchkamp (2000).

Huck and Kosfeld (2001) use a similar local interaction framework to show how punishment of defection can survive and sustain cooperation in the prisoners' dilemma game. In their model some players in the population are so-called "social controllers" who punish defection by any neighbor. The effect of punishment is that first, it may deter defection, and second, it may induce other players to become social controllers, as well. The latter effect is called "socialization". Huck and Kosfeld show that if the socialization effect is strong enough, a condition which turns out to be relatively easy to be fulfilled, punishers can survive in the population. A crucial driving force behind this result is the fact that neighborhoods overlap in the given network structure. This allows punishment to diffuse in the population, similar to a contagion process (cf. Morris 2000). Finally, Huck and Kosfeld show that if punishers survive, they can generate a large degree of cooperation.

A recent theoretical paper considering endogenous cooperation networks is Vega-Redondo (2002), who studies the formation of networks among players bilaterally involved in infinitely repeated prisoners' dilemma games. In addition to specifying which pairs of players in the population play the game, in the model of Vega-Redondo a network also determines how strategic information diffuses among the players and how cooperation opportunities of the players are found. Assuming that payoffs in the prisoners' dilemma game fluctuate over time, Vega-Redondo analyzes the notion of pairwise-stable cooperation networks, where, intuitively, two players are directly connected with each other only if both players have an incentive to use the connection for cooperation in the prisoners' dilemma game. The main results are that players can sustain a dense social network, i.e., a network with sufficiently many individual connections, only if payoff volatility is not too high. Moreover, higher payoff volatility increases the cohesiveness of the network, i.e., the average distance between two players in the network declines as payoffs fluctuate more strongly.

Experiments

Although the experimental literature on cooperation in prisoners' dilemma games is extensive, there are only few recent papers considering the role of networks on

cooperation. These are Kirchkamp and Nagel (2001), Cassar (2002) and Riedl and Ule (2002).

Kirchkamp and Nagel (2001) are interested in the prediction of Eshel *et al.* (1998) showing that cooperation can be sustained by local interaction and imitation. They consider a similar experimental design as Keser *et al.* (1998), which consists of two treatments. In the first treatment 18 subjects interact around a circle, each subject playing the prisoners' dilemma game with his four nearest neighbors, i.e., two neighbors to the left and two neighbors to the right. In the second treatment subjects interact in isolated groups of size five. Subjects play 80 periods in every treatment. In each period subjects observe the strategies and payoffs of each of their interaction partners. They must use the same strategy against all partners.

Different from the theoretical prediction, which says that there should be more cooperation in the local interaction treatment as subjects can learn from neighbors, Kirchkamp and Nagel find that cooperation rates are higher if subjects interact in isolated groups than if they locally interact on a circle. While initially, cooperation rates are close to 30 percent in both treatments, they decline to below five percent in the local interaction treatment. In contrast, cooperation rates stay at about the same level in the group treatment. This result clearly contradicts the theoretical prediction.

Cassar (2002) reports a similar decline in cooperation for the local interaction network, the small-world network, and the random network. She finds no major differences between these networks.

A possible explanation for the instability of cooperation in these networks is that subjects do not learn the way it is assumed by models of imitation. Indeed, Kirchkamp and Nagel find that in both treatments learning is not driven by imitation of neighbors' successful strategies, but mainly by positive reinforcement of one's own successful strategies. Hence, the main mechanism which makes cooperation survive on the circle but not in isolated groups does not seem to be at work in the laboratory. This alone, however, does not explain why subjects cooperate more in the group treatment, since the success, and hence the reinforcement of one's own strategy to cooperate should be the same (namely, zero) in both treatments. Yet, the surprising and interesting finding is that the strategy to cooperate increases one's own payoff in groups but not on the circle. Subjects seem to interact more in a reciprocal manner, i.e. in particular they cooperate more if others cooperate, when they interact in isolated groups, than if they interact in locally overlapping groups of the same size. So far it remains unclear what precisely generates this result, whether it is some sort of strategic reasoning that plays a role, as suggested by Kirchkamp and Nagel, or other more psychological factors. It thus presents an interesting problem for future research.

The experiment of Riedl and Ule (2002) considers a different question. They are interested in endogenous network formation when players play a repeated prisoners' dilemma game, as, e.g., studied in the model of Vega-Redondo (2002). Riedl and Ule take groups of size six. In the control treatment of their experiment the network structure is exogenously fixed to the complete network, i.e. each subject plays the prisoners' dilemma game against every other group member. In the other treatments subjects can themselves decide whether to form links with other group members.

A link is established if both parties agree to form a link (pairwise stability, cf. Jackson and Wolinsky 1996). Links are cost free. Each pair that is linked plays the prisoners' dilemma game. If some party rejects a link both parties earn an outside-option payoff. Subjects have to choose the same strategy against all partners. Every treatment consists of 60 periods of play.

The analysis presented in Riedl and Ule (2002) is still preliminary. However, some interesting results can already be summarized. First, cooperation rates are significantly higher in the endogenous-network treatments compared to the exogenous-network treatment. While initially, cooperation rates are around and beyond fifty percent in all treatments, cooperation declines if the network is fixed, whereas cooperation remains stable almost until the end if subjects can choose their partners themselves. Only in the final five periods cooperation decreases also in the endogenous-network treatments. Second, cooperation rates are highest if the outside-option payoff lies between the Nash-equilibrium and the cooperation payoff in the prisoners' dilemma game and subjects can observe the strategies of all other subjects in the group.⁵ Assuming that some of the subjects are reciprocators, who cooperate if others cooperate as well, this suggests that the value of the outside option might serve as a signaling device for cooperative play. Third, cooperators are more likely to propose links to subjects who cooperated in the previous period than to previous-period defectors. This holds, even if the outside-option payoff is lower than the Nash-equilibrium payoff, i.e. when exclusion of defectors is costly.

The results of Riedl and Ule clearly show that network structure plays an important role for cooperation in prisoners' dilemma like situations. In particular, the possibility to exclude players who defect, even if this is costly, turns out to be a powerful instrument having striking effects on behavior. It is interesting to compare this finding to a recent experiment by Brown, Falk, and Fehr (2002), who do not study network formation explicitly, but allow subjects to form bilateral relations in a two-person incomplete contract setting. They also find that the threat to terminate a relation represents a powerful discipline device, which induces partners to cooperate and consequently enhances efficiency.

3 Buyer-seller networks

Theory

Buyer-seller networks represent an area where both theoretical and experimental research in economics has been initiated recently.

Kranton and Minehart (2001) are interested in the individual motives of buyers and sellers to form particular network structures, as documented, e.g., in the Japanese electronics industry (Nishigushi 1994) and the Italian garment industry (Lazerson 1993). In particular, they ask what may lead buyers and sellers to establish links to multiple trading partners, and whether these networks can be expected to be efficient.

⁵Hauk and Nagel (2001) find a similar increase in cooperation rates in a related prisoners' dilemma experiment with the possibility of partner choice. In this experiment, however, no networks are considered.

To answer these questions they consider a number of buyers and sellers, where each seller has an indivisible object for sale and buyers have i.i.d. random valuations for the object. A buyer can purchase from a seller if and only if the two are linked. Links are established by the buyers, who face costs for these links.⁶ Transactions and prices are determined by an English, i.e., ascending-bid, auction. Buyers drop out of the bidding as the price exceeds their valuation. This process continues until demand equals supply. Kranton and Minehart show that competition generates an efficient allocation of goods in the network. Furthermore, prices reflect the link pattern in the sense that a buyer's profit equals the marginal social value of his participation in the network. From this it follows that efficient network structures are always an equilibrium outcome.

The model of Kranton and Minehart (2001) emphasizes two reasons why buyer-seller networks may emerge, one economic, the other strategic. First, networks may allow buyers and sellers to pool uncertainty in demand, which in the present model is caused by buyers' random valuations. Second, multiple links of a trader can enhance the competitive position of this trader.

Network effects on competition are also addressed in the network model of Corominas-Bosch (1999). However, different from the model of Kranton and Minehart (2001), in this model prices are determined by a bargaining process rather than an English auction, and a buyer's valuation of the seller's good is certain. Again, a link between two trading partners is necessary for possible transaction. Consequently, if an individual has several links, he has several possible partners he can trade with. Thus, the network structure directly determines the bargaining power of individual players.

Bargaining follows a variation of the Rubinstein alternating-offer protocol. In the first period each seller calls out a price. Buyers then simultaneously choose to accept at most one of the prices offered by a seller to whom they are linked, with ties broken randomly. If a buyer and a seller trade their links are removed from the network. In the next period the situation reverses and buyers call out prices, which are then accepted or rejected by the sellers connected to them. This process repeats itself until all remaining buyers and sellers are not linked to each other. Future periods are discounted according to a common discount factor.

Corominas-Bosch shows that depending on the given network structure the subgame-perfect equilibrium of the bargaining game has the following properties. If the network is "competitive" the short side of the market receives all surplus. For example, if there is only one buyer who is linked to two sellers, competition between the sellers will reduce the price such that the buyer extracts all surplus. This is reversed for a single seller who is linked to two buyers. If, however, the network is "even", i.e., the number of buyers and sellers linked to each other is the same, traders split the surplus evenly. In the case of a single seller linked to a single buyer this result corresponds exactly to the Rubinstein bargaining model. Corominas-Bosch then shows that any network can be decomposed into a union of subnetworks that are either competitive or even, plus some extra links. Furthermore, subgame-perfect equilibrium outcomes of the bargaining game are such that all the surplus

⁶Jackson (2001) generalizes the model such that links are costly to buyers and sellers.

is given to the short side of the market in every competitive subnetwork, while the surplus is divided evenly in every even subnetwork.

Experiments

The model of Corominas-Bosh (1999) has recently been tested by Charness, Corominas-Bosch, and Frechette (2001), who are particularly interested in the predictive power of the subgame-perfect equilibrium outcome in competitive versus even buyer-seller networks. In the initial phase of their experiment subjects interact on either one of two different network structures: a three-player network, where a single buyer is connected to two sellers, or a four-player network, where two buyers and two sellers are each connected to each other. The three-player network is competitive whereas the four-player network is even. In both networks subjects are engaged in alternating bargaining over five to six rounds. Sellers start offering a proposal to divide 2500 with any of their linked buyers, who then simultaneously decide to accept at most one of the proposals. If a transaction is made, the corresponding pair of a buyer and a seller is removed from the network. The game proceeds to the second round if any links remain. In the second round buyers offer a division of 2400 to all of their linked sellers, who choose to accept or not to accept any of these offers. If needed, a third round is implemented, etc. A coin is flipped to determine whether bargaining ends after five or six rounds. All unmatched subjects receive 200. This bargaining protocol is repeated for four periods with subjects' roles reversed after each period.

After the fourth period the second phase of the experiment begins, in which the two separate networks are merged to obtain a single seven-player network. This is done by adding a link either from the short (i.e., buyer) side or from the long (i.e., seller) side of the competitive network to the corresponding other side of the even network, respectively. If the two networks are connected via the short side of the competitive network, theory predicts that equilibrium-outcomes remain the same for each player. If, however, the two networks are connected via the long side of the competitive network, the new seven-player network becomes competitive. Consequently, equilibrium-outcome predictions change for all subjects from the former even subnetwork. The second phase of the experiment consists of six periods of bargaining, each containing again up to six alternating rounds.

Charness *et al.* find that while exact equilibrium-point predictions fail, qualitatively, the behavior in the experiment is consistent with theoretical predictions. For example, seller's payoffs are lower in the competitive three-player network than in the even four-player network. Hence, buyers seem to successfully make use of their bargaining power. Furthermore, the way in which the two subnetworks are connected with each other has a significant effect on what bargaining outcomes are realized in subsequent periods. Theory predicts that payoffs between buyers and sellers in the four-player subnetwork should diverge if the large network becomes competitive, while they should be roughly the same if the subnetwork remains even. This prediction is nicely confirmed by the data.

The fact that any precise equilibrium prediction fails, may be due to insufficient time for subjects to learn optimal behavior. However, as Charness *et al.* show, the

willingness to offer or accept certain shares is strongly affected by allocations in the past. Subjects learn from each other and develop a social norm for appropriate bargaining outcomes. This suggests that learning may not necessarily lead into the direction predicted by the theory. Finally, that bargaining outcomes tend to be less extreme than predicted by standard economic theory is consistent with extensive experimental research on ultimatum and dictator games (cf. Roth 1995, Camerer and Thaler 1995). In these games observations can be explained by assuming that players behavior is driven by considerations of fairness and equity (e.g., Fehr and Schmidt 1999). It seems plausible that fairness considerations may also play a role in the present experiment of a buyer-seller network. However, more experimental research is certainly needed.

4 Network formation

Theory

Next to the effects on coordination, cooperation, and bargaining, one of the most important questions with respect to research on economic networks is, of course, the question how networks emerge. In recent years, several theoretical approaches have been proposed to address this question, using techniques from cooperative and non-cooperative game theory.⁷

The paper of Myerson (1977) is, probably, one of the first important contributions to this area. Myerson analyzes a cooperative game that is enriched by a network structure describing the possibilities for communication or cooperation among different players. Individuals can act as a coalition if and only if they are connected through links in the network. While this idea constitutes an important step forward, it leaves several issues unsolved. Because the value function is still defined on coalitions and not on the network directly, the theory does not distinguish between different networks that connect the same players but differ in the way these players are connected. In consequence, many interesting details of the network formation process, for example costs and benefits of particular links, can not be analyzed in the model.

Jackson and Wolinsky (1996) follow a different approach. They consider value functions that are defined on networks directly. The main issue they address is the conflict between efficiency, i.e., value maximization, and stability. With regard to the latter they analyze the notion of pairwise stability, which assumes that links are formed if and only if both players that are connected by that link agree to form the link. On the other hand, links are severed if any of the two individuals decides to do so.

A particular network model they consider is the so-called connections model, where individuals receive benefits from being connected to other individuals and bear costs for maintaining direct links. Jackson and Wolinsky show that the set of efficient networks reduces to only three different types of networks: the complete network if costs are low, the star network if costs are intermediate, and the empty

⁷See Jackson (2003) for a detailed survey of the theoretical literature on network formation.

network if costs are high. While the complete and the empty network are pairwise stable if they are efficient, the star network may fail to be pairwise stable. As Jackson and Wolinsky show, this conflict between efficiency on the one hand and pairwise stability on the other, is no unique feature of the connections model but extends to more general network settings, as well.

The work of Myerson (1977) and Jackson and Wolinsky (1996) has attracted much interest in economic models of network formation. Subsequent models include Aumann and Myerson (1988), Dutta and Mutuswami (1997), Dutta, van den Nouweland, and Tijs (1998), Slikker and van den Nouweland (2000) and Johnson and Gilles (2000). All these models study the formation of networks in a static setting. Watts (2001) departs from this tradition and analyzes the connections model in a dynamic setting, where individuals meet over time and decide to form or sever links between each other. Similarly, Jackson and Watts (2001) consider the evolution of more general network models.

Somewhat parallel to the above literature that has clear origins in cooperative game theory, Bala and Goyal (2000a,b) develop models of network formation that use tools from non-cooperative game theory. Rather than considering pairwise stability, Bala and Goyal assume that individuals can form and sever links unilaterally, i.e., in particular no mutual consent is needed to form a link between two individuals. Clearly, this assumption changes the incentives of the players, hence the analysis in Bala and Goyal (2000a,b) differs substantially from the analysis in the models mentioned above. A central implication of unilateral link formation is that with regard to stability it leads to the concept of Nash equilibrium. Papers that follow a similar methodological approach include Goyal and Moraga-Gonzalez (2001), Goyal and Joshi (2002), Haller and Sarangi (2001) and Sarangi, Kannan, and Ray (2003).

The main idea of the network model in Bala and Goyal (2000a) is similar to the connections model of Jackson and Wolinsky (1996): players earn benefits from being connected to other players and bear costs for maintaining direct links. Benefits are regarded as resulting from valuable, non-rival information that flows through the network. Bala and Goyal distinguish between two different scenarios of information flow. In the first scenario (the 1-way flow model) information flows only to the player who maintains the link. In the second scenario (the 2-way flow model) information flows both ways. Independent of the information flow, Bala and Goyal assume that players simultaneously decide with whom to form a direct link, a link being costly to the individual who forms it.

Assuming that information flows through the network with no decay, Bala and Goyal prove that Nash equilibria of the network-formation game look as follows.⁸ In the 1-way flow model a Nash equilibrium is either the empty network, where no player maintains any connection to any other player, or minimally connected, i.e. it has a unique component that splits if one link is severed. Analogously, in the 2-way flow model a Nash equilibrium is either the empty network or minimally 2-way connected, i.e., it has a unique component, no cycle, and no two individuals both maintain a link with each other. Intuitively, in both models a network is Nash if (i)

⁸Bala and Goyal prove also results for the general case, where they allow for decay. However, results are more clear-cut if no decay exists.

either none or all players are connected and (ii) no redundant links are maintained.

Depending on the number of players the number of Nash (equilibrium) networks can be quite large. Therefore, a reasonable refinement to consider is the notion of strict Nash equilibrium, where each player plays his unique best response to the strategy profile of the other players. Surprisingly, the set of strict Nash networks is much more restrictive. In the 1-way flow model Bala and Goyal show that the only strict Nash networks are the empty network and the circle (or, as Bala and Goyal call it, the wheel). In the 2-way flow model only the empty network and the center-sponsored star are strict Nash networks. The center-sponsored star is the network, where one individual (the center) maintains a direct link to every other individual, and no other individual maintains any link.

Both the circle and the center-sponsored star are shown to be efficient networks, where efficiency is defined in terms of maximizing the sum of players' payoffs. Thus, contrary to the model of Jackson and Wolinsky (1996), in the model of Bala and Goyal (2000a) no conflict between efficiency and stability exists. This suggests that the circle and the center-sponsored star may serve as a powerful prediction for network formation.

Experiments

Recent economic experiments that consider network formation are Deck and Johnson (2002), Callander and Plott (2003), and Falk and Kosfeld (2003). Vanin (2002) presents a pilot study for the model of Jackson and Wolinsky (1996). In his experiment three groups of four subjects each collectively bargain about what network to form. Links between two subjects are established based on pairwise stability, i.e. bilateral agreement is required. Each group forms a network in three different scenarios: the connections model with and without side payments and the co-author model.⁹ Vanin finds that groups tend to form pairwise unstable but efficient networks both in the co-author model and, two out of three times, also in the connections model with side payments. In the connections model without side payments groups form networks that are inefficient but equalize payoffs among the subjects.

The experimental study of Deck and Johnson (2002) is inspired by the network-formation model of Johnson and Gilles (2000), which introduces a spatial cost topology in the connections model of Jackson and Wolinsky (1996). Without going into details, in this model players are located on a line and the cost for a direct connection between two players monotonically increases with the distance between the two players. Deck and Johnson are interested in a comparison of three different institutions for network formation in this model, and consequently implement one treatment for each institution. In the first institution, called "Split", players select those direct links, for which they are willing to pay exactly half of the cost. If both players agree to pay half of the cost of their connecting link, the link is formed. The second institution, "Primary", allows players to bid between zero and the total cost

⁹The co-author model assumes that an additional link generates a negative externality on individuals already connected. Jackson and Wolinsky (1996) show that pairwise stable networks are generally over-connected and hence inefficient in this model.

of a link for all their direct links. Finally, in the institution “Secondary” players can bid for all possible links. In particular, they can bid for a link between any two of the other players. Under the latter institutions, a link is formed if and only if the total sum of players’ bids for that link exceeds the cost of the link. Deck and Johnson use the notion of Nash equilibrium as a stability concept for a network.

In each treatment of the experiment subjects consecutively interact in three different payoff environments, two of them involving groups of five players. In the third environment players interact in pairs. In the following, I focus on the five player environments only. A particular feature of the experimental design of Deck and Johnson is that the situation is described to the subjects as a decision task faced by managers of a train station. Precisely, in the experiment a subject is put into the role of a station manager, who has to decide (i.e., to bid) on the connections between different stations. Just like in the connections model, the value of a network to a station manager is determined by the minimum number of stops a fictitious passenger would have to make on his trips from the manager’s station to the other stations. In the first payoff environment, the parametrization is such that the unique efficient network requires each player to be directly connected to his nearest and to his second nearest neighbor(s). This network is supportable by a Nash equilibrium under all three institutions. In the second payoff environment, parameters are chosen such that the chain is the unique efficient network, where each player is connected to his nearest neighbor only. The chain is supportable by a Nash equilibrium only under the Primary and the Secondary institution, but not under the Split institution.

Subjects play 15 periods of the first environment, following 10 periods of the second environment. The results reported in the paper are based on the last two thirds of all periods in each environment. Deck and Johnson find that in the first environment the Primary institution performs best, yielding an average level of efficiency of 89 percent while Split and Secondary, on average, achieve only an efficiency of 83 percent and 81 percent, respectively. The main reason the authors give is the following: while in all treatments subjects form too many of the long links, connecting stations that are far away from each other, in the Primary treatment subjects are more successful in forming the necessary short links. In the second environment all three institutions show a similarly bad performance. No group achieves a positive surplus under any institution. Note that in this environment, the chain is the unique efficient network, where a significant percentage of the network’s value is provided by indirect links. Since subjects do not succeed in coordinating on this network, the authors conclude that network formation driven by individual decision-making is a difficult process that is likely to produce inefficient outcomes. However, in the environment the authors consider the individual strategy to form a chain is a high-risk strategy that is likely to generate a negative payoff to each player if players do not successfully coordinate. It remains to be shown whether it is the individual decision-making itself or rather the riskiness of the efficient equilibrium-strategy that leads to the coordination failure. In fact, the studies of Callander and Plott (2003) and Falk and Kosfeld (2003) show that if this riskiness is reduced, individuals are well able to coordinate on an efficient network. These studies are discussed in the remainder of the paper.

Callander and Plott (2003) report on — as they call it — an “exploratory” experiment, in which they study the evolution of information networks under various treatment conditions. In each treatment they consider groups of six players. The first treatment involves between 10 and 20 periods of a 1-way flow network-formation game à la Bala and Goyal (2000a), using a random stopping rule to determine the end of the experiment. In this experiment, subjects sit together in a room. In each period, after each subject has recorded his direct connections a physical process is used to calculate the payoffs in the resulting network. Precisely, different physical signs are placed in front of each subject that represent individual links and are used to determine subjects’ profits. Costs and benefits in this treatment are such that the unique strict Nash network prediction is the circle, which is also efficient. In the second treatment that is studied by Callander and Plott, individual decisions are made via the computer. Moreover, whereas in the first treatment the decision making of the subjects is simultaneous, in the second treatment decisions are made continuously over two minute rounds, where choices can be adjusted repeatedly in real time. Subjects are continuously updated about the choices of the other group members and each choice adjustment is costly. Each session lasts between 15 and 20 rounds, using again a random stopping rule to determine the end of a session. Different games are played in this treatment. At the beginning, each session starts with a 1-way flow network-formation game with benefits being slightly different from the game played in the first treatment but the theoretical prediction being the same (i.e., the circle). In case subjects form the same network for three consecutive periods the game is changed in such a way that from then on direct links to nearest neighbors are twice as costly than links to other subjects.¹⁰ This modification has the effect that only particular circles, where no nearest neighbors are directly connected, are efficient. If under the new setting subjects form again the same network for three consecutive periods, the game is changed yet another time. Precisely, one subject is selected to/from which direct links are free. This renders the circle inefficient and predicts a star network as the unique efficient (and strict Nash) network. Subjects are not informed about the potential change of the games at the beginning of the experiment.

The main findings of Callander and Plott are as follows: first, in both treatments networks often converge to Nash equilibrium. If convergence is observed, in all cases except for one group subjects form the efficient and strict Nash network, which is the circle.¹¹ Second, decision making with real-time choice adjustment facilitates the coordination on a Nash network. While in the first treatment two out of five groups eventually converge to Nash, in the second treatment six out of seven groups eventually form a Nash network. On the individual level, Callander and Plott reject the hypothesis that subjects adjust their strategies based on a best-response rule with additional error term. Instead, the authors conclude that many subjects seem to exhibit so-called “simple strategic behavior”, meaning that subjects form exactly one direct link that is part of a focal (e.g., a clockwise) circle network.

The experiment of Callander and Plott (2003) is inspired by the network for-

¹⁰This is the opposite of what is assumed in the model of Johnson and Gilles (2000).

¹¹Only one group participated in the final parameter setting of the second treatment. No convergence was observed in this case.

mation model of Bala and Goyal (2003a). The authors, however, consider only the 1-way flow model and in this model only treatments, where the circle is the unique efficient strict Nash network. Falk and Kosfeld (2003) present a more detailed analysis of the Bala-Goyal model, where both 1-way and 2-way flow networks are studied and several treatment conditions are implemented yielding different theoretical predictions.

In their experiment, Falk and Kosfeld (2003) consider groups of four. In every period subjects anonymously and independently decide with whom of the other group members they want to form a direct link. After decisions are made subjects learn the realized network and earn a payoff that depends on the number of links they form and the number of group members to whom they are connected in that period. Groups stay together for five periods, after which they are randomly re-composed. Overall, subjects play 15 periods of the network formation game and participate in three different groups.

The experiment consists of five treatments, three 1-way flow and two 2-way flow treatments. In every treatment the benefit from an additional (direct or indirect) connection is equal to 10. In the 1-way flow model the cost of a direct link is either 5, 15, or 25. In the 2-way flow model the cost of a direct link is either 5 or 15. Table 1 shows the predictions of the Bala and Goyal model based on strict Nash equilibrium and efficiency in the five different treatments of the experiment.

	1-way flow			2-way flow	
Cost of connection	5	15	25	5	15
Strict Nash network	circle	circle, \emptyset	circle, \emptyset	cs-star	\emptyset
Efficient network	circle	circle	circle	m2c	m2c

Note: In all treatments the benefit from an additional connection is equal to 10. (\emptyset = empty network, cs-star = center-sponsored star, m2c = minimally 2-way connected.)

Table 1: Treatments and predictions in Falk and Kosfeld (2003).

The circle is a strict Nash network in all 1-way flow treatments. For costs of a direct link equal to 5 it is the unique strict Nash network, while for costs equal to 15 or 25 the empty network is also a strict Nash network. In all 1-way flow treatments the circle is the unique efficient network. In the 2-way flow model the center-sponsored star is the unique strict Nash network if costs equal 5 and the empty network is the unique strict Nash network if costs equal 15. In both treatments, efficient networks are those that are minimally 2-way connected. This includes the center-sponsored star but not the empty network.

See Falk and Kosfeld (2003) for details of the experiment. The main finding is the following. While in the 1-way flow model more than fifty percent of the networks formed by the subjects are strict Nash networks (i.e., either the circle or the empty network), in the 2-way flow model no strict Nash network (i.e., neither the center-sponsored star nor the empty network) is formed at all. Hence, there is a significant difference between the two scenarios of information flow. Whereas

the notion of strict Nash equilibrium serves as a good prediction in the 1-way flow scenario, it has no predictive power at all in the 2-way flow scenario. This difference is no artefact of the refinement of strict Nash equilibrium, but remains true if we look at all Nash networks instead. Although subjects form Nash networks in the 2-way flow model, they do so significantly less than in the 1-way flow model.

What explanation can be offered to account for this finding? One reason might be that the 2-way flow model is more complex and that it might take more time for the subjects to coordinate on equilibrium. However, simulations of Bala and Goyal (2002a) suggest that convergence times of both models are similar if subjects have the chance to revise their strategy in every period, as is the case in the experiment. Moreover, following the complexity argument, one should not see any difference in behavior once an equilibrium is reached. However, Falk and Kosfeld do find such difference in the data. Given that subjects form a Nash network in period t , the probability to form a Nash network in period $t + 1$ is significantly smaller in the 2-way flow model than in the 1-way flow model. Thus, even if subjects succeed in coordinating on a Nash network, equilibria are less stable in the 2-way flow model.

Another explanation might be that the behavior of the subjects in the experiment is guided by fairness considerations. Note that a crucial difference between the 1-way and the 2-flow model is that in the first model every subject has to form a link himself if he wants to receive any benefits from the network. In contrast, in the 2-way flow model a subject can earn large benefits even if he does not form a link himself. If links are costly, this may cause individual payoffs to be very unequal. For example, in the circle every individual maintains a link and thus everyone earns exactly the same payoff. On the other hand, in the center-sponsored star the individual in the center, who maintains all links, earns a payoff of 25, while individuals on the periphery earn a payoff of 40. If subjects dislike unequal payoffs (as, e.g., in Fehr and Schmidt 1999), they may be unwilling to form such networks.

A first test, which is consistent with this hypothesis, is given by a probit regression measuring the impact of payoff inequity on inertia. The latter is defined as any instance where a subject plays the same strategy as in the previous period. Payoff inequity is measured by the sum of absolute payoff differences between one own and the other players' payoff. Controlling for best replies, it is found that payoff inequity has a significantly negative impact on inertia. The higher the payoff inequity between subjects in a given network is, the less willing subjects are to maintain that network, even if (according to monetary rewards) the maintenance of the network is an individual best reply. As Nash networks in the 2-way flow model generate higher payoff inequity than in the 1-way flow model, this implies that subjects form less Nash networks in the 2-way compared to the 1-way flow model, which is exactly what is found in the experiment.

One conclusion from this result is that networks have to be fairness compatible in order to be stable. Otherwise the notion of Nash equilibrium together with standard money-maximizing preferences generates wrong predictions. A second implication is that mechanisms that help overcome the conflict between fairness and stability (e.g., compensation of the central player or rotation within the network) may play an important role in the formation and maintenance of social networks. In the experiment subjects did not have such possibilities. In reality, however, there

are many of these possibilities, and, as evidence from sociology, psychology, and anthropology suggests, these mechanisms are widely used. Clearly, these mechanisms provide an interesting area for further experimental research, too.

5 Conclusion

The experiments discussed in this paper belong to a recent wave in experimental economics focusing on social and economic networks. Present work emphasizes individual incentives for network formation, as well as network effects on equilibrium selection, competition, and cooperation. Given the interest in the topic and the existing theoretical literature in this area, it seems clear that more experimental studies are on the way.

References

- Aumann, R. and R. Myerson (1988) "Endogenous Formation of Links Between Players and of Coalitions: an Application of the Shapley Value," in: Roth, A.E., ed., *The Shapley Value*, Cambridge University Press, 175-191.
- Bala, V. and S. Goyal (2000a) "A Noncooperative Model of Network Formation," *Econometrica*, 68, 1181-1229.
- Bala, V. and S. Goyal (2000b) "A Strategic Analysis of Network Reliability," *Review of Economic Design*, 5, 205-228.
- Bavelas, A. (1950) "Communication Patterns in Task-Oriented Groups," *Journal of the Acoustical Society of America*, 22, 725-730.
- Berninghaus, S.K., K.-M. Ehrhart, and C. Keser (2002) "Conventions and Local Interaction Structures: Experimental Evidence," *Games and Economic Behavior*, 39, 177-205.
- Bertrand, M., E.F.P. Luttmer, and S. Mullainathan (2000) "Network Effects and Welfare Cultures," *Quarterly Journal of Economics*, 115, 1019-1055.
- Bhaskar, V. and Vega-Redondo, F. (2002) "Migration and the Evolution of Conventions," *Journal of Economic Behavior and Organization*, forthcoming.
- Bienenstock, E.J. and P. Bonacich (1993) "Game-Theory Models for Exchange Networks: Experimental Results," *Sociological Perspectives*, 36, 117-135.
- Bienenstock, E.J. and P. Bonacich (1997) "Network Exchange as a Cooperative Game," *Rationality and Society*, 9, 37-65.
- Blume, L.E. (1993) "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5, 387-424.
- Brown, M., A. Falk, and E. Fehr (2002) "Contractual Incompleteness and the Nature of Market Interactions," mimeo, University of Zurich.

- Callander, S. and C.R. Plott (2003) "Networks: An Experimental Study," mimeo, Northwestern University.
- Camerer, C. and R. Thaler (1995) "Ultimatums, Dictators, and Manners," *Journal of Economic Perspectives*, 9, 209-219.
- Cassar, A. (2002) "Coordination and Cooperation in Local, Random and Small World Networks: Experimental Evidence," in: *Proceedings of the 2002 North American Summer Meetings of the Econometric Society: Game Theory*, edited by D.K. Levine, W. Zame, L. Ausubel, P.-A. Chiappori, B. Ellickson, A. Rubinstein and L. Samuelson, available at: <http://lev0201.dklevine.com/proceedings/game-theory.htm>.
- Charness G., M. Corominas-Bosch, and G.R. Frechette (2001) "Bargaining and Network Structure: An Experiment," mimeo, University of California, Santa Barbara.
- Cook, K.S. and R.M. Emerson (1978) "Power, Equity, and Commitment in Exchange Networks," *American Sociological Review*, 43, 721-739.
- Cook, K.S., R.M. Emerson, M.R. Gillmore, and T. Yamagishi (1983) "The Distribution of Power in Exchange Networks: Theory and Experimental Results," *American Journal of Sociology*, 89, 275-305.
- Corbae, D. and J. Duffy (2002) "Experiments with Network Economies," mimeo, University of Pittsburgh.
- Corominas-Bosch, M. (1999) "On Two-sided Network Markets," PhD Thesis, Universitat Pompeu Fabra.
- Deck, C. and C. Johnson (2002) "Link Bidding in a Laboratory Experiment," mimeo, University of Arkansas.
- Droste, E., R.P. Gilles, and C. Johnson (2000) "Evolution of Conventions in Endogenous Social Networks," mimeo, Virginia Tech.
- Dutta, B. and S. Mutuswami (1997) "Stable Networks," *Journal of Economic Theory*, 76, 322-344.
- Dutta, B., A. van den Nouweland, and S. Tijs (1998) "Link Formation in Cooperative Structures," *International Journal of Game Theory*, 27, 245-256.
- Ellison, G. (1993) "Learning, Local Interaction, and Coordination," *Econometrica*, 61, 1047-1071.
- Ely, J.C. (2002) "Local Conventions," *Advances in Theoretical Economics*, 2(1).
- Eshel, I., L. Samuelson, and A. Shaked (1998) "Altruists, Egoists, and Hooligans in a Local Interaction Model," *American Economic Review*, 88, 157-179.
- Fafchamps, M. and S. Lund (2001) "Risk-Sharing Networks in Rural Philippines," mimeo, Oxford University.

- Falk, A. and M. Kosfeld (2003) "It's all about Connections: Evidence on Network Formation," mimeo, University of Zurich.
- Fehr, E. and K.M. Schmidt (1999) "A Theory of Fairness, Competition and Cooperation," *Quarterly Journal of Economics*, 114, 817-868.
- Freeman, L.C. (1979) "Centrality in Social Networks: Conceptual Clarification," *Social Networks*, 1, 215-239.
- Freeman, L.C. (1980) "Centrality in Social Networks: II. Experimental Results," *Social Networks*, 2, 119-141.
- Goyal, S. and S. Joshi (2002) "Networks of Collaboration in Oligopoly," mimeo, Queen Mary, University of London.
- Goyal, S. and J.L. Moraga-Gonzalez (2001) "R&D Networks," *RAND Journal of Economics*, 32, 686-707.
- Goyal, S. and F. Vega-Redondo (2000) "Learning, Network Formation and Coordination," mimeo, University of Alicante.
- Granovetter, M. (1974) *Getting a Job*, University of Chicago Press, Chicago.
- Haller, H. and S. Sarangi (2001) "Nash Networks with Heterogeneous Agents," mimeo, Virginia Polytechnic Institute.
- Hauk, E. and R. Nagel (2001) "Choice of Partners in Multiple Two-Person Prisoner's Dilemma Games," *Journal of Conflict Resolution*, 45, 770-793.
- Holzer, H. (1987) "Hiring Procedures in the Firm: Their Economic Determinants and Outcomes," NBER Working Paper No. 2185.
- Huck, S. and M. Kosfeld (2001) "Local Control: The Private Enforcement of Public Rules," mimeo, University of Zurich.
- Jackson, M.O. (2001) "The Stability and Efficiency of Economic and Social Networks," forthcoming in: Sertel, M. and S. Koray, eds., *Advances in Economic Design*, Springer, Heidelberg.
- Jackson, M.O. (2003) "A Survey of Models of Network Formation: Stability and Efficiency," mimeo, California Institute of Technology.
- Jackson, M.O. and A. Watts (2001) "The Evolution of Social and Economic Networks," *Journal of Economic Theory*, forthcoming.
- Jackson, M.O. and A. Wolinsky (1996) "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory*, 71, 44-74.
- Johnson, C. and R.P. Gilles (2000) "Spatial Social Networks," *Review of Economic Design*, 5, 273-300.
- Kandori, M., G.J. Mailath, and R. Rob (1993) "Learning, Mutation and Long Run Equilibria in Games," *Econometrica*, 61, 29-56.

- Keser, C., K.-M. Ehrhart, and S.K. Berninghaus (1998) "Coordination and Local Interaction: Experimental Evidence," *Economics Letters*, 58, 269-275.
- Kirchkamp, O. (2000) "Spatial Evolution of Automata in the Prisoners' Dilemma," *Journal of Economic Behavior and Organization*, 43, 239-262.
- Kirchkamp, O. and R. Nagel (2001) "Local and Group Interaction in Prisoners' Dilemmas," mimeo, University of Mannheim.
- Kosfeld, M. (2002) "Stochastic Strategy Adjustment in Coordination Games", *Economic Theory*, 20, 2002, 321-339.
- Kranton, R. and D. Minehart (2001) "A Theory of Buyer-Seller Networks," *American Economic Review*, 91, 485-508.
- Lazerson, M. (1993) "Factory or Putting-Out? Knitting Networks in Modena," in Grabher G., ed., *The Embedded Firm: On the Socioeconomics of Industrial Networks*, 203-226, Routledge, New York.
- Leavitt, H.J. (1951) "Some Effects of Certain Communication Patterns on Group Performance," *Journal of Abnormal and Social Psychology*, 46, 38-50.
- Lévi-Strauss, C. (1963) *Structural Anthropology*, Basic Books, New York.
- Mailath, G., L. Samuelson, and A. Shaked (2001) "Endogenous Interactions," in Nicita, A. and U. Pagano, eds., *The Evolution of Economic Diversity*, Routledge, 300-324.
- Markovsky, B., D. Willer, and T. Patton (1988) "Power Relations in Exchange Networks," *American Sociological Review*, 53, 220-236.
- McMillan, J. and C. Woodruff (1999) "Interfirm Relationships and Informal Credit in Vietnam," *Quarterly Journal of Economics*, 114, 1285-1320.
- Montgomery, J. (1991) "Social Networks and Labor Market Outcomes: Towards an Economic Analysis," *American Economic Review*, 81, 1408-1418.
- Morris, S. (2000) "Contagion," *Review of Economic Studies*, 67, 57-79.
- Myerson, R. (1977) "Graphs and Cooperation in Games," *Mathematics of Operations Research*, 2, 225-229.
- Nishiguchi, T. (1994) *Strategic Industrial Sourcing*, Oxford University Press, New York.
- Nowak, M.A. and R.M. May (1992) "Evolutionary Games and Spatial Chaos," *Nature*, 359, 826-829.
- Ochs, J. (1995) "Coordination Problems," in Kagel, J.H. and A.E. Roth, eds., *Handbook of Experimental Economics*, Princeton University Press, 195-251.
- Riedl, A. and Ule, A. (2002) "Exclusion and Cooperation in Social Network Experiments," mimeo, University of Amsterdam.

- Roth, A.E. (1995) "Bargaining Experiments," in: Kagel, J.E. and A.E. Roth, eds., *Handbook of Experimental Economics*, Princeton University Press, Princeton.
- Ruys, P. (1997) "From National Public Utilities to European Network Industries," *Annals of Public and Cooperative Economics*, 68, 35-452.
- Ruys, P. and R. van den Brink (1999) "Positional Abilities and Rents on Equilibrium Wages and Profits," in Herings, P.J.J., G. van der Laan, and A.J.J. Talman, eds., *The Theory of Markets*, North-Holland, 261-279.
- Ruys, P. and R. van den Brink (2000) "The Internal Organization of the Firm and its External Environment," mimeo, Tilburg University.
- Sarangi, S., R. Kannan, and L. Ray (2003) "The Structure of Information Networks," mimeo, Louisiana State University.
- Slikker M. and A. van den Nouweland (2000) "Network Formation Models with Costs for Establishing Links," *Review of Economic Design*, 5, 333-362.
- Shaw, M.E. (1964) "Communication Networks," in Berkowitz, L., ed., *Advances in Experimental Social Psychology*, Academic Press, New York, 111-147.
- Van Huyck, J.B., R.C. Battalio, and R.O. Beil (1990) "Tacit Coordination Games, Strategic Uncertainty, and Coordination Failure," *American Economic Review*, 80, 234-249.
- Vanin, P. (2002) "Network Formation in the Lab: A Pilot Experiment," mimeo, Universitat Pompeu Fabra.
- Vega-Redondo, F. (2002) "Building Up Social Capital in a Changing World," mimeo, University of Alicante.
- Watts, A. (2001) "A Dynamic Model of Network Formation," *Games and Economic Behavior*, 34, 331-341.
- Watts, D.C. (1999) *Small Worlds*, Princeton University Press.
- Watts, D.C. and S.H. Strogatz (1998) "Collective Dynamics of 'Small-World' Networks," *Nature*, 393, 440-442.
- Weibull, J.W. (2001) "Testing Game Theory," mimeo, Stockholm School of Economics.
- Young, H.P. (1993) "The Evolution of Conventions," *Econometrica*, 61, 57-84.

MARKTWERKING IN DE MONDZORG

Edward Droste*

Essent N.V.

1 Inleiding

De huidige organisatie van de gezondheidszorg in Nederland kenmerkt zich door aanbodregulering. Zo bepaalt de overheid in belangrijke mate welke producten en diensten door de zorgsector worden voortgebracht, wie de betreffende zorg mogen verlenen, welke prijs de zorgverleners voor de producten en diensten in rekening mogen brengen en wie voor de kosten van de zorg opdraait. Het systeem van aanbodregulering is er op gericht om door middel van direct overheidsingrijpen de publieke belangen kwaliteit, toegankelijkheid en betaalbaarheid van de gezondheidszorg te realiseren. Het ingrijpen van de overheid betreft daarbij zowel de marktstructuur als het marktgedrag.

De overheid streeft er naar om het huidige stelsel van aanbodregulering waar mogelijk te gaan vervangen door een systeem van gereguleerde marktwerking.¹ De introductie van marktwerking moet het mogelijk maken om de bovengenoemde publieke belangen op een meer efficiënte wijze te realiseren. In een systeem van gereguleerde marktwerking zijn het de consumenten die uiteindelijk bepalen aan welke producten en diensten ze behoefte hebben en hoeveel ze bereid zijn daarvoor te betalen. In aanvulling daarop laat de overheid de bepaling van het aanbod over aan de zorgverleners zelf en laat ze de prijsvorming over aan de vrije markt. De overheid kan bij gereguleerde marktwerking bijvoorbeeld wel kwaliteitseisen blijven stellen aan de zorgverleners en de door hen geleverde producten en diensten. Ook kan de overheid regels blijven stellen aan het marktgedrag van zorgverleners en verzekeraars om direct de toegankelijkheid en betaalbaarheid van de zorg te waarborgen.

Dit paper gaat in op de mogelijkheden voor meer marktwerking in een specifiek onderdeel van de gezondheidszorg, namelijk de tandheelkundige zorg oftewel mondzorg. Specifieke kenmerken van de markt voor mondzorg maken deze markt vanuit economisch oogpunt interessant. Bovendien komen de meeste Nederlanders als gevolg van de periodieke controle met enige regelmaat met de markt voor mondzorg in aanraking.

Paragraaf 2 definieert de markt voor mondzorg. Paragraaf 3 geeft een overzicht van de relevante eigenschappen van deze markt. De efficiëntie van de markt en

* *Adres:* Essent N.V., Postbus 268,6800 AG Arnhem *Email:* edward.droste@essent.nl

¹Zie de beleidsbrief *Modernisering Curatieve Zorg* (Tweede Kamer, vergaderjaar 2000-2001, 23 619, nr. 18) en de nota *Vraag aan bod* (Tweede Kamer, vergaderjaar 2000-2001, 27 855, nr.2).

de realisatie van de publieke belangen komen respectievelijk in paragraaf 4 en 5 aan de orde. Paragraaf 6 inventariseert de belemmeringen voor de introductie van meer marktwerking in de mondzorg. De conclusies komen tenslotte aan de orde in paragraaf 7.

2 Marktafbakening

Deze paragraaf stelt de afbakening van de markt voor mondzorg aan de orde. Daarbij komt ter sprake welke producten en diensten op deze markt verhandeld worden, wie deze producten en diensten leveren, wie voor de kosten van mondzorg opdraait en hoe de financieringsstructuur van invloed is op de deelmarkten die binnen de markt voor mondzorg onderscheiden kunnen worden.

Mondzorg omvat een groot aantal producten en diensten: van het vullen en trekken van tanden en kiezen tot het aanbrengen van kronen, bruggen en prothesen, en van het geven van een instructie mondhygiëne en het verwijderen van tandsteen tot het uitvoeren van wortelkanaalbehandelingen en orthodontische behandelingen. Deze producten en diensten worden aangeboden door verschillende zorgverleners, namelijk tandartsen, mondhygiënist, tandprotheticen en tandartsspecialisten (kaakchirurgen en orthodontisten).

Tandartsen zijn ten opzichte van de andere aanbieders van mondzorg op een relatief breed terrein binnen de mondzorg actief. Samen met mondhygiënist en tandprotheticen vormen tandartsen de eerstelijnszorg. Desondanks werken mondhygiënist en tandprotheticen voor een deel op verwijzing van tandartsen. De tweedelijnszorg wordt geleverd door de tandartsspecialisten die voornamelijk op verwijzing van tandartsen werken. Tandartsen vormen daarmee de centrale spil in het aanbod van mondzorg.

Mondzorg valt voornamelijk in het zogenaamde tweede en derde compartiment van de gezondheidszorg.² Voor zover het gaat om mondzorg in het tweede compartiment zijn het de zorgverzekeraars die — namens hun verzekerden — optreden als vrager op de zorginkoopmarkt. Naast deze zorginkoopmarkt bestaat de markt voor mondzorg dan nog uit een zorgverleningsmarkt en een zorgverzekeringsmarkt. Voor zover het gaat om mondzorg die niet onder een verzekering valt of onder een (aanvullende) particuliere verzekering valt, maakt de consument zelf de afweging om al dan niet een beroep te doen op een zorgverlener. In dat geval bestaat de markt voor mondzorg uit twee deelmarkten: een zorginkoopmarkt (die samenvalt met de zorgverleningsmarkt) en een zorgverzekeringsmarkt.

²Het eerste compartiment van de gezondheidszorg bevat de langdurige zorg en de zogeheten onverzekerbare medische risico's. De verstrekking en financiering van deze zorg wordt grotendeels geregeld door de overheid via de AWBZ. Het tweede compartiment bevat kortdurende noodzakelijke geneeskundige zorg die voor iedereen toegankelijk moet zijn. Zowel de overheid als de ziekenfondsen en de particuliere verzekeraars hebben een rol in de verstrekking en financiering van deze zorg: de sociale ziekenfondsverzekering en de reguliere particuliere ziektekostenverzekering bestaan hier naast elkaar. Het derde compartiment bevat de zorg die niet in het eerste of tweede compartiment is ondergebracht en waarvoor iedere burger zich vrijwillig aanvullend kan verzekeren.

3 Karakteristieken van de markt voor mondzorg

In deze paragraaf komen de karakteristieken van de markt voor mondzorg ter sprake die van invloed zijn op de efficiëntie van de markt en het realiseren van de publieke belangen bij de introductie van marktwerking. Daarbij wordt onderscheid gemaakt tussen intrinsieke kenmerken van mondzorg en karakteristieken van de marktstructuur en het marktgedrag.

Intrinsieke kenmerken

Mondzorg valt uiteen in preventieve en curatieve zorg. Met uitzondering van een klein deel van de curatieve zorg is mondzorg niet spoedeisend. Preventieve mondzorg kent een hogere prijselasticiteit dan curatieve mondzorg: als de prijs van mondzorg stijgt dan zullen consumenten de periodieke controle relatief gemakkelijk laten schieten, terwijl men met kiespijn wel naar de tandarts gaat ongeacht de kosten. Het toenemende gebitsbewustzijn van de Nederlandse bevolking maakt de vraag naar preventieve mondzorg overigens wel minder gevoelig voor prijsveranderingen. Prijselasticiteit speelt verder ook binnen de curatieve mondzorg een rol, namelijk als een bepaalde behandeling niet wordt vergoed door de verzekering of als er sprake is van een eigen risico.

Op de markt voor mondzorg is sprake van asymmetrische informatie. Ondanks dat er gebruikelijk overleg plaatsvindt tussen de patiënt en de zorgverlener over de te volgen behandeling hebben patiënten op twee punten een informatieachterstand ten opzichte van de aanbieder van mondzorg. Ten eerste, wat betreft het nut en de noodzaak van een door de zorgverlener aanbevolen tandheelkundige behandeling. De omvang van de informatieachterstand wordt hier beperkt door de relatief eenduidige combinaties van diagnoses en behandelingen binnen de mondzorg: als patiënten kiespijn hebben dan weten ze dat de kies gevuld of getrokken moet worden en als hun tanden scheef staan dan weten ze dat een orthodontische behandeling een optie is. Ten tweede, de informatieasymmetrie met betrekking tot de kwaliteit van een uitgevoerde tandheelkundige behandeling. Patiënten kunnen natuurlijk wel beoordelen of na de behandeling de kiespijn is verdwenen of dat hun tanden minder scheef staan, maar als een vulling na een jaar uitvalt dan weten ze niet of dit aan hen ligt of aan de manier waarop de tandarts de vulling heeft aangebracht.

Marktstructuur en marktgedrag

In theorie biedt de mondzorg mogelijkheden voor aanbodsubstitutie. Bijvoorbeeld, mondhygiënische zorg en prothetische zorg kunnen zowel door mondhygiënist en tandprothetici als door tandartsen worden geleverd. Voor wat betreft eenvoudige orthodontische en kaakchirurgische behandelingen is aanbodsubstitutie mogelijk tussen enerzijds tandartsen en anderzijds respectievelijk orthodontisten en kaakchirurgen. De mate waarin aanbodsubstitutie zich in de praktijk voordoet is echter beperkt aangezien de werkzaamheden van aanbieders van mondzorg veelal worden bepaald door hun beroepsmatige interesse. Alleen verschillen in werkdruk tussen de verschillende aanbieders van mondzorg lijken voor het optreden van aan-

bodsubstitutie te zorgen.³

Wet- en regelgeving zijn van grote invloed op de structuur van de markt voor mondzorg en het gedrag van de betrokken marktpartijen. Zoals gezegd brengen wet- en regelgeving mondzorg voor een belangrijk deel onder in het derde compartiment van de gezondheidszorg. Hierdoor is de betekenis van de zorginkoopmarkt beperkt en speelt de consument zelf een belangrijke rol bij de keuze van een zorgverlener.

Ook de Ziekenfondswet (ZFW) is direct van invloed op de markt voor mondzorg. De ZFW specificeert dat mondzorg voor ziekenfondsverzekerden alleen wordt vergoed als ze wordt geleverd door of in onderaanneming van tandartsen of tandartsspecialisten. Hierdoor worden mondhygiënist en tandprotheticen belemmerd in hun optreden op de markt voor mondzorg.

Verder valt mondzorg geleverd door tandartsen of tandartsspecialisten onder de Wet tarieven gezondheidszorg (WTG). De WTG specificeert voor mondzorg gedetailleerde verrichtingen en bijbehorende maximum tarieven. Als gevolg van de voorgeschreven verrichtingen en de scherpe maximum tarieven bestaan er voor deze zorgverleners weinig mogelijkheden om zich in hun dienstverlening te onderscheiden van hun concurrenten.

De Wet beroepen in de individuele gezondheidszorg (Wet BIG) definieert een zogenaamd domeinmonopolie door een aantal handelingen exclusief voor te behouden aan tandartsen en tandartsspecialisten. Daarnaast regelt de Wet BIG onder andere de titelbescherming van tandartsen, mondhygiënist, tandprotheticen, orthodontisten en kaakchirurgen. Regels van de beroepsorganisatie stellen bovendien dat tandartsspecialisten zich in hun werkzaamheden dienen te beperken tot hun specialisme.

Tenslotte wordt de capaciteit van het aanbod van mondzorg in Nederland gereguleerd. Zo bestaan er *numeri fixi* voor de (vervolg)opleidingen tot tandarts, kaakchirurg, orthodontist en mondhygiënist. Samen met de opleidingseisen voor aanbieders van mondzorg vormen de *numeri fixi* toetredingsbelemmeringen tot de markt. Sinds een aantal jaren staat het aanbod van mondzorg onder druk. De belangrijkste oorzaak daarvan is dat tandartsen steeds meer parttime werken en daardoor minder patiënten kunnen behandelen. Aanbieders van mondzorg werken van oudsher vaak in solopraktijken of kleine groepspraktijken al is mede als gevolg van meer parttime werken een trend van schaalvergroting en het aannemen van meer assistenten waarneembaar.

4 Efficiëntie

Het functioneren van een markt wordt typisch beoordeeld op basis van zijn efficiëntie. De efficiëntie is afhankelijk van de prikkels en mechanismen die voortvloeien uit de wisselwerking tussen de intrinsieke kenmerken, de marktstructuur en het marktgedrag zoals die in de vorige paragraaf zijn besproken. In deze paragraaf bespreken we de prikkels en mechanismen in de markt voor mondzorg die van invloed

³In het geval van meer specialistische of meer complexe behandelingen kenmerkt de relatie tussen enerzijds de tandartsen en anderzijds de overige aanbieders van mondzorg zich daarentegen meer door complementariteit.

zijn op de efficiëntie. Daarbij maken we onderscheid tussen technische efficiëntie (een kostenefficiënte bedrijfsvoering), allocatieve efficiëntie (het voorzien in de behoefte van de consument) en dynamische efficiëntie (de mate van innovatie).

Technische efficiëntie

Een significant deel van de maatschappelijke kosten van mondzorg bestaat uit de opleidingskosten van aanbieders van mondzorg. Met uitzondering van tandprotheticen dragen de zorgverleners zelf echter maar een zeer beperkt deel van de opleidingskosten. Daardoor bestaat er voor een individuele zorgverlener geen prikkel om lang fulltime te blijven werken en zo deze kosten terug te verdienen. De trend dat steeds meer tandartsen parttime gaan werken heeft een negatief effect op de technische efficiëntie.

De markt bevat wel een prikkel tot schaalvergroting en herverdeling van arbeid binnen praktijken. Voor zorgverleners is het financieel aantrekkelijk om assistenten in hun praktijk aan het werk te stellen. Verrichtingen die door assistenten onder verantwoordelijkheid van de zorgverlener worden uitgevoerd kunnen namelijk tegen het tarief van de zorgverlener bij de zorgverzekeraar worden gedeclareerd. De waarneembare schaalvergroting en herverdeling van arbeid binnen praktijken heeft een positieve prikkel op de technische efficiëntie. Al neemt het aantal solopraktijken ook af als gevolg van het toenemende aantal zorgverleners dat parttime werkt.

Tenslotte is er een negatieve invloed op de technische efficiëntie doordat veel aanbieders van mondzorg werkzaamheden verrichten die net zo goed kunnen worden uitgevoerd door aanbieders met een lagere of minder specialistische opleiding. Zo hebben tandartsen bij de huidige krapte op de markt de prikkel om patiënten voor relatief eenvoudige behandelingen (die ze ook zelf kunnen verrichten) toch te verwijzen naar een tandartsspecialist. In het geval van een ruime markt zouden tandartsen deze patiënten eerder wel zelf behandeld hebben. Op vergelijkbare wijze hebben tandartsen ook de verkeerde prikkels voor het doorsturen van patiënten naar mondhygiënist en tandprotheticen.

Allocatieve efficiëntie

De asymmetrische informatie tussen zorgverleners en consumenten vormt een risico voor de allocatieve efficiëntie van de markt voor mondzorg. Zoals gezegd is het voor consumenten moeilijk het nut, de noodzaak en de kwaliteit van tandheelkundige behandelingen te beoordelen. Hierdoor bestaat de mogelijkheid van supply induced demand, oftewel dat de vraag naar mondzorg wordt bepaald door de voorkeuren van de zorgaanbieders. Het is duidelijk dat supply induced demand niet overeen hoeft te komen met de behoefte van de consumenten.

Dynamische efficiëntie

De innovatie binnen de mondzorg wordt afgeremd door de WTG die stelt dat zorgverleners alleen verrichtingen mogen declareren die zijn gedefinieerd in de WTG.

Het opnemen van aanvullende verrichtingen in de WTG is op het moment nog een langdurig en moeizaam proces.⁴

Een positieve prikkel op de dynamische efficiëntie van de markt voor mondzorg gaat uit van het kwaliteitsbeleid van de beroepsorganisaties van de verschillende zorgverleners. In het kader van hun kwaliteitsbeleid bieden deze organisaties verplichte en/of vrijwillige bij- en nascholing aan voor hun leden.

5 Publieke belangen

In de gezondheidszorg heeft de overheid niet alleen belang bij een efficiënte markt, maar ook bij de realisatie van publieke belangen. In de gezondheidszorg worden kwaliteit, toegankelijkheid en betaalbaarheid als publieke belangen onderscheiden. In deze paragraaf geven we ten eerste aan wat er onder kwaliteit, toegankelijkheid en betaalbaarheid wordt verstaan. Vervolgens gaan we in op de realisatie van deze publieke belangen binnen de mondzorg.

Kwaliteit, toegankelijkheid en betaalbaarheid

Onder kwaliteit wordt verstaan de veiligheid en effectiviteit van de zorgverlening, zowel volgens de zorgverlener als de patiënt. De vraag is vervolgens welke instrumenten in de markt worden ingezet om te komen tot kwalitatief hoogstaande zorg, waarbij deze instrumenten kunnen voortkomen uit voorschriften van de overheid of de beroepsgroepen, of uit de handelswijze van individuele zorgverleners. Voorbeelden van dergelijke instrumenten zijn het in kaart brengen van de prestaties van zorgverleners en de tevredenheid van patiënten aan de hand van indicatoren en normen.

Toegankelijkheid van de zorg valt uiteen in een viertal aspecten, namelijk bereikbaarheid en beschikbaarheid van de zorgverlening en de toegangstijd en financiële toegankelijkheid voor de patiënt:

- bereikbaarheid van de zorgverlening wordt bepaald door de geografische spreiding van de dienstverlening, oftewel de afstand die moet worden afgelegd van een willekeurige plaats in Nederland naar de dichtstbijzijnde dienstverlener of de tijd die het kost om deze afstand af te leggen. Een ander aspect van bereikbaarheid is op welke tijdstippen van de dag de zorgverleners bereikbaar zijn;
- beschikbaarheid van de zorgverlening is afhankelijk van het aantal gekwalificeerde zorgverleners per duizend inwoners dat de betreffende zorg kan aanbieden;

⁴Het opnemen van aanvullende verrichtingen in de WTG is een langdurig en moeizaam proces mede als gevolg van het volgende kip-ei probleem. Het College Tarieven Gezondheidszorg (CTG) waaraan de uitvoering van de WTG is opgedragen, is bereid verrichtingen aan de WTG toe te voegen zodra voldoende zorgverleners deze verrichtingen aanbieden. Zorgverleners wachten daarentegen met het aanbieden van aanvullende verrichtingen totdat ze zijn opgenomen in de WTG. Alleen dan kunnen zij de verrichtingen namelijk declareren bij de zorgverzekeraars. Los van de dynamische efficiëntie leveren de gespecificeerde verrichtingen en bijbehorende maximum tarieven wel een positieve bijdrage aan de transparantie van de markt voor mondzorg.

- de toegangstijd voor de patiënt betreft de tijd dat een patiënt op een behandeling moet wachten vanaf het moment dat de zorgbehoefte ontstaat dan wel het moment van indicatiestelling;
- financiële toegankelijkheid voor de patiënt draait om de vraag hoeveel consumenten vanwege financiële redenen geen gebruik kunnen maken van een bepaalde vorm van zorg.

Bij betaalbaarheid van de zorg worden microbetaalbaarheid en macrobetaalbaarheid onderscheiden. Microbetaalbaarheid hangt samen met de prijs van een behandeling, oftewel het bedrag dat een zorgverlener voor een behandeling in rekening brengt bij de patiënt of zijn verzekeraar. Macrobetaalbaarheid betreft de totale maatschappelijke kosten van het betreffende type zorg en wordt daarmee dus niet alleen bepaald door de prijs van behandelingen maar ook door het aantal behandelingen (volume).

Realisatie van publieke belangen in de mondzorg

Hieronder komt de realisatie van publieke belangen in de markt voor mondzorg aan de orde, waarbij achtereenvolgens wordt ingegaan op kwaliteit, toegankelijkheid en betaalbaarheid.

Uit onderzoek onder consumenten blijkt een grote tevredenheid over mondzorg, wat wordt ondersteund door het feit dat de algemene kwaliteit van mondzorg in Nederland in internationale vergelijkingen hoog scoort. Zo beschikt in Nederland bijvoorbeeld een relatief hoog percentage van de bevolking over zijn eigen tanden. Het meten van de kwaliteit van mondzorg op het niveau van individuele zorgverleners vindt nauwelijks plaats bij gebrek aan goede indicatoren voor het meten van de kwaliteit van hun output. Kwaliteit op het niveau van individuele zorgverleners wordt dan ook afgemeten aan de input en het proces dat tot het leveren van mondzorg leidt. Aan het daarvoor benodigde kwaliteitsbeleid wordt invulling gegeven door de beroepsorganisaties en door individuele zorgverleners. Daarbij wordt onder andere gebruik gemaakt van intercollegiaal overleg, peer-reviews en bij- en nascholing.

Wat betreft de toegankelijkheid van mondzorg kan worden opgemerkt dat op het moment de beschikbaarheid enigszins onder druk staat, met name als gevolg van tekorten aan tandartsen in sommige steden en regio's. Voor tandartsspecialisten en mondhygiënisten geldt in een aantal regio's in Nederland een relatief lange toegangstijd. Om deze problemen op te lossen is inmiddels besloten de opleidingscapaciteit van de (vervolg)opleidingen tot tandarts, mondhygiënist, orthodontist en kaakchirurg te vergroten. De financiële toegankelijkheid van mondzorg wordt gegarandeerd door de ZFW, tenminste voor zover de zorg is ondergebracht in het tweede compartiment van de gezondheidszorg. Mondzorg in het derde compartiment wordt voor een groot deel gedekt door aanvullende particuliere verzekeringen.

De WTG regelt de microbetaalbaarheid voor een groot deel van de mondzorg, te weten, de mondzorg die wordt geleverd door of in onderaanneming van tandartsen en tandartsspecialisten. Met de WTG ligt dus ook de prijscomponent van de macrobetaalbaarheid in belangrijke mate vast. De volume component van de macrobetaalbaarheid wordt gecontroleerd door middel van de opleidingscapaciteit voor

tandartsen, tandartsspecialisten en mondhygiënisten.

6 Belemmeringen voor marktwerking

De voorafgaande paragrafen geven een overzicht van de factoren die een rol spelen bij het functioneren van de markt voor mondzorg. Op basis van dit overzicht zijn belemmeringen voor een verdere introductie van marktwerking in de mondzorg te identificeren. In het onderstaande gaan we achtereenvolgens in op belemmeringen die te maken hebben met contracteervrijheid, belang van zorgverzekeraars, transparantie van de markt en keuzemogelijkheden.

Contracteervrijheid

Zorgverzekeraars zijn in principe vrij om naar eigen inzicht aanbieders van mondzorg te contracteren voor het verlenen van zorg aan hun verzekerden. Hierbij geldt echter wel de beperking vanuit de ZFW dat tandheelkundige zorg voor ziekenfondsverzekerden alleen door of in onderaanneming van tandartsen of tandartsspecialisten geleverd mag worden. Ziekenfondsverzekeraars kunnen dus niet rechtstreeks terecht bij bijvoorbeeld mondhygiënisten.

Een bijkomend probleem is de relatief grote afhankelijkheid van veel praktijken voor tandheelkundige zorg van een contract met de regionaal dominante ziekenfondsverzekeraar. Een beslissing van een ziekenfondsverzekeraar om met een aanbieder van mondzorg geen contract aan te gaan, kan het bestaansrecht van de betreffende aanbieder direct aantasten. Dit geldt met name voor praktijken van tandprothetici die een relatief groot aantal ziekenfondsverzekerden als klant hebben.

Zoals blijkt uit het bovenstaande is de contracteervrijheid op de zorginkoopmarkt voor mondzorg niet optimaal. Zowel voor zorgverzekeraars als voor zorgverleners zijn er met betrekking tot dit aspect significante beperkingen. Hieruit komt een belemmering naar voren voor een verdere introductie van marktwerking.

Belang van zorgverzekeraars

Zorgverzekeraars hebben een groot belang bij het afsluiten van contracten met aanbieders van mondzorg. Verzekerden hechten in hun keuze voor een bepaalde zorgverzekeraar namelijk veel waarde aan de beschikbaarheid van mondzorg en de vergoedingen voor mondzorg. Het grote belang van zorgverzekeraars heeft in de eerste plaats betrekking op orthodontische zorg en dat belang komt met name voort uit de collectieve contracten die zorgverzekeraars met bedrijven en overheden afsluiten. Van de consumenten die onder collectieve contracten vallen, heeft een relatief groot deel kinderen in de leeftijdscategorie van 10 tot 14 jaar. Veel kinderen in deze leeftijdscategorie ondergaan langdurige en kostbare orthodontische behandelingen. Logischerwijze spelen de vergoedingen voor orthodontische zorg dan ook een belangrijke rol bij de keuze van een zorgverzekeraar.

Het belang van zorgverzekeraars bij mondzorg verzwakt hun positie op de zorginkoopmarkt. Dit is een belemmering voor de introductie van marktwerking. Daarnaast geldt dat het grote aantal contracten dat afgesloten moet worden om over

voldoende mondzorg te beschikken hoge transactiekosten met zich meebrengt, zowel voor de zorgverzekeraars als de zorgverleners.

Transparantie van de markt

De beperkte transparantie van de markt is een belangrijke belemmering voor de introductie van marktwerking. Zowel zorgverzekeraars als consumenten kunnen als gevolg van asymmetrische informatie nut, noodzaak en kwaliteit van tandheelkundige behandelingen maar moeilijk beoordelen. Het beoordelen van de kwaliteit wordt nog eens extra bemoeilijkt door het ontbreken van indicatoren en normen voor het meten van kwaliteit, met als gevolg dat ook contracten moeilijk in termen van prestaties te specificeren zijn.

Keuzemogelijkheden

De mogelijkheden voor verzekeraars en patiënten om een aanbieder van mondzorg te kiezen worden in ernstige mate beperkt door de huidige tekorten aan zorgverleners. Hiervoor is een aantal oorzaken aan te wijzen. Ten eerste het gereguleerde aanbod van zorgverleners door middel van een beperkte opleidingscapaciteit. Een tweede oorzaak is de systematiek van contractering waarin besloten ligt dat ziekenfondsverzekerden voor zorg die is ondergebracht in het tweede compartiment alleen terecht kunnen bij zorgaanbieders die door hun ziekenfondsverzekeraar zijn gecontracteerd. Ten derde is er nauwelijks sprake van potentiële concurrentie als gevolg van het ontbreken van aanbodssubstitutie. Of een tandarts bijvoorbeeld zelf orthodontische zorg verricht of de patiënt doorverwijst naar een orthodontist, hangt voornamelijk af van zijn beroepsmatige interesse en werkdruk. Ook worden de mogelijkheden voor aanbodssubstitutie beperkt door het domeinmonopolie van tandartsen en tandartsspecialisten en het feit dat tandartsspecialisten zich dienen te beperken tot hun specialisme.

7 Conclusie

Op grond van de voorafgaande analyse kunnen de volgende conclusies worden getrokken met betrekking tot de efficiëntie van de markt, de realisatie van de publieke belangen en de belemmeringen voor marktwerking.

De prikkels voor een efficiënte markt voor mondzorg zijn zwak. Als gevolg van asymmetrische informatie zijn nut, noodzaak en kwaliteit van tandheelkundige behandelingen door zorgverzekeraars en consumenten moeilijk vast te stellen. Ook ontbreken de juiste prikkels om te komen tot een optimale taakverdeling tussen de verschillende zorgverleners.

De publieke belangen van de markt voor mondzorg worden in redelijke mate gerealiseerd. De algemene kwaliteit van de Nederlandse mondzorg is goed. De financiële toegankelijkheid wordt ten dele gegarandeerd door de ZFW en daarnaast gerealiseerd doordat mondzorg in veel aanvullende particuliere verzekeringen is opgenomen. De betaalbaarheid wordt geregeld door de WTG. Alleen de beschikbaarheid van mondzorg staat enigszins onder druk.

Belemmeringen voor marktwerking in de mondzorg komen voornamelijk voort uit een beperkte contracteervrijheid voor zorgverzekeraars, de afhankelijkheid van aanbieders van mondzorg van de regionaal dominante ziekenfondsverzekeraar, de informatieachterstand van zorgverzekeraars en consumenten en het gereguleerde aanbod van zorgverleners.

NATIONAL CULTURE, ECONOMIC BEHAVIOR AND ORGANIZATION, AND ECONOMIC THEORY

Radislav Semenov*

University of Nijmegen

I first met Pieter Ruys in 1995, when I arrived at Tilburg University to start my PhD research into the reasons for cross-country differences in financial systems, and he was appointed my supervisor. During the course of my research I discovered that cultural differences between countries provide by far the best explanation of cross-country differences in financial systems. I became interested in influence of cultural differences on economy, and undertook the analysis of such influence on several areas of economic organization. Pieter was very enthusiastic and supportive of this research, suggested many novel ideas, but also gave a very substantial space for my initiative. This resulted in my thesis "Cross-Country Differences in Economic Governance: Culture as a Major Explanatory Factor" (Semenov, 2000).

I was not educated primarily in mainstream economics, was very conscious about the fact that neoclassical economics makes unrealistic assumptions, and was initially quite skeptical about the ability of formal theory to explain many of the aspects of the complex reality. During my work in Tilburg, to a large extent under the influence of Pieter, I came to a better understanding of philosophical and methodological behind mainstream economics and of the way its approach is treated by economists themselves, as well as of the power and beauty of formal theory. Still, I was convinced that more attention to many complexities inherent in reality, in particular those related to social and cultural embeddedness of economic life, were necessary for understanding of economic processes. Pieter shared the same view; some of his PhD students worked on incorporating social organization into formal economic models. During my stay in Tilburg we together conducted some joint work in this direction (together with Rene van den Brink), and this work is continuing at present.

This contribution presents some thoughts on possibilities of incorporating cultural analysis into mainstream economic theory, as well as some examples illustrating, and making first steps at exploring, these possibilities. Many of the ideas presented crystallized when I did research in Tilburg, but all of them are inspired by the experience I had at that time, most importantly our conversations with Pieter.

* *Address:* Department of Economics, Nijmegen School of Management, University of Nijmegen, Nijmegen. *Email:* r.semenov@bw.kun.nl

1 Introduction

Culture, values held by people, cultural differences between societies attract very little attention of economists. Most economists working within the neoclassical tradition treat human behaviors as manifestations of universal human characteristics. Cultural patterns that are time- and place-specific rarely play a role in economic models, and then usually only as constraints preventing the achievement of the most efficient outcome, which is universal for all modern cultural systems.

In recent years cross-country differences in economic organization and economic performance became a topic of significant interest of scholars. Similar transactions are organized very differently in different countries, even in relatively similar ones, such as developed Western countries. In many cases institutions and organizational forms that would be deemed obsolete and inefficient by traditional theory lead to more efficient results than more 'modern' and more *laissez-faire* institutions and organizational patterns (see e.g. Streeck, 1997). These facts are impossible to explain and understand without paying attention to cultural differences.

Culture significantly influences economic behavior and organization. Cultural values influence mental models with which individuals interpret various aspects of their environment. These models, as the research in this field has shown, have a layered structure, with deeper levels more difficult to change than more surface levels. The deepest level of the mental model is genetically determined, but for the most part the mental model should be developed from the experience, and constantly reevaluated from the new experience. It is impossible for an individual to build the mental model solely through a direct experience: the world is far too complex for that, and chances of survival of such an individual until the time such a model will be reasonably effective would be minimal. Most of the knowledge required to build an effective mental model is thus transformed through indirect experience, through an ongoing communication with other individuals. For this reason, people communicating with each other (possibly through written sources) will have mental models that are similar in many significant aspects. These shared aspects become codified in culture. The basic components of the mental models are learned from parents in the early childhood and form individual's values, which permeate less basic attitudes and interpretative frameworks of the individual. Learning a language alone provides a person with important inputs for building a mental model since different aspects of language reflect significantly the mental models of cultural groups using it. Differences in shared mental models may affect how agents perceive their environment, what characteristics they take into account, and what criteria they use in choosing among options.

Culture provides legitimation for participants of social activity (individuals and organizations) as 'actors', that is, as sovereign and capable agents for themselves and others. As a rule, only legitimated elements are able to obtain resources and cultural, scientific, legal and administrative support from the wider society, which is crucial for their functioning. Culture defines in what types of activities particular actors can engage, and what strategies are appropriate for actors in pursuing these activities. Many empirical investigations have found that national culture influences many aspects of economic behavior and organization. In particular, culture

influences consumption patterns, types and models of organizations used, preferred coordination mechanisms in organizations, motivation of people working in organizations, the modes of entry of businesses into foreign countries and the chances of its success of such entries, the rate of innovation in the country (see Hofstede (1991, 2001) for a review of these influences), the rate of inflation, the degree of central bank independence (De Jong, 2002), the character of financial systems (Semenov, 2000; Stulz and Williamson, 2000; Licht e.a., 2001; De Jong and Semenov, 2002), the character of economic policy formation and implementation, the character of industrial relations (Semenov, 2000).

This has significant implications for economic theory. Since cultural characteristics are a major determinant of economic phenomena, we cannot obtain a real understanding of these phenomena without incorporating culture in the analysis. The standard assumptions currently used in economic theory may be reasonably correct for members of some societies, and we can explain empirically what happens in these societies. The analysis under standard assumptions, however, is unlikely to be sufficient for understanding economic life in societies for which this is not the case.

Till relatively recently, it was perceived as technically very difficult to incorporate values in economic analysis. If one took into account realistic complexities in human behavior and interaction, it had to be done at the expense of formal rigor, and thus remained at the margins of the discipline. However, the research in recent years indicates that technical problems involved are solvable, and such models can lead to interesting and important insights into human behavior.

This paper offers some thoughts on possibilities of incorporating cultural analysis into mainstream economic theory (Section 3), and, after introducing the cultural variables we use in the analysis (Section 4), considers some applications illustrating these possibilities (Sections 5-7). Section 8 concludes.

2 Cultural differences and economic theory: general considerations

Culture is “the collective programming of the mind” (Hofstede, 1980, p. 21). It influences peoples’ choices (and features of social organization formed through these choices) through their mental models. Thus, the most important way of incorporating cultural values in formal economic analysis is through modifying assumptions on preferences of agents.

Economists have long been suspicious of explaining economic phenomena by changing assumptions on preferences. It was argued that everything (and thus nothing) can be explained by assuming the ‘right’ preferences. It was also argued that preference-based explanations cannot be empirically tested. This has long been true. In particular, cultural explanations of economic phenomena were in the past often based on ‘common wisdom’ or personal impressions of researchers about cultures of different societies. This route to explaining behavior is understandably regarded as dangerously ad hoc (see e.g. Stigler and Becker, 1977). However, empirical work conducted in the last two of three decades seriously undermines the notion

that 'preferences explain everything and therefore nothing'. Research in cultural differences between countries has identified a small number of basic dimensions on which cultures differ, and a larger but still finite number of attitudes related to these dimensions. These dimensions have been found again and again in subsequent empirical research. Rather than dismissing preference-based explanations with their 'endless degrees of freedom', it is possible now to operationalize preferences as a short list of major deeply rooted values and related attitudes.

Moreover, new experimental techniques enable us to examine the nature of preferences in a scientifically rigorous way. Thus, there has been much progress in the empirical investigation of fairness preferences during the past decade. This research clearly shows that it is possible to discriminate between theories based on different preference assumptions (see e.g. Fehr and Schmidt (2000) for a survey).

This suggests that the methodological objections against incorporating cultural explanations in economic theory, which were valid two or three decades ago, can not be maintained anymore. In the view of theoretical and empirical arguments presented in the Introduction, incorporating cultural explanations in economic theory is important. How can it best be achieved?

I would suggest that such research may retain most of methodological assumptions and techniques of mainstream economic theory and fruitfully utilize its achievements to date. There could be two starting points from which this work may proceed. One is the empirical issues. The empirical facts not explained very well by existing theories, especially those related to cross-country differences in organization and behavior, may be addressed using cultural variables as potential explananda. Another starting point is the existing theory. In cases where one can expect that variations on empirically relevant cultural characteristics are likely to lead to variations in results, cultural characteristics can be incorporated into already existing economic models.

From the technical point of view, there are three main ways in which differences in values and attitudes can be incorporated into economic theory. First, formal analysis explicitly operationalizing cultural characteristics can be undertaken. It is not immediately obvious, of course, how cultural dimensions could be usefully operationalized. The work on operationalizing culture and using such operationalizations to obtain insights about economic phenomena may be a fruitful avenue of future research. Van den Brink, Ruys and Semenov (1999) and Ruys, van den Brink and Semenov (2000) suggest one possible way of operationalizing cultural characteristics (see the contribution of René van den Brink in this volume).

Second, given the fact that formal modelling may be difficult to accomplish in some cases (not only because of difficulties of operationalizing culture as such, but also because the full investigation of effects of culture may require the modelling of some aspects of social organization (see Section 5), and such modelling may prove difficult), non-formal theories can be build on the basis of research in economics and other social disciplines, and predictions based on them can be tested. Third, formal analysis without culture can be combined with non-formal cultural analysis. Thus, when a formal model has multiple equilibria, non-formal cultural analysis may be used to arrive at conclusions about which equilibrium is likely to be achieved in a particular type of societies. Aoki (1998) has made a similar suggestion, although

he proposed to use historical and institutional rather than cultural factors in such a non-formal analysis.

The rest of this contribution aims at taking first steps in and illustrating the approaches to incorporating national culture in economic theory outlined above. After introducing in the next section the cultural variables that would be used in this paper, I show how non-formal cultural analysis can identify differences across countries in attitudes that already have been incorporated in economic analysis, such as the degree of altruism, the degree of trust and the importance of status (Section 5), how non-formal cultural analysis can supplement formal analysis when in the latter multiple equilibria are obtained (Section 6), and how a model with formally introduced cultural characteristics can help in explaining cross-country differences in one of the characteristics of financial systems (Section 7).

3 Hofstede's cultural dimensions

As indicators of societies' cultural characteristics I use the cultural variables derived by Hofstede (1980) on the basis of a survey of work-related values, which was conducted in 1968-1973 and embraced about 117,000 native employees of IBM's subsidiaries in 40 countries. Hofstede identified four dimensions on which national cultures differ. A country could be given a score on each of the dimensions on the basis of scores on each of the orientations which reflect this dimension.

On the basis of the responses to a questionnaire Hofstede could identify many of values and orientations correlated with a particular dimension. Analysis of the work of other authors has shown that the results of many surveys on specific issues correlate with countries' scores on dimensions which a priori could be expected to be relevant. Thus, a range of additional values and orientations comprising a particular societal norm could be discerned. Below I briefly characterize the four cultural dimensions and list those of the associated beliefs and attitudes which are relevant for economic organization.

Power Distance refers to the extent to which the society accepts that power within its organizations and the society as a whole is distributed unequally. People in societies with a large Power Distance believe that inequality is existential: a majority of people perceives that there should be an order of inequality in which everybody has a rightful place, and such an order provides the best protection for everyone. People perceive that a few are (and should be) independent, but most people are and should be dependent; superiors consider subordinates as being of a different kind; subordinates also consider superiors as being of a different kind; other people are a potential threat to one's power and rarely can be trusted; cooperation is thus difficult to sustain. People in societies with a small Power Distance believe that hierarchy in society is an inequality of roles established for convenience; all people are and should be interdependent; superiors consider subordinates as people like them; subordinates also consider superiors as people like them; people of different power levels feel less threatened and more prepared to trust others and engage in cooperation with them.

Collectivism-Individualism reflects whether people look only after themselves

and their immediate family, or belong to in-groups which look after them in exchange for loyalty. People in collectivist societies have a ‘We’ consciousness; they emphasize belonging, for them collective interests prevail and group decisions are better than individual ones (not because of negation of individual interests, but because of the perception that individual interests could best be served through increasing the group’s well-being). Peoples’ involvement with organizations in such societies is more moral than calculative. In individualist societies people have an ‘I’ consciousness: for them individual interests prevail, individual decisions are better than group ones. People in such societies show universalistic attitudes and emphasize individual achievement (rather than belonging). Involvement with organizations is more calculative than moral.

Uncertainty Avoidance refers to the extent to which people feel threatened by uncertainty and ambiguity and try to avoid it and/or protect themselves against it. People in societies high on Uncertainty Avoidance perceive uncertainty inherent in life as a continuous threat than must be fought. Thus, achieving security provides a powerful motivator for them. People in these societies have a strong need of consensus; conflict and competition are felt undesirable. People in societies with a low Uncertainty Avoidance accept uncertainty inherent in life much more easily. Achievement is defined in terms of recognition by others (rather than security). Conflict and competition are perceived as natural patterns of conduct, which can be contained on the level of fair play and used constructively.

Femininity-Masculinity deals with a relative emphasize in society on achievement and success, on the one hand, and caring for others and quality of life, on the other. Feminine societies stress equality and solidarity and feel that needy should be helped; managers in these societies strive for consensus. Masculine societies stress competition and feel that strong should be supported; conflicts are resolved by fighting them out; managers make decisions on their own.

Later a fifth dimension — Long-Term Orientation — was identified from another survey. It refers to the extent to which a society exhibits a pragmatic future-oriented perspective or past- and present-perspective.

Examination of a large amount of studies from different disciplines allows Hofstede to show that the position of society on a given dimension has many important consequences in areas of family life, educational system, religious life, organizations, political system, dominant traditions in philosophy and dominant ideologies, and other aspects of life.

4 National culture, reciprocity, altruism, trust, and cooperation

Most of economic research assumes that preferences of economic agents are self-oriented and concern material goods. However, research adopting more extended preference functions, while (largely for the reasons outlined in Section 3) remained on the margins of the discipline, has a long tradition and has intensified and become more accepted in the last decade. The main reasons for this are the developments in game theory and experimental economics. Predictions of game-theoretic mod-

els based on players behaving in strictly self-interested and rational ways do not conform well with many observers' intuitions about how interactions take place in the real world and with the findings of experiments. Modifying the concept of rationality has helped to go some distance in reconciliation of theoretical predictions with empirical findings, but it proved to be insufficient; taking into account motivations of players beyond self-interest has also been found necessary. Three types of attitudes received a particular attention of researchers: reciprocity, altruism, and the importance of status and approval/respect from other people. The mentioned developments in game theory are the primary motivation for research taking the first two attitudes into account. The models taking into account status preferences help explain, in particular, the demand for fashion goods, the occupational choice of school-leavers, and the character of interaction of employees and employers in bargaining over wages.¹

All these models investigate single societies. The outcomes of these models depend on the relative importance for the agent of the attitude considered relative to other arguments in her utility function, or on the proportion of agents in the society for whom the attitude is important or is not important. Thus, if real societies differ along these lines, the predictions of the models for them will be different. For example, if in one society status is relatively more important to people than in another society, the models predict that the character of bargaining over wages will be different in these societies. However, do societies differ in the relative importance of reciprocity, altruism and status for agents? If yes, in which real societies are these attitudes more pronounced? We expect that societies do differ, and these differences have a lot to do with differences in deeply rooted cultural values.

We may expect people in more collectivist societies to show higher reciprocity towards other people in their collectivities than the reciprocity people in individualist societies show towards other people. What people consider as a 'collectivity' in this sense will differ across nations. In some nations the only collectivities playing an important role may be extended families, while in others colleagues, people living in the same village, members of the same profession, people of the same ethnic origin, or even all members of the society may be perceived as a 'collectivity'. Thus, in Japan collectivities such as firms with dozens of thousands employees, and industrial groups (where CEOs are a part of collectivity because they manage the firm which is a group member, and not because they have established close personal ties with all other CEOs), are possible, while in Taiwan it seems possible to form a collectivity only from at most a few dozens employees, or from entrepreneurs who know each other closely. The Individualism score alone does not allow to infer the importance of different types of collectivities. Furthermore, we can expect that 'positive' reciprocity (propensity to respond to actions perceived to be kind in a kind manner, even if this leads to material costs) will be higher in more feminine societies, with their relationship orientation, than in masculine societies, with their ego orientation.

Altruism for members of the nuclear family is probably a universal characteristic,

¹See the review of Fehr and Fischbacher (2002) for models incorporating reciprocity, the reviews of Rose-Ackerman (1996) and Hammond (1998) for models incorporating altruism, and the review of Weiss and Fershtman (1998) for models incorporating preferences for status and respect.

and the level of such altruism probably does not differ much across societies. People in more collectivist societies will be more altruistic towards other people in their collectivities than people in individualist societies will be towards other people, but people in more collectivist societies may be less altruistic towards non-members of collectivities than people in individualist societies will be towards other people. Altruism towards people in general is related to Femininity/Masculinity dimension: people in more feminine societies, with their stress on solidarity and support for the weak, are likely to be more altruistic.

It may be suggested that the extent to which the members of a society act cooperatively will be related to factors such as trust, the evaluation of costs of non-cooperation (the perceived necessity cooperate), and perhaps a general propensity to cooperate with other people.

To our knowledge only one paper (Zak and Knack, 2001) formally models trust. The authors assume that an inclination to cheat in the transaction is higher when social differences between agents (or 'social distance') is lower, and show formally that trust is negatively related to 'social distance'. While from a formal point of view this result is rather trivial, the authors take an important step in introducing social patterns and their implications into economic models. The empirically important question is exactly what social differences lead to less trust, and whether this differs across societies. A more general question is whether societies differ in the overall level of trust, and if yes, how can these differences be explained. We would expect that the answers are related to differences in basic values held in societies.

People in more collectivist societies will trust more to other people in their collectivities than people in individualist societies. This, however, does not extend to non-members of collectivities, to whom people in more collectivist societies may well be more distrustful than people in more individualist societies. A person in a society with higher Uncertainty Avoidance will trust less to people who have characteristics which differ from her own, and about which she knows little. People in societies with larger Power Distance perceive other people are a potential threat to one's power and have difficulty trusting them, while in societies with smaller Power Distance people of different power levels feel less threatened and more prepared to trust others.

As for the overall level of trust of a person towards other members of her society, we would expect that people in more feminine societies will a priori expect other members of their societies to have relationship orientation rather than ego orientation, and thus will trust them more. Given the preceding discussion, however, it is rather questionable if the overall level of trust of a person towards other members of her society is an empirically useful concept, because the degree of trust towards people with different characteristics will differ sharply. Nevertheless, there has been an attempt to measure this overall level through surveys (in particular, the World Value Survey; see Inglehart e.a., 1998), and the results have been used in empirical economic research (e.g. Keefer and Knack, 2000; Zak and Knack, 2002). We may expect that the overall level of trust will be higher in more collectivist societies, in societies with lower Uncertainty Avoidance, in societies with smaller Power Distance, and especially in more feminine societies. While there is not enough variance on the Collectivism/Individualism dimension in the sample to investigate the

issue, the predictions for other dimensions are confirmed: trust is highest in very feminine countries (there is a big gap in scores on this dimension between a group of Scandinavian countries and the Netherlands, on the one hand, and the rest of the countries, on the other hand, while scores on other dimensions are more evenly distributed). The level of trust decreases with increase in Uncertainty Avoidance and, for relatively high levels of Uncertainty Avoidance, with increase in Power Distance.

The costs of non-cooperation may be of several kinds. The open conflict imposes significant costs on the party which are avoided through compromise. Even if the competition does not result in an open conflict and parties negotiate the outcome on the basis of their relative power, the relationships between them become adverse. Thus, a potential for an open conflict and associated costs will be higher in future dealings with disputes on the same issue. Finally, suppose a party A is weaker on a particular issue than party B. However, party A may be stronger on other issues with which the parties will have to deal with in the future. If a compromise is achieved now on the issue at hand, party A may promise to reciprocate by compromising on these future issues. Thus, if party B refuses to compromise it bears the costs of losing the advantages from compromises on future issues. Whether the promise of party A is credible will, however, depend on the level of trust between the parties.

Many factors would influence the 'true' size of the costs of non-cooperation and the perception of the magnitude of these costs by agents. In particular, agents in different countries may perceive these costs differently because of differences in their mental model induced by cultural differences. In a collectivist societies such costs are likely to be perceived as very high when dealing with people from the same collectivity. In an uncertainty avoiding society a majority of people have a strong need for consensus and seek to avoid conflict and competition (which may unleash aggression and in general increases uncertainty), while in a society low on Uncertainty Avoidance it is believed that conflict and competition is natural, can be contained at the level of fair play and used constructively. In an uncertainty avoiding society people also worry more about the future, while in a society low on Uncertainty Avoidance they are more prepared to live day by day. This implies a greater readiness to sacrifice immediate gains on specific issues if this may increase stability, or bring benefits in the future on some other, perhaps more important issues, and there is a reasonable assuredness that these benefits will be forthcoming.

Finally, a general propensity to cooperate may differ with cultural differences. In particular, people in more feminine societies, which have relationship orientation, may have a higher general propensity to cooperate than people in more masculine societies, which have ego orientation.

Investigation of these issues is likely to be a fruitful direction of economic research.

Another attitude which captured the interest of economists and which has been included as a component of preferences in formal models is the importance of status, of respect of other people. The importance of status, of respect of other people will be influenced by deeply-rooted cultural values. Status is more important to people in societies with larger Power Distance than to people in societies with smaller

Power Distance. People in more uncertainty avoiding societies define achievement in terms of achieved security, while people in less uncertainty avoiding societies define achievement in terms of recognition by others. Status and respect of others will thus be more important to them. People in more collectivist societies will likely care more about their status in their collectivities and respect from members of these collectivities, than people in individualist societies will care about their status, but people in more collectivist societies may care less about their status with, and respect by, non-members of collectivities, than people in individualist societies will care about their status.

This discussion indicates that for a thorough understanding of economic organization and performance it is important to include the characteristics of social organization in the model, to model explicitly how social organizations interact in economic process, how people interact with people within and outside organizations. This has been recognized for some time already, independently of cultural differences. The fact that the behavior towards people with different social characteristics is influenced by culture in different ways underscores this.

5 National culture and legislative protection of investors and employees

This section considers an application in which non-formal cultural analysis supplements formal analysis. In particular, I consider a formal model that has multiple equilibria, and use cultural analysis to predict which equilibrium is likely to be arrived at in particular (real) societies. The model is that of Pagano and Volpin (2000), who analyze the process of adoption of legislative provisions protecting investors and employees. There are three classes in the economy - rentiers, workers and entrepreneurs. All agents have initial wealth endowments, equal for all members of the same class. Workers also have (equal) endowments of labor time, and entrepreneurs have human capital which is essential to set up a firm. The entrepreneurs use their wealth endowments to set up firms, which operate for two production cycles. Rentiers and workers can invest their wealth endowments in shares of firms or in a riskless asset. All firms are identical and require an investment of capital larger than the initial endowment of an entrepreneur. Entrepreneurs obtain private benefits from the company by diverting a part of the cash flow, which reduces the value of the firm for shareholders. Entrepreneurs maximize the sum of their final wealth and private benefits. Rentiers maximize their final wealth. Workers' utility function is $U_W = c - u(l_1) - v(l_2)$, where c is wealth and l_i is the time spent working in the corresponding production cycle. A crucial assumption of the model is that disutility of work at the same firm in the second production cycle is smaller than in the first production cycle: $u(l_1) - v(l_2) = \varepsilon > 0$.

Workers are hired for two production cycles, but are paid wages after each cycle, and their contracts can be renegotiated (or terminated) after the first cycle. External capital can be obtained by entrepreneurs only by selling equity stakes to rentiers and workers, which are held for two periods. The fractions of shares owned by rentiers, workers and entrepreneurs are denoted correspondingly β_R, β_W and β_E ,

with $\beta_R + \beta_W + \beta_E = 1$. Dividends are paid at the end of the second period. The output in each period is deterministic.

After the first production cycle is completed, a technological innovation takes place which makes some workers more productive. The workers that worked at the firm at the first production cycle, however, remain as productive as they were. The entrepreneurs thus have to fire them and hire new workers to take advantage of the innovation. Such restructuring is always profitable.

After the contracts for the first production cycle are signed, legislation is introduced on the degree of investor protection and the degree of employment protection. Investor protection decreases the private benefits the entrepreneur obtains from diverting a given amount of cash flow. More precisely, these benefits are represented by a function $B = B(D, \lambda)$, where D is the amount of cash flow diverted and λ is the degree of investor protection, with $B_\lambda < 0$ and $B_{D\lambda} < 0$. (D is an endogenous variable in the model, which in particular depends on β_E and λ .) Employee protection makes it more difficult for firms to fire workers. In particular, the law can determine the fraction f of workers that can be fired after the first period. The legislation is adopted by a majority vote. It is assumed that no single class has the majority, while any two classes together have the majority of votes.

The authors show that rentiers will always be in favor of greater investor protection and of as little employee protection as possible. Entrepreneurs will be in favor of as little investor protection and as little employee protection as possible. Workers will always be in favor of greater investor protection, and will be in favor of greater employee protection, unless they hold a large amount of equity and/or unless the difference in disutility of work between the two periods is large enough, more precisely if $\varepsilon > \beta_W \Delta$. From this it is obvious that if the two issues are voted separately, these votes would result in maximum investor protection and no employee protection ($\lambda = 1, f = 1$). Two classes, however, can commit to a joint vote on both issues. In case such commitment is made, and if

$$\varepsilon > \left(1 + \frac{\beta_E}{\beta_W} \frac{D_\lambda(\beta_E, 1)}{B_\lambda(\beta_E, 1)} \right) \cdot \beta_W \Delta, \quad (1)$$

employees and workers both can get higher utility if some employee protection, and a lesser than maximal degree of investor protection, are chosen ($\lambda < 1, f < 1$). Since the two classes have a majority, this would be the outcome of the political process.

Thus, two political outcomes are possible in this model: either workers and entrepreneurs strike a deal in which low investor protection is exchanged for high labor protection, or no such deal is made and the legal regime features high investor protection and low employee protection. Using a sample of 21 countries, the authors show that the extent of labor protection and the extent of investor protection are indeed inversely related empirically .

But why some countries end up with high investor protection and low employee protection, and others - with low investor protection and high employee protection? In the model the outcome depends on three parameters: the difference in the disutility of work in the first and the second period, ε , the proportion of shares held by workers, β_W , and way political decisions are arrived at, in particular the propensity

of political agents to cooperate.

These parameters are crucially influenced by cultural characteristics of countries. First, from the theoretical point of view, the propensity of people to stay in the same job and thus to commit to the job, to enter implicit contracts with management and make firm-specific investments is likely to be higher in societies with higher Uncertainty Avoidance and (to a lesser extent) with lower Masculinity. This is confirmed empirically: average employee tenure with one firm is higher in these societies (see Semenov, 2000). The propensity of employees to stay at the same job and commit to the firm is clearly related to ε in the model.

Second, in societies where people have higher Uncertainty Avoidance (small investors are likely to have a lower propensity to invest in stock (rather than in safer mediums) than in societies with lower Uncertainty Avoidance. This propensity will, of course, influence β_W in the model.

Third, the differences among countries in the way decisions on economic policy are arrived at are related to cultural differences. Countries where labor does not participate in the process of economic and social policy formation and implementation, the intensity of labor-capital interaction is low, and industrial order is conflicting, have cultural characteristics different from countries where organized interests actively participate in the process of economic and social policy formation and implementation, the intensity of labor-capital interaction is high, and industrial order is peaceful. This can be accounted for theoretically. In particular, different cultural characteristics induce a different attitude to various mechanisms of handling disputes: 'fighting them out' or cooperation and searching for compromise. The latter option would be more appealing to people in societies with a higher Uncertainty Avoidance, societies with lower Masculinity and societies with smaller Power Distance (see Semenov, 2000). This would, of course, influence the likelihood of cooperation of political agents.

The direction of influence of all these factors is generally the same: in societies with higher Uncertainty Avoidance and in societies with lower Masculinity the propensity to stay at the same job is higher, and the propensity of political agents to seek cooperation is higher. In addition, in societies with higher Uncertainty Avoidance small investors are more reluctant to invest in stock. Thus, in societies with higher Uncertainty Avoidance and in societies with lower Masculinity the outcome with a stronger employee protection and a weaker investor protection is more likely according to the model.

Empirically, it can be shown (see Semenov, 2000) that the level of employment protection in developed Western countries is related positively to the level Uncertainty Avoidance and (to a lesser extent) negatively to the level of Masculinity. On the other hand, the level of shareholder protection is related negatively to the level Uncertainty Avoidance and positively to the level of Masculinity.

6 National culture and the character of relationships of firms and investors

In this section I present a model of choice of contractual forms by entrepreneurs and investors, where differences in one of agents' characteristics related to culture - the degree of uncertainty aversion - are explicitly formalized and play an important role in determining the outcome.

Interests of shareholders diverge in several important ways from interests of creditors. First, if two projects are identical in their expected return but differ in their riskiness, shareholders prefer the riskier project with identical expected returns, since in case of success they obtain all the benefits but in case of failure part of the losses is borne by creditors. This choice is undesirable for creditors because it increases the probability of their losses on the project. Second, shareholders (and managers acting in their interests) will not apply efforts to improve the returns in bankrupt states because the increased returns will accrue to creditors. Third, when a risk of bankruptcy is present, shareholders may pay excessive dividends and thus reduce potential assets available for creditors' compensation, thereby redistributing wealth from the creditors to themselves. Fourth, shareholders will have an incentive to issue a new debt senior to the existing one, thus increasing the risk for the existing creditors (Jensen and Meckling, 1976).

These agency problems between creditors and shareholders can be reduced in three ways. First, shareholders can make an obligation to transfer the rights on some assets to the creditors in case they do not repay the debt, that is, provide a collateral. Second, shareholders can agree on establishing mechanisms which will allow creditors to obtain extensive information about the firm, monitor shareholders' and managers' decisions, and perhaps give creditors some rights in making these decisions. Third, shareholders (or managers) can convince creditors that they will further creditors' interests, because they have always done so in the past, or because shareholders belong to the same social group as creditors, and this group possesses sufficient mechanisms of social sanction to make shareholders honor their obligations. While the third mechanism had been very important historically and may be important in many countries at present, it can play only a supplementary role in modern developed countries, and agents rely mostly on the first two mechanisms. In each particular transaction one of these two mechanisms plays a more important role. When the first mechanism predominates, creditors distance themselves from the firm as long as the results are satisfactory, and appropriate collateral if they are not. Thus, the relationship of firms and borrowers is arm's-length. When the second mechanism predominates, creditors and borrowers are engaged in a close relationship.

While both of these mechanisms are used in all developed countries, in each of them a reliance on one of the mechanisms predominates (see Semenov, 2000). Thus, in Germany banks provide companies with long-term capital, underwrite equity issues, control a large share of equity in many corporations, hold positions on boards of directors, often take an active role in corporate management when dissatisfied with corporate policy, and organize and manage corporate restructuring. Market for corporate bonds (where arm's-length relationships of investors and firms

predominate) is very thin. In contrast, in the USA relationships between banks and large firms throughout the post-war period have been arm's-length. Firms typically borrow from a large number of banks. Banks are not involved in ownership or management changes in the firms (except in the case of financial crisis). Market for corporate bonds is well developed.

Why have these differences emerged and why do they persist? In my Ph.D. thesis I offer a (non-formal) theoretical explanation based on cultural characteristics of the countries, which is confirmed by empirical evidence. Here I present a simple model formalizing one of the arguments of this analysis. Before I present the model, I introduce the concept of uncertainty aversion which plays a significant role in the model.

6.1 Uncertainty aversion

The theory of uncertainty aversion attempts to generalize the expected utility theory by distinguish between quantifiable 'risks' and unknown 'uncertainties'. The theory of uncertainty aversion argues that people do not necessarily behave as if they have reduced all ambiguity to a belief represented by a unique probability distribution. It further argues that people are uncertainty averse: the more ambiguous is their information the more conservative is their choice. The model of decision making incorporating uncertainty aversion was pioneered by Gilboa (1987), Schmeidler (1989) and Gilboa and Schmeidler (1989). The theory has provided fresh insights in, among others, asset pricing and portfolio choice (Dow and da Costa Werlang, 1992; Tallon, 1998), incompleteness in financial markets (Mukerji and Tallon, 2001), strategic interactions in game theory (Dow and da Costa Werlang, 1994) and contract theory (Mukerji, 1998).

The theory has suggested two closely related (and for a large class of situations equivalent) representations of decision making under uncertainty. First, the decision-maker's subjective beliefs can be represented by a non-additive probability function. Let $\Omega = \{s_i\}_{i=1,\dots,S}$ be a set of states and let Σ be the algebra of this set. A function $P : \Sigma \rightarrow [0, 1]$ is a nonadditive probability if (1) $P(\emptyset) = 0$, (2) $P(\Omega) = 1$, (3) $A \subset B \Rightarrow P(A) \leq P(B)$, and (4) $\forall A, B \in \Sigma : P(A \cup B) + P(A \cap B) \geq P(A) + P(B)$. The property (4) distinguishes the nonadditive probability from a standard probability, which is characterized by the equality in (4). The second representation suggests that the decision-maker evaluates the expected utility for a set of prior probability distributions and acts to maximize the minimum of expected utility over these priors. Let $\mathcal{F} = \{f_i\}_{i=1,\dots,F}$ be the set of actions, $\mathcal{X} = \{X^j\}_{j=1,\dots,F}$, $X^j = \{x_1^j, \dots, x_S^j\}$ be the set of outcomes for the agent, $\mathcal{P} = \{P^k\}$, $P^k = \{p_1^k, \dots, p_S^k\}$ be the set of probability distributions, and $u : \mathbb{R} \rightarrow \mathbb{R}_+$ be the agent's utility function. Then the uncertainty averse agent chooses the action maximizing

$$\max_{\mathcal{F}} \min_{\mathcal{P}} \sum_i p_i^k u(x_i^j)$$

As an example, consider a risk-neutral investor with utility function $u(W) = W$, where W is final wealth. The investor contemplates two actions: (1) investing a in some project which may be either success (bringing the return R) or failure

(bringing nothing), (2) investing in a riskless asset with the rate of return $r < R$. The investor does not know a precise probability distribution of these outcomes. She perceives the probability of success as lying within an interval $[p_1, p_2]$. The uncertainty averse investor will invests in the project if $p_1 a(1 + R) > a(1 + r)$, and in the riskless asset otherwise.

This rule reflects an extreme degree of uncertainty aversion. A less assumption would be that an agent can assign relative weights to alternative probability distributions (reflecting the available information, his past experience and intuition), thus arriving at a composite probability distribution \tilde{p} . The confidence in these weights, however, will depend on her judgement about the reliability, credibility and adequacy of her information. These judgements cannot be expressed in terms of relative likelihoods. Out of all possible probability distributions ($[0, 1]$ in our example) she rejects some as unreasonable. The set $[p_1, p_2]$, however, still seem reasonable, reflecting the situations that she cannot on the basis of her information rule out, judgements that she 'might almost as well' have made. For this reason, she will also take into account the worst possible outcome. The weight z she, in making her decision, assigns to the worst possible outcome, as opposed to the 'composite' distribution she arrived at, will depends on her confidence in judgements she made. Thus, in our example, the investor will choose the risky project if

$$(zp_1 + (1 - z)\tilde{p})a(1 + R) > a(1 + r) \quad (2)$$

More generally, the decision maker will maximize

$$\max_{\mathcal{F}} \left(z \min_{\mathcal{P}} \left(\sum_i p_i^k u(x_i^j) \right) + (1 - z) \sum_i \tilde{p}_i u(x_i^j) \right)$$

This has been suggested by Ellsberg (1961) as an ad hoc decision rule for interpreting some experimental evidence; Dow and Werlang (1992) provide a theoretical rationale for this rule.

It can be suggested that z would reflect not only the degree of confidence of an agent in her information, but also the degree of her uncertainty aversion. We can expect that the degree of uncertainty aversion of agents will be larger in more uncertainty avoiding societies.

In the next subsection I apply this idea to a simple model of relationships between entrepreneurs and investors.

6.2 The model

The model is largely based on Holmstrom and Tirole (1997). An entrepreneur is endowed with a project, which requires the initial investment I . The entrepreneur's wealth W is smaller than I so he has to obtain external funds to realize the project. The project with probability p is successful, bringing Y ; if it is a failure it brings nothing. The entrepreneur and the investors can invest their money either in the project or in a riskless asset, bringing a return which is normalized to zero. The project has a positive net present values: $pY > I$. Both entrepreneurs and lenders are risk-neutral. The supply of loanable funds to investors is unlimited at some exogenously given interest rate which will be normalized to zero. Investors offer debt

contracts, and the entrepreneur accepts the most favorable offer. Through competition, this offer generates zero expected profits to investors. Entrepreneurs know the probability of success of their projects. Investors do not know this probability for certain, but know that it lies in the interval $[p_l, p_h]$, where $p_l < p_h$.² Investors assign relative weights to alternative probability distributions (reflecting the available information, past experience and intuition), thus arriving at a composite probability distribution \tilde{p} . I assume, to simplify exposition, that this probability is the true probability of success: $\tilde{p} = p$. Relaxing this assumption does not change the results. Investors can, at a cost m , monitor the firm and learn the true probability of success of the project. I adopt the following assumptions:

Assumption 1. Monitoring costs are not so large as to make the project non-viable: $m < pY - I$.

Assumption 2. Investors have good enough information and experience to correctly estimate in their composite probability distributions \tilde{p} that the projects are viable: $\tilde{p}Y > I$.

Investors are uncertainty averse. In making their decisions they use the decision rule represented by (2). In a given economy all investors have the same degree of uncertainty aversion z ; investors in different economies may differ in the degree of uncertainty aversion, reflecting cultural influences.

Now I describe the entrepreneur's choices. Consider informed finance first. The investors' participation constraint is $p(1+r)(I-W) \geq I-W$, which implies that it will ask the interest rate $r = 1/p - 1$. The entrepreneur's participation constraint is $p(Y - (1+r)(I-W) - m) \geq W$. Substituting r from the investor's participation constraint produces $m < pY - I$. Thus, the entrepreneur will undertake projects under informed finance rather than invest in a riskless asset, provided that the monitoring costs are small enough. Assumption 1 ensures this to be the case in our model.

Consider now uninformed finance. The investors' participation constraint is $(zp_l + (1-z)p)(1+r)(I-W) \geq I-W$, which implies that it will ask the interest rate $r = 1/(zp_l + (1-z)p) - 1$. It is easy to see that r is increasing function of z , the degree of uncertainty aversion. The entrepreneur's participation constraint is $p(Y - (1+r)(I-W)) \geq W$. Substituting r from the investor's participation constraint and solving for W produces

$$W \geq pY - \frac{p(pY - I)}{(p - p_l)z} \equiv \tilde{W}$$

Thus, the entrepreneur will undertake the project under uninformed finance, rather than invest in a riskless asset, only if he has a large enough initial endowment ($W > \tilde{W}$).

²A similar setup is used, e.g., by Blazenko (1987). We assume, to simplify exposition, that the limits of this interval are the same for all projects probability is the true probability of success. Relaxing this assumption does not change the results.

Entrepreneurs will prefer informed over uninformed finance if the required repayments to the investors in this case are lower, that is, if

$$\frac{I - W + m}{p} > \frac{I - W}{zp_l + (1 - z)p}$$

or

$$W < I - m \left(\frac{p}{(p - p_l)z} - 1 \right) \equiv \hat{W}$$

It is easy to see that if investors are not uncertainty averse ($z = 0$), for any initial endowment the entrepreneur will prefer to undertake the project under uninformed finance.

Differentiating \tilde{W} by z we obtain

$$\frac{\partial \tilde{W}}{\partial z} = \frac{p(pY - I)}{(p - p_l)z^2} > 0$$

This implies that with increase in uncertainty aversion of investors less entrepreneurs will prefer to undertake the project using uninformed finance compared with investing in a riskless asset. Differentiating \hat{W} by z we obtain

$$\frac{\partial \hat{W}}{\partial z} = m \frac{p}{(p - p_l)z^2} > 0$$

Thus, with increase in uncertainty aversion of investors less entrepreneurs will prefer to undertake project under uninformed finance, and more will prefer to undertake it under informed finance.

This analysis indicates that entrepreneurs with $W < \max\{\tilde{W}, \hat{W}\}$ will choose informed finance to finance their projects, while entrepreneurs with larger endowments will choose uninformed finance. Both these critical values are higher when the degree of uncertainty aversion of investors is higher. This implies that if two societies have the same distribution of projects by the probability of success and the same distribution of entrepreneurs by initial endowments, but investors in the two societies have different degrees of uncertainty aversion, in the society where investors are more uncertainty averse less entrepreneurs will prefer to undertake project under uninformed finance, and more will prefer to undertake it under informed finance.

To get an idea about the magnitudes involved, suppose that $m = 0.01I$. To guarantee that all entrepreneurs choose uninformed finance, \tilde{W} should be non-positive. This implies $\frac{p}{z(p - p_l)} \geq 99$. This implies either a very precise knowledge of probability distribution by investors (very low $p - p_l$) or very low degree of uncertainty aversion. Thus, if $p = 0.7$ and $p - p_l = 0.2$, z can be at most slightly larger than 0.035. On the other hand, to guarantee that all entrepreneurs choose informed finance, \hat{W} should be greater or equal to I . This implies $\frac{p}{z(p - p_l)} \leq 1$. The only case when it is possible is when investors have maximum uncertainty aversion ($z = 1$) and always expect that there is a possibility that a project is a failure with probability 1, that is, $p_l = 0$. This situation, of course, is not very realistic. This

suggests that in most cases \hat{W} would lie in between 0 and I , that is, that some entrepreneurs will chose informed, and some - uninformed finance.

This analysis suggests that close relationships of investors and firms are likely to be more widespread in societies with higher Uncertainty Avoidance than in societies with lower Uncertainty Avoidance. This is indeed the broad pattern we observe in practice (although the empirical pattern is also influenced by of values related to other cultural dimensions; see Semenov (2000)).

The model, without modifications or only with minor modifications, can be given other interpretations. I mention two of them. First, it can be assumed that entrepreneurs cannot invest their endowments, but can offer them as collateral, giving a link to a large literature on the role of collateral. Second, instead of assuming that investors do not know the probability of success, we can assume that they are uncertain about the propensity of the entrepreneur to divert cash flow to obtain private benefits, or about the extent of effort aversion of the entrepreneur. The models thus becomes that of a moral hazard.

7 Conclusion

This paper argues that without incorporating the diversity of national cultures in economic analysis, it is impossible to achieve a real understanding of many economic phenomena, and practical recommendations in many cases may be counterproductive. It also argues that methodological objections against incorporating cultural explanations in economic theory, which were valid two or three decades ago, can not be maintained anymore. The paper suggests some ways of incorporating cultural analysis into economic theory, and illustrates several of them. I argue that incorporating cultural differences by modifying assumptions on preferences may be a fruitful way of going forward. These should not be ad hoc modifications but should be based on empirically obtained cultural characteristics. More generally, I suggest that the economic research that takes cultural factors into account may retain most of methodological assumptions and techniques of mainstream economic theory and fruitfully utilize its achievements to date.

Other authors may come with different suggestions. Which approaches prove the most promising will be known only as the analyses of the kinds illustrated in this paper are further developed, criticized, and tested. The important point is that such efforts have to be made if economic theory is to explain the mechanisms governing economic transactions in various countries.

References

- Aoki M., 1998, "The Subjective Game Form and Institutional Evolution as Punctuated Equilibrium", Paper presented at the Second Conference of the International Society for New Institutional Economics.
- Blazenko G.W., 1984, Managerial Preference, Asymmetric Information, and Financial Structure, *Journal of Finance*, 42, 839-862.

- Brink R. van den, P.H.M. Ruys and R. Semenov, 1999, Governance of Clubs and Firms with Cultural Dimensions, Tilburg University, Center for Economic Research, Discussion paper.
- Dow J. and S.R. da Costa Werlang, 1992, Uncertainty Aversion, Risk Aversion, and the Optimal Choice of Portfolio, *Econometrica*, 60, 197-204.
- Dow J. and S.R. da Costa Werlang, 1994, Nash Equilibrium under Knightian Uncertainty: Breaking Down Backward Induction, *Journal of Economic Theory*, 64, 305-324.
- Ellsberg D., 1961, Risk, Ambiguity and the Savage Axioms, *Quarterly Journal of Economics*, 75, 643-669.
- Fehr E. and U. Fischbacher, 2002, Why Social Preferences Matter - The Impact of Non-selfish Motives on Competition, Cooperation and Incentives, *Economic Journal*, 112, C1-C33.
- Fehr E. and K.M. Schmidt, 2000, Theories of Fairness and Reciprocity: Evidence and Economic Applications, in: M. Dewatripont, L.P. Hansen, S. Turnovsky, eds., *Advances in Economic Theory*, Cambridge, Cambridge University Press.
- Gilboa I., 1987, Expected Utility Theory with Purely Subjective Non-Additive Probabilities, *Journal of Mathematical Economics*, 16, 65-88.
- Gilboa I. and D. Schmeidler, 1989, Maxmin Expected Utility with a Non-Unique Prior, *Journal of Mathematical Economics*, 18, 141-53.
- Hammond P.J., 1998, Altruism, in: J. Eatwell, M. Milgate and P. Newman, eds., *The New Palgrave: A Dictionary of Economics*, London, Macmillan, pp. 85-86.
- Holmstrom B. and J. Tirole, 1997, Financial Intermediation, Loanable Funds, and the Real Sector, *Quarterly Journal of Economics*, 112, 663-692.
- Hofstede G., 1984, *Culture's Consequences*, Beverly Hills, Sage.
- Hofstede G., 1991, *Cultures and Organizations*, London, McGraw-Hill.
- Hofstede G., 2001, *Culture's Consequences*, 2nd edition, Beverly Hills, Sage.
- Inglehart R., M. Basañez and A. Moreno, 1998, *Human Values and Beliefs: A Cross-Cultural Sourcebook*, Ann Arbor, University of Michigan Press.
- Jensen M.C. and W.H. Meckling, 1976, Theory of the Firm, Managerial Behavior, Agency Costs and Ownership Structure, *Journal of Financial Economics*, 3, 305-60.
- Jong E. de, 2001, Why Are Price Stability and Statutory Independence of Central Banks Negatively Correlated?, Groningen University, SOM Research report.

- Knack S. and P. Keefer, 1997, Does Social Capital Have an Economic Payoff? A Cross-Country Investigation, *Quarterly Journal of Economics*, 112, 1251-1288.
- Licht A.N., C. Goldschmidt and S. H. Schwarz, 2001, Culture, Law, and Finance: Cultural Dimensions of Corporate Governance Laws, mimeo.
- Mukerji S., 1998, Ambiguity Aversion and Incompleteness of Contractual Form, *American Economic Review*, 88, 1207-1231.
- Mukerji S. and J.-M. Tallon, 2001, Ambiguity Aversion and Incompleteness of Financial Markets, *Review of Economic Studies*, 68, 883-908.
- Pagano M. and P. Volpin, 2000, The Political Economy of Corporate Governance, CEPR Discussion paper.
- Rose-Ackerman S., 1996, Altruism, Nonprofits, and Economic Theory, *Journal of Economic Literature*, 34, 701-728.
- Ruys P.H.M., R. van den Brink and R. Semenov, 2000, Values and Governance Systems, in: C. Menard (ed.), *Institutions, Contracts and Organizations: Perspectives from New Institutional Economics*, Cheltenham, Edward Elgar, pp. 422-446.
- Schmeidler D., 1989, Subjective Probability and Expected Utility without Additivity, *Econometrica*, 57, 571-87
- Semenov R., 2000, *Cross-Country Differences in Economic Governance: Culture as a Major Explanatory Factor*, Ph.D. thesis, Tilburg University.
- Stigler G and G. Becker, De Gustibus Non Est Disputandum, *American Economic Review*, 67, 76-90.
- Streeck W., 1997, Beneficial Constraints: On the Economic Limits of Rational Voluntarism, in: J.R. Hollingsworth and R. Boyer, eds., *Contemporary Capitalism: The Embeddedness of Institutions*, Cambridge, Cambridge University Press.
- Stulz R.M. and R. Williamson, 2001, Culture, Openness, and Finance, NBER working paper.
- Tallon J.-M., 1998, Asymmetric Information, Nonadditive Expected Utility, and the Information Revealed by Prices: An Example, *International Economic Review*, 39, 329-342.
- Weiss Y., Fershtman C., 1998, Social Status and Economic Performance: A Critical Analysis, *European Economic Review*, 42, 801-820.
- Zak P.J. and S. Knack, 2001, Trust and Growth, *Economic Journal*, 111, 295-321.

THE RIVER, THE STONES AND THE GRADUALIST REFORM IN CHINA'S BANKING SECTOR*

Yuan Ju[†]

Tilburg University

Before the first meeting with Pieter Ruys in April 2000 when he formally visited Shandong University, China, we had already kept up correspondence with each other for nine months as I participated in the joint research project *Competition, Regulation, and Reform of State-owned Enterprises and National Utilities* (sponsored by the KNAW and the Ministry of Education, China) between the Department of Econometrics of Tilburg University and the School of Economics of Shandong University and was in charge of a sub-project “Competition and Regulation in the Banking Sector in China”. I still clearly remembered Pieter’s enthusiastic and encouraging words for my research on that topic in the first letter from him in 1999. Since then, his continuous direction and support to me has expanded but not limited to study, life, Ph.D. research and other personal or social affairs. To be honest, I not only met an excellent supervisor but also got a good friend despite a large difference in age. In respect of scientific research, it is Pieter who has brought me to the frontier of economics, which at least can be seen from the following three turning points. The first one is about the above research project. From then on, I made the resolution to devote to academic research. Secondly, he suggested me to come to Tilburg and induced me into the world of modern economics by discussing the new developments in Mathematical Economics, Game Theory, New Institutional Economics, Public Sector Economics, Network Economics, etc. Together with the papers, books, seminars and conferences he suggested to me, I was truly moved by and absorbed into the beauty of economics. Once again, under the influence and supervision of him, my Ph.D. research focuses on the economics of *cooperation and compensation*, which opened a new door for my further progress. I am greatly indebted to Pieter for the excellent supervision and gratefully appreciate his wisdom and kindness. To my pleasure, I have the opportunity to dedicate this paper to him on the occasion of his 65th birthday and retirement celebration.

*I thank Pieter Ruys, Liangchun Yu and Henk van Gemert for helpful conversations and comments on the related versions of the paper. Of course, all the remaining errors or deficiencies of the paper are mine.

[†]*Adres:* CentER and Department of Econometrics and Operations Research, Tilburg University, Warandelaan 2, P.O.Box 90153, 5000 LE, Tilburg, the Netherlands. *E-mail:* Y.Ju@uvt.nl

Although currently my research interests and focus shift to Game Theory and Network Economics, this contribution still relates to the starting point of my academic career, which was also the beginning of my friendship with Pieter. This short paper presents my thoughts on the link in terms of approach and strategy between the Banking sector reforms and the whole economic transition in China. By briefly retrospectively and analyzing the reform measures and process in the banking sector—a miniature of the whole economic reform, the paper provides a window to look into the basic ideas and features of China’s reform and transition: gradualism in process, experimentalism in method, pragmatism in attitude and evolutionism in nature.

“Crossing the river by touching the stones.”

Deng Xiaoping, Chinese leader

1 Introduction

The transition from plan to market in former planned economies is one of the main economic events of the 20th century. Not only does it affect the lives of approximately 1.65 billion people, but it is contributing to a shift in emphasis in economics from standard price and monetary theory to contracting and its institutional environment. In particular, the gradualist reform and transition process in China issue ample challenges for economic research (see Chow (1997)). To see how reform worked in China is not only necessary for understanding the reality of the world economy but also helpful to make theoretical explorations for researchers. However, it is almost impossible to comprehensively study and answer this “big” problem by writing a paper only. Therefore, this short essay can not give a full view about China’s economic reform, but will focus on the banking sector to partially illustrate the measures of the reform and the process of the transition so as to show exactly how the reform goes along a unique way and reveal the ideas and features behind that.

Why I choose the banking sector? The main reason lies in its importance in the economy and the urgency of reforming. The bottom-up approach and the sequence of the economic reform — rural reform → enterprise/city reform → financial reform — resulted in a long-term transition period and the fact that the later, the harder, more and more challenges, even difficulties have crowded into the road to further development, such as unemployment, deflation, stagnating in the reform of the stated-owned enterprises (SOE) and financial reform, widening wealth disparity, disparity in district developments, and so on since the mid-1990s. The process seems not Pareto-neutral any more. As the crucial component and late-reformed¹ part of the economy, nowadays, the banking sector attracts more attention. People expect the reform in this sector will change its low efficiency and huge amount of bad debts or non-performing loans (NPLs), and then foster the economic development.

¹The related indicators of structural reform and institutional quality can be found in EBRD Report 1999 and IMF Report 2000.

Another reason for being chosen is that the strong representativeness of the sector helps to portray the profile of the reform. Besides, the clear boundary of the sector makes it easier to analyse and grasp the essence of the problem.

Much work has been done in the field of China's reform and transition. Among others, some important works which may closely relate to this paper are as follows. Roland (2000) provides an overview of current research raised by transition and concludes the transition economics in his book *Transition and Economics*, especially, of which the third part looks at the economic behavior of firms in the transition from state to private ownership and compares the effects of privatization, restructuring, and financial reform. Dewatripont and Roland (1992) analyze the virtues of gradualism and legitimacy in the transition and extensively examine the impact of political constraints on economic reform plans in the next paper in the same year. Some economists regard China's gradualist approach as self-defeating (Murphy, Schleifer, and Vishny (1992)). Feltenstein and Nsouli (2001) empirically analyzed the strategy of gradualism in China. Other economists argue that it is neither gradualism nor experimentation but rather China's unique initial conditions—namely, a large agricultural labor force, low subsidies to the population, and a rather decentralized economic system—that have contributed to China's success (Woo (1993); Sachs and Woo (1993); Qian and Xu (1993)). Moreover and specifically, Lau, Qian and Roland (2000) develop a simple model to analyze the dual-track approach to market liberalization as a mechanism for implementing efficient Pareto-improving economic reform. To further understand how reform works in a developing and transition economy that has great growth potential, Qian (2001) suggests that one should study how feasible, imperfect institutions have evolved to complement the initial conditions and to function as stepping stones in the transition toward the goal. However, all those studies did not put emphasis on the general and basic ideas and features of China's reform but only give a partial or incomplete and sometimes even incoherent conclusion in this respect. In this paper, I offer a fundamental but comprehensive perspective on the topic although only relying on a case study in the banking sector. As for the analysis about the reform in China's banking sector, especially from the market structure point of view, few work can be found except some papers in Chinese which are helpful, e.g. Yu and Ju (1999, 2000), Pang and Ju (2000).

In addition to this section introducing the purpose of the paper and reviewing the literature briefly, the remaining part is structured as follows. The next section presents a retrospectus about the reforms in the banking sector and provides a timeline of the market evolution. Based on explaining the issue of “the river and the stones” and showing the consistency between the banking sector reform and the whole economic transition, the third section discusses and summarizes the basic ideas and features of China's transition model. Concluding remarks follow in the last section.

2 From plan to market: reforms in the banking sector

The P.R.China has nationalized after 1949 when it established all means of production, including the banking system. It introduced a Planned Economy in which final decisions were centralized. From 1949 to 1978², the banking system was governed by a national monopoly, called the People's Bank of China (PBC) which operated all kinds of financial businesses with the primary function of serving government to allocate capital resources through budgetary grants. As the financial businesses were quite limited due to the planned economy itself, PBC simply provided credit needed by the SOEs for their production plans and provided cash used principally to cover labor costs and purchases of agricultural products. Obviously, the system was not modern at all and played only a limited role in promoting economic growth.

Since 1978, the policy of reform and opening up has fundamentally changed China's economy and society. As a crucial component of the economy, the banking sector has also experienced a from-plan-to-market transition process and has gone through three phases of reform.

The first phase of the banking reform, from 1979 to 1995, consisted of two parts. The main part focused on the establishment of a two-tier banking system that comprised primarily a central bank and four state-owned specialized banks (Big-four): the Agricultural Bank of China (ABC), the Bank of China (BOC), and China Construction Bank (CCB) which were established in 1979 and the Industrial and Commercial Bank of China (ICBC) established in 1984. Although the market was still a (specialized) monopolistic one because of the identical ownership (wholly state-owned) and the strict market segmentation and large market share (those four banks divided the banking market into sub-markets with clear boundaries in terms of specific banking businesses. One bank could hardly enter another bank's market and each bank dominated its own market excluding the others from taking over. For instance, the Agricultural Bank of China covered more than 95% of the rural banking business), it was the first step to break the monopoly status of the PBC. Since 1986, the PBC has gradually focused on its role as central bank by transferring its commercial banking activities to the Big-four. In 1995, with the enacts of the Central Bank Law and the Commercial Bank Law, the government completed the two-tier banking system. The second part was the gradual removal of entry barrier, which resulted in an increasing number of new entrants. Before 1987, a first attempt was made in this respect but with a very limited scope. Besides the Big-four, only a few new banks were set up: the state-owned China Investment Bank (1981), the joint-stock Bank of Communications (1986), CITIC Industrial Bank (1987) owned by China Investment and Trust Cooperation, etc. From 1987, especially from 1992, 3 policy-lending banks, 10 commercial joint-stock banks and a large number of urban credit cooperatives and rural credit cooperatives were established. As for the opening up policy in this sector, we can also see it developed step by step. Since 1979, foreign financial institutions started to set up representative agencies

²The Third Plenum of the Eleventh National Party Congress, held December 18-22, 1978, is considered a major turning point in modern Chinese political history.

in China. In 1982, the government started to license foreign banks to engage in banking businesses in China on an experimental basis, and authorized a Hong Kong based bank, the Nanyang Commercial Bank, to establish a branch in Shenzhen Special Economic Zone (SEZ). In 1985, other three SEZs were also opened to foreign financial institutions. By 1995, the number of cities opening up increased to 24.

Clearly, the process of the banking reforms in this period was rather slow, especially compared to other sectors in the economy. The reason is twofold. On the one side, the role of banks in economic development was not significant and even overlooked as China just started its marketization. On the other side, the emphases of the whole reform in this phase were rural reform (1978-1986) and the SOE reform (1987-1997). However, the domestic NPL problem of the state-owned banks (SOBs) and the financial crisis in Southeast Asia made the government realize the importance and urgency of accelerating the banking reforms in the next periods.

The second phase of banking reforms also had two major parts: further institutional building and the management of NPLs. The former focused on the commercialization of specialized banks and a separation between policy and commercial lending activities, which were started from 1995. The measures included improving the management of the SOBs by granting their managers autonomy and employees profit incentives, reducing local government intervention (e.g. the Central Bank system was re-organized by merging 31 provincial branches into 9 regional ones), the removal of credit allocation, limited interest rate deregulation, entry deregulation, a narrowing of the scope of business³ and a gradual tightening of accounting and prudential regulations. Responsibilities of financial regulation and supervision were separated into three mutually independent regulatory agencies with the PBC responsible for bank supervision, the China Securities and Regulatory Commission for equities market, and the China Insurance Regulatory Commission for the insurance industry. The latter included the recapitalization of the Big-four SOBs by injecting 270 billion RMB into them, the disposal of NPLs held by SOBs through setting up four assets management companies (as a result, 1.4 trillion RMB NPLs were carved out from the Big-four), the merger and closure of problematic banks, the transformation of urban credit cooperatives into city commercial banks, and the promotion of debt-equity swaps. Moreover, in 1999, China abolished the geographic limitations on the establishment of business institutions for foreign banks, but still with the limitation in operating specific businesses. Especially, local currency business was restricted to a few cities and banks. As of 2001, China's banking system consists of the Central Bank (PBC), the Big-four SOBs, three policy-lending banks, 110 commercial banks (99 of which are city commercial banks, one is home savings bank and the rest joint-stock commercial banks), about 3,000 urban credit cooperatives, some 42,000 rural credit cooperatives, and about 190 foreign banks with 164 branches and 233 representative offices.

As China officially joined the WTO on December 11, 2001, the third phase of banking reforms started. A detailed timeline is given as below.

- December 11, 2001 The government removed restrictions on the foreign ex-

³In effect, the government has since adopted a Glass-Steagall Act to separate commercial banking from investment banking business.

change clients of foreign banks and allowed them to conduct local currency (RMB) business in four cities: Shanghai, Shenzhen, Tianjin and Dalian;

- December 1, 2002 Foreign-funded financial institutions would be allowed to conduct RMB business in five more cities: Guangzhou, Zhuhai, Qingdao, Nanjing and Wuhan;
- 2003 Foreign banks can serve Chinese enterprises;
- 2005 Foreign banks can conduct RMB business in 20 cities;
- 2006 National treatment to foreign banks in China. Opening up in all dimensions in the banking sector.

In 2006, the Chinese banking market will be open to foreign competition. Thus, the third phase will stress the roles of the market structure and property right reform: increasing the market competition by gradually removing barriers to entry, especially for foreign banks and diversification of SOBs' ownership through introducing non-state sources of capital so as to increase their competitiveness.

From the above, although there are still many problems left unsolved, we have to admit that in the past twenty years or so, China's banking sector has evolved from a planned and undeveloped system to a market and modern system through the gradual reform.

3 The river, the stones and the strategy

3.1 A popular and perfect metaphor

China's approach of reform is indeed original and unique⁴. Although foreign advice is frequently sought, China's reforms are genuinely homegrown, based on domestic actual conditions and on learning by doing. The Chinese leader Deng Xiaoping, the general "designer" for China's reform and opening up, described the process as "crossing the river by touching the stones"⁵. This perfect metaphor vividly illustrates all the key elements of the reform strategy in China.

No doubt, generally, the starting point on the one side of the river is the planned, undeveloped and agriculture-based economy, while the opposite bank is a market-oriented, developed and modern economy. The river is the division in between the two sides. Crossing it implies the transition process from plan to market, from backwardness to modernization, from the poor to the wealthy and from isolation to opening up.

However, on which exact place to start the "wading"? How to cross the river? Shall we take the old stuff as crossing? How far is the opposite side? Where is shallow to be easy for wading or where is deep and full of dangers? Which stones are firm and stable that can be relied on and which are loose? Where is the right

⁴Of course, it has influenced some other transition economies. For example, Vietnam has adopted a similar strategy since early 1990s. The spill-over effect is greater and greater.

⁵A literal (Western) interpretation would be, "if you are on an unfamiliar course, look carefully and take one step at a time."

point for berth on the other side? ... All those questions were not clear at all during the course of the reform, especially in the beginning.

In reality, the initial “wading” started from rural sectors, which is the relatively simple and efficient program of economic liberalization as it immediately changed the situation of shortage in agricultural products by reconstructing incentive mechanisms. The most well-known “effective” “stone” was the joint-production land contract responsibility system. Now when looking back at that period, we might see that as a correct starting point as the result justified it. However, that was not the case in 1978. Farmers had no choice except this way. So, in fact, the government did not have a well-designed blueprint for reform.

The looking while going and learning by doing propelled the reform. The next step is city/enterprise reform. As for opening door policy, in 1980, the government first established a special economic zone in Shenzhen as an experimental field for the market-based activities such as international trade, foreign investment and setting up new factories with non-state ownership. Later on, other four SEZs were set up. With the successful experiences in those districts, the reform went further by opening 14 coastal cities. Nowadays, the open door policy applies in the hinterland.

Gradual process in SOE reforms is more obvious. In contrast to Russia and the Eastern European countries which adopted the Shock Therapy to privatize the state-owned sectors over one night, China has not completed the SOE reforms until today. The role of the dual-track approach⁶ is significant and efficient in this respect because of its Pareto-improving property: developing the market track without affecting the plan track. In such a way, the reform accomplished the goals of achieving economic development and keeping social stability simultaneously. Similar to the above, the reform also went forward by touching the “stones”. The first and most common one “touched” from the early 1980s was the so-called contract responsibility system, by which the management of SOEs were strengthened. As later on it ran into trouble, a second stone found in the first half of the 1990s was the separation between government and enterprises. Currently, the third one which can be treated as an obstacle as well is the ownership reform.

To further understand the ideas and features behind China’s reform strategy, I will go back to the example of the banking sector in the following to make some more detailed investigation.

3.2 The philosophy of the reform

As shown in the second section, the practice of banking sector reforms is in line with the whole economic reform. Investigating this sector provides a suitable window to watch the panorama of China’s reform and transition.

Apparently, the banking reform aims to go across the river between a planned, backward and inefficient system and a market-oriented, developed and efficient system. In order to fulfill the transition, the reform still explores the way by touching the stones, step by step. The stones, the path and the direction show us the philosophy of the reform.

⁶For details, refer to Lau, Qian and Roland (2000).

- In process: *gradualism*

The key and basic feature of China's reform is gradualism, which is an approach or policy of advancing toward the goal of constructing the market system by gradual, often slow stages. In fact, this strategy is naturally formed during the course of processing rather than pre-designed. The ultimate goal of the reform evolved over time and has only gradually come into sharper focus. In the very beginning of the reform, no body, at least no one in China or in the government that formally started the reform, realized that the final target would be constructing the market economy, simply because it was a bottom-up reform instead of a top-down one. The first goal was rather direct: increasing the output of agricultural products to get out of starvation. The second goal was still temporary: to construct a planned-based while market-supplemented economic system, which corresponded to the dual-track system period. Only after a long time of exploration and various kind of tests, did the final target came into being clearly. The other side of the river is visible now.

As for the evidences in the banking sector, in addition to the general three phases mentioned in Section 2, four lines of the reforms further demonstrate this feature.

1. Market: Big-four → new SOBs and Joint-stock banks → foreign banks;
2. Measure: management reform → ownership reform;
3. Opening: SEZs → coastal cities → hinterland;
4. Foreign banking business: representative agency → foreign currency business → local currency business.

The gradualism approach provides ample time and space for experimenting all the measures related to reforms.

- In method: *experimentalism*

Here, experimentalism means that using experimental methods to determine the validity of the measures of the reform. China always tests a new reform measure in some districts, enterprises or sectors before it is formally and nationwide carried out. Because it is prudential and mistake-allowed, this idea has advantage in minimizing the cost of reform.

Some examples in banking reform show this feature. After accumulating enough experience to supervise foreign banks' activities in the SEZs, the government would open other coastal cities, even western provinces to foreign entrants. When introducing a kind of new business, the government always first issues license to some bank(s) to take an experiment and see the effect and then determines whether or not to apply and develop it. The event of the bankruptcy of China Agricultural Trust and Investment Company made the government realize the necessity of separating commercial banking from investment banking business. As there is no well-designed blueprint and no

former footprints can be referred to, taking experimental methods is then a natural choice when confronting uncertainty.

- In attitude: *pragmatism*

Pragmatism is a practical, matter-of-fact way of approaching or assessing situations or of solving problems. Gradualism and experimentalism have characterized the practical attitude all through the reform.

Given the initial economic and political conditions, China had no basis to carry out single-track full marketization in the banking sector. As a compromised and pragmatic choice, the dual-track system helped to foster the transition from plan to market as it minimizes political opposition to reform *ex ante* and maximizes political opposition to reversal of reform *ex post*. In such a way, the market gradually came into being without much instability or obvious obstruction. Another example lies in the gradual removal of entry barriers and business restrictions for foreign banks. This process provides enough time for the banks to make a sufficient domestic competition as the rehearsal for the future international competition through learning by doing. As the spill-over effect of know-how from foreign banks strengthened the competitiveness of domestic banks, we can anticipate that in 2006, when the door is fully open, the SOBs will grow up to a highly competitive standard. So, the strategy is successful both in developing the economy and in keeping social stability.

This feature can be further understood from another well-known saying by Deng Xiaoping: "No matter whether it is black or white, a cat that catches mice is a good cat." It reflected that the government would explore a way by practice rather than by dogmatism.

- In nature: *evolutionism*

Evolutionism in reform implies a belief that reform is a gradual process in which the economy changes into a different and usually better form through a developing and continuously revising way rather than a well pre-designed or one-shot approach. Exactly speaking, this feature is not an approach but a belief and conclusion.

Importantly and fortunately, evolutionism is not alone. The path-dependent property and self-enforcing mechanism keep the reform go forward to the market instead of backward to the planned economy. That is, once stepping on the road, then no way back.

For instance, nobody was against the establishments of the Big-four and other new SOBs. A limited scope of experiment of joint-stock commercial banks and foreign banks did not incur much objection either, as the existing interest groups can be relatively protected. However, gradually, the competition from the new track has forced the old system to make change and approach to the market system. The entry to the WTO is a milestone. Even if someone wants to go back, that's not possible, which guarantees the direction of the reform. The external pressure and incentive provide new engines for the reform.

A natural question concerning those distinguishing features of China's reform and transition is why it initially and mainly happened in China instead of the other transition economies? Economists usually look for explanations from the initial economic and political constraints and actual conditions at the beginning of the reform in China. Yes, indeed, that's the direct reason. However, the cultural dimension and even the historical/traditional or philosophical dimension as the fundamental factors were neglected. An excuse might be that it's hard to introduce them into scientific models. However, cutting edge findings usually come from breakthroughs. Scientific research need the spirit of "crossing the river by touching the stones" as well. In short, if taking the doctrine of the mean of the Confucian school and China's more than two thousand years' feudal history and stable social structure into consideration, we will find a more reliable answer. Put it another way, why is it easier to practise the "Big-Bang" strategy in European countries? Fundamentally, it's easier to be understood and accepted by the minds of the people living there.

4 Concluding remarks

In addition to the significant effects in the transition and development of China's economy and society, the gradualist reform has far reaching influences in two aspects at least. First, the success sets an example for other countries, especially the transition economies and developing countries. Vietnam took the same strategy. Some CEE countries have already adopted similar approaches and measures in specific sectors to slow down the radical paces. Even in North Korea, the government started to set up the SEZs to make some experiments. Moreover, Japanese also realized the advantages of the strategy and started to learn from it for reforming its stagnant and depressed economy. Another effect is in academic research. Not only in economics, but also in the related social sciences, China's reform provides a platform and a new path to explore the unknown world and to further update our knowledge.

Moreover, this paper does not make any predictions regarding the eventual success or failure of China's transition. By the way, what's the exact meaning of "eventual" in respect of societal evolution? The most important thing in Economics is not prediction but lies in explanation.

References

- Chow, Gregory C. (1997), "Challenges of China's Economic System for Economic Theory", *The American Economic Review*, 87, 321–327.
- Dewatripont, Mathias, and Gérard Roland (1992), "The Virtues of Gradualism and Legitimacy in the Transition to a Market Economy", *The Economic Journal*, 102, 291–300.
- Dewatripont, Mathias, and Gérard Roland (1995), "The Design of Reform Packages under Uncertainty," *The American Economic Review*, 85, 1207–1223.

- Ju, Yuan, and Juan Pang (2000), “Chinese Model of Deregulation”, *Journal of Southwest University for Nationalities*, Vol. 21, No. 11, Nov. 2000 (in Chinese).
- Qian, Yingyi, Lawrence Lau and Gérard Roland (2000), “Reform without Losers: An Interpretation of China’s Dual-Track Approach to Transition”, *Journal of Political Economy*, February 2000, 108(1), 120–143.
- Roland, Gérard (2000), *Transition and Economics: Politics, Markets, and Firms*, Cambridge: MIT Press.
- Ruys, Pieter H.M., Réne van den Brink and Radislav Semenov (2000), “Values and Governance Systems”, in C. Menard (ed.), *Institutions, Contracts and Organizations: Perspectives from New Institutional Economics*, Cheltenham, Edward Elgar, 422–446.
- Shirai, Sayuri (2002), “Banking Sector Reforms in the Case of the People’s Republic of China-Progress and Constraint”, Chapter III in *Rejuvenating Bank Finance for Development in Asia and the Pacific*, United Nations 2002.
- Yu, Liangchun, and Yuan Ju (1999), “Monopoly and Competition: Banking Reform and Development in China”, *Journal of Economic Research*, No. 8, Aug. 1999 (in Chinese).
- Yu, Liangchun, and Yuan Ju (2000), “Banking Development and Competition Policy in China During the 10th Five Year Plan Period”, *Reform*, Vol. 103, May 2000 (in Chinese).

PUBLIEK-PRIVATE SAMENWERKING

Gerard van der Laan* en Dolf Talman†

Vrije Universiteit en Universiteit van Tilburg

De bijdragen in deze bundel geven een goede karakterisering van de wiskundig econoom Pieter Ruys. Aan de ene kant zien we bepaalde aspecten steeds weer opnieuw terugkomen, aan de andere kant valt het op dat de diverse auteurs ook heel verschillende onderwerpen uit Pieter's wetenschappelijke leven als kenmerkend voor zijn werk noemen. Zo begint de bijdrage van Michael Kosfeld met 'Networks play an important role in the research agenda of Pieter Ruys', kenschetst Ton Storcken 'Tripolariteit en Dualiteit' als twee van Pieter's meest bekende stokpaardjes, en noemt Cees Withagen de thema's 'publieke goederen, algemeen evenwicht en diensteneconomie' als kenmerkend voor Pieter's werk. Hoewel enige samenhang tussen deze onderwerpen niet kan worden ontkend, geeft dit toch een beeld van de breedte van Pieter's interesses. Daarnaast komt in de bijdragen ook een ander belangrijk kenmerk van Pieter's werk als wiskundig econoom naar voren, namelijk zijn voorliefde voor abstracte modellen van de werkelijkheid. Rob Gilles zegt: 'Pieter is één van de meest abstracte denkers die ik onder economen ontmoet heb' en Cees Withagen merkt op dat het verzoek van de redactie om de bijdragen 'een persoonlijk karakter te geven en daarom relatief weinig formeel moeten zijn', moeilijk te begrijpen is. Dit betekent echter niet dat de wiskunde in Pieter's werk centraal staat. Integendeel, Pieter gebruikt de wiskunde als formele taal, zie daarover bijv. zijn oratie 'Rationeel gedrag' (1981), om economische vraagstukken te modelleren, maar juist deze vraagstukken staan voorop. Terecht merkt Rob Gilles op: 'Wiskunde is voor Pieter niet een doel op zich, (...) maar vooral een middel om ideeën precies uit te drukken en volgens strikte logische deductie conclusies te formuleren'. Het werk van Pieter overziende moeten we dan ook concluderen dat dit vooral wordt gekenschetst door vernieuwende ideeën over belangrijke economische vraagstukken, waarbij de wiskunde weliswaar een hulpmiddel is om de ideeën in een model te verwoorden, maar waarbij vooral zijn originele denkwijze en zijn brede visie vanuit verschillende invalshoeken de aandacht trekken. Zijn innovatieve bijdragen bestrijken een breed terrein, met name het tripolaire model en zijn modellen over semi-publieke goederen, de hiërarchische organisatie van bedrijven en de diensteneconomie moeten worden genoemd als parels van zijn originele denkwijze. We doen Pieter dan ook tekort als we hem kenschetsen als 'wiskundig econoom'. Integendeel, Pieter is voluit

*Adres: Departement Econometrie en Tinbergen Instituut, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam. *Email:* glaan@feweb.vu.nl

†Adres: Departement Econometrie & Operations Research en CentER, Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg. *Email:* talman@uvt.nl

de econoom, die zich niet laat beperken door de wiskunde, maar voor wie het wiskundig model slechts een voertuig is om de ‘weidse vergezichten’ ook daadwerkelijk te bereiken. Hierbij nemen twee modellen een centrale plaats in, namelijk het tripolaire model en het algemeen evenwichtsmodel. Dat deze modellen niet los van elkaar kunnen worden gezien, wordt duidelijk gemaakt in hoofdstuk 7 van zijn proefschrift (1974), waar het ‘tripolaire diagram’ een evenwicht weergeeft als ‘a realization of a relation between values and resources’. In deze bijdrage willen we met name het algemeen evenwichtdenken van Pieter centraal stellen. Daartoe passen we het algemeen evenwichtskoncept toe op een model van een economie met publiek-private diensten, een door Pieter geïntroduceerd concept, dat nauw verwant is met zijn werk op het gebied van de publieke goederen en de diensteneconomie. In de afsluitende paragraaf willen we kort ingaan op de acceptatie van het concept van algemeen evenwicht binnen Nederland en de rol van Pieter daarin.

1 Model van een economie met publiek-private diensten

Het algemeen evenwicht staat centraal in de moderne economische theorie en wordt hier gebruikt als referentiekader voor de implementatie van een systeem voor de financiering van infrastructuur. In de economische literatuur wordt infrastructuur gekarakteriseerd als een omvangrijk, ondeelbaar, niet-rivaliserend kapitaalgoed dat diensten produceert voor haar gebruikers. Consumenten ontlenen in het algemeen geen direct nut aan de aanwezigheid van infrastructuur maar wel indirect nut, omdat de beschikbaarheid van infrastructuur in combinatie met het gebruik van een privaat goed diensten genereert. Voorbeeld is het openbaar wegennet. Zo geeft de beschikbaarheid van het wegennet de consument in combinatie met het gebruik van de auto de mogelijkheid van vervoer. Dikwijls zijn met de aanleg en het onderhoud van infrastructuur hoge kosten gemoeid en kan iedereen er vrijelijk gebruik van maken. Vanwege de niet-rivaliteit en de niet-uitsluitbaarheid van gebruik wordt infrastructuur gewoonlijk als een publiek goed gezien. Aan de andere kant genereert infrastructuur samen met het gebruik van een privaat goed diensten en heeft het daarmee ook kenmerken van een privaat goed. Infrastructuur is dan ook een schoolvoorbeeld van een kapitaalgoed dat een ‘publiek-private’ dienst genereert.

Van oudsher is de publieke sector verantwoordelijk voor aanleg, onderhoud, beheer en financiering van infrastructuur, zoals bijv. het openbaar wegennet. Echter, ontwikkelingen in de moderne informatie-technologie openen nieuwe mogelijkheden om de behoefte van gebruikers aan een specifiek soort publieke infrastructuur te achterhalen, door het gebruik ervan per individu te monitoren. Als gevolg daarvan kan de financiering van uitbreiding van bestaande publieke infrastructuur geprivatiseerd worden door de gebruikers te laten meebetalen aan de beschikbaarheid ervan. Het basisidee hier is dat het gebruik van infrastructuur beperkt wordt door de omvang van de infrastructuur, waardoor gebruikers die zich beperkt voelen door een te kleine voorziening aan infrastructuur bereid zijn te betalen voor uitbreiding ervan. Dat vormt de basis voor publiek-private samenwerking om meer infrastructuur aan te leggen en te onderhouden.

Wij beschouwen een economisch model met één type infrastructuur die een publiek-private dienst (pp-dienst) genereert, d.w.z. de infrastructuur is een publiek goed dat complementair is aan het gebruik van een specifiek privaat goed, bijvoorbeeld het openbaar wegennet, dat complementair is met het gebruik van auto's, waarbij het gebruik van auto's bijvoorbeeld gemeten kan worden aan de hand van het aantal verbruikte liters benzine. De combinatie van de publieke beschikbaarheid van het wegennet en de consumptie van het complementaire private goed benzine levert de pp-dienst 'vervoer'. We zullen in het vervolg meestal dit voorbeeld gebruiken. Met de omvang van het wegennet (infrastructuur) bedoelen we alle relevante factoren die het individuele gebruik beïnvloeden, dus niet alleen het aantal km's asfalt, maar bijvoorbeeld ook de kwaliteit van het wegennet en het gebruik door anderen. Een individuele beslissing om gebruik te maken van de auto hangt niet alleen af van de beschikbaarheid van een weg tussen A en B, maar ook van de kwaliteit van die weg (snelweg, rijksweg, provinciale weg, B-weg), alswel de verwachte drukte (files). Al deze factoren bepalen de 'omvang' van de infrastructuur. Het model is dus niet een specifiek congestiemodel, maar bevat congestie wel als één van de mogelijke factoren die de omvang bepalen. We gaan er van uit dat de omvang van de infrastructuur op een 1-dimensionale schaal wordt gemeten.

Ter vereenvoudiging en zonder verlies van algemeenheid nemen we aan dat het gebruik van één eenheid pp-dienst overeenkomt met de consumptie van één eenheid van het complementaire private goed. De door een gebruiker genoten pp-dienst (vervoer) wordt dus gegeven door een combinatie van de omvang van de publieke infrastructuur (het wegennet) en het gebruik ervan, waarbij het laatste wordt gemeten door de consumptie van het complementaire private goed (benzine). Naast de infrastructuur en het complementaire private goed, zijn er in de economie n zuiver private goederen, die niet gerelateerd zijn aan de infrastructuur. Verder zijn er m consumenten, geïndiceerd $h = 1, \dots, m$, en twee producenten, $j = 1, 2$.

Consument h beschikt over een initiële voorraad $\omega^h \in \mathbb{R}_+^n$ van zuiver private goederen, $h = 1, \dots, m$. De benutting van de pp-dienst door consument h wordt gegeven door het paar (s^h, z) , met $z \geq 0$ de omvang van de infrastructuur en $s^h \geq 0$ de consumptie van het complementaire private goed. De maximale consumptie van het complementaire private goed door consument h wordt bepaald door de omvang van de infrastructuur en wordt gegeven door een niet-negatieve, concave, stijgende functie $q^h: \mathbb{R}_+ \rightarrow \mathbb{R}_+$, d.w.z. voor elk toegelaten paar (s^h, z) van de pp-dienst voor consument h geldt de restrictie

$$s^h \leq q^h(z), \quad h = 1, \dots, m. \quad (1)$$

De functie q^h is een individuele subjectieve beschikbaarheidsfunctie, d.w.z. consument h voelt zich vanwege de omvang van het wegennet beperkt in haar gebruik van het wegennet en daarmee in de consumptie van benzine. Hoewel een hoger gebruik in theorie mogelijk is, zal consument h bij omvang z van de infrastructuur de consumptie van benzine beperken tot maximaal $q^h(z)$, bijvoorbeeld omdat vanwege de geboden kwaliteit of congestie een hoger gebruik van het wegennet te veel tijd kost, te veel irritatie veroorzaakt, een (goede) weg tussen A en B ontbreekt, of eventueel andere al dan niet gevoelsmatige factoren meespelen.

De voorkeuren van consument h worden gegeven door de nutsfunctie $u^h(x^h, s^h)$,

waarbij $x^h \in \mathbb{R}_+^n$ de consumptie van de n zuiver private goederen is en s^h de consumptie van het complementaire private goed en dus tevens het gebruik van de infrastructuur weergeeft. We veronderstellen hierbij dat de aanwezigheid van het publiek goed zelf geen direct nut oplevert voor de consument. Uiteraard heeft de omvang van de infrastructuur via restrictie (1) wel invloed op het indirecte nut van de individuele consument. Voor een meer algemeen model waarin het nut ook direct van de infrastructuur afhangt, dus de nutsfunctie van de vorm $u^h(x^h, s^h, z)$ is, wordt verwezen naar van der Laan, Ruys en Talman (2001).

Wat de producenten betreft, produceert producent 1 het complementaire private goed volgens de transformatiefunctie $T^1: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$. De bijbehorende verzameling van technologisch toegelaten input-output combinaties (y^1, v) wordt gegeven door

$$T^1(y^1, v) \leq 0, \quad (2)$$

met $y^1 \in \mathbb{R}_-^n$ een vector van inputs van de zuiver private goederen en $v \geq 0$ de output van het complementaire private goed. In van der Laan, Ruys en Talman (2001) wordt toegelaten dat deze producent ook de pp-dienst als input gebruikt, hetgeen we hier achterwege laten om het model niet nodeloos gecompliceerd te maken. Producent 2 produceert ‘infrastructuur’ volgens de transformatie functie $T^2: \mathbb{R}^{n+1} \rightarrow \mathbb{R}$. De bijbehorende verzameling van technologisch toegelaten input-output combinaties (y^2, z) wordt gegeven door

$$T^2(y^2, z) \leq 0, \quad (3)$$

met $y^2 \in \mathbb{R}_-^n$ een vector van inputs van de zuiver private goederen en $z \geq 0$ de output van infrastructuur. De output z moet hier worden gezien als toevoeging (of onderhoud) aan de bestaande infrastructuur, d.w.z. de maat voor de omvang van infrastructuur wordt zo genormaliseerd dat de omvang van de initiële hoeveelheid nul is.

In het vervolg nemen we aan dat de bovengeformuleerde economie $\mathcal{E} = \{(T^j)_{j=1,2}, (u^h, q^h, \omega^h)_{h=1}^m\}$ regulier is, d.w.z. alle nutsfuncties u^h en beschikbaarheidsfuncties q^h , $h = 1, \dots, m$, en transformatiefuncties T^1 en T^2 zijn continu-differentieerbaar, de nutsfuncties zijn stijgend en strikt quasi-concaaf, de beschikbaarheidsfuncties zijn niet-negatief, concaaf en stijgend, de transformatie functies zijn strikt concaaf en voldoen aan $T^j(\underline{0}) = 0$, $j = 1, 2$, en alle vectoren ω^h , $h = 1, \dots, m$, zijn strikt positief. Al deze aannames zijn standaard in de algemeen evenwichtsliteratuur, uiteraard met uitzondering van de aannames op de beschikbaarheidsfuncties, waarvan we later zullen zien dat met name de concaviteit een belangrijke rol speelt.

Een allocatie e is een vector van consumptieplannen (x^h, s^h) , $h = 1, \dots, m$, en productieplannen (y^1, v) en (y^2, z) . Een allocatie $e = ((x^h, s^h)_{h=1}^m, (y^1, v), (y^2, z))$ is toegelaten als wordt voldaan aan de restricties

$$\left. \begin{array}{l} \text{(i)} \quad T^1(y^1, v) \leq 0 \text{ en } T^2(y^2, z) \leq 0 \text{ (technologierestricties),} \\ \text{(ii)} \quad \sum_{h=1}^m (x^h - \omega^h) - y^1 - y^2 \leq \underline{0} \text{ (geen overbenutting private goederen),} \\ \text{(iii)} \quad \sum_{h=1}^m s^h \leq v \text{ (voldoende productie complementair privaat goed),} \\ \text{(iv)} \quad s^h \leq q^h(z), \quad h = 1, \dots, m, \text{ (beschikbaarheidsrestricties).} \end{array} \right\} (4)$$

Zij A de verzameling van toegelaten allocaties en zij $u^h(e) \equiv u^h(x^h, s^h)$ het nut van consument h , $h = 1, \dots, m$, in allocatie $e = ((x^h, s^h)_{h=1}^m, (y^1, v), (y^2, z)) \in A$.

Dan is $e \in A$ een Pareto-efficiënte allocatie als er geen allocatie $e' \in A$ is, zodanig dat $u^h(e') > u^h(e)$ voor elke $h = 1, \dots, m$. Zij nu $W: \mathbb{R}^m \rightarrow \mathbb{R}$ een sociale welvaartsfunctie, die voor gegeven coëfficiënten $\alpha_h > 0$, $h = 1, \dots, m$, welvaartsniveau $W(u_1, \dots, u_m) = \sum_{h=1}^m \alpha_h u_h$ toekent bij nutsniveaus u_h , $h = 1, \dots, m$. De eerste orde voorwaarden volgen dan uit het welvaartsmaximalisatieprobleem

$$\max_{e \in A} \sum_{h=1}^m \alpha_h u^h(e). \quad (5)$$

In de oplossing van dit probleem wordt voldaan aan de restricties (4) met, vanwege de regulariteitsaanname, gelijkheid voor de condities (i), (ii) en (iii). Voor elk tweetal zuiver private goederen kunnen de Kuhn-Tucker voorwaarden voor de eerste afgeleiden van de bijbehorende Lagrange functie worden gereduceerd tot de standaardvoorwaarde dat voor elke consument en elke producent de marginale substitutieverhouding gelijk is aan de marginale transformatieverhouding: voor elk paar $k, \ell \in \{1, \dots, n\}$ geldt

$$\frac{\partial u^h / \partial x_k^h}{\partial u^h / \partial x_\ell^h} = \frac{\partial T^j / \partial y_k^j}{\partial T^j / \partial y_\ell^j}, \quad h = 1, \dots, m, \quad j = 1, 2. \quad (6)$$

De eerste orde voorwaarden voor het publieke goed infrastructuur en het complementaire private goed zijn aan elkaar gekoppeld via de pp-dienst. Voor deze goederen volgt uit de Kuhn-Tucker voorwaarden voor de eerste afgeleiden van de Lagrange functie dat in een efficiënte allocatie voor een willekeurig gekozen zuiver privaat goed ℓ moet worden voldaan aan de complementariteitsrestricties

$$\frac{\partial u^h / \partial s^h}{\partial u^h / \partial x_\ell^h} - \frac{\partial T^1 / \partial v}{\partial T^1 / \partial y_\ell^1} \geq 0 \quad \perp \quad s^h - q^h(z) \leq 0, \quad h = 1, \dots, m, \quad (7)$$

en

$$\sum_{h=1}^m \left(\frac{\partial u^h / \partial s^h}{\partial u^h / \partial x_\ell^h} - \frac{\partial T^1 / \partial v}{\partial T^1 / \partial y_\ell^1} \right) \frac{\partial q^h}{\partial z} - \frac{\partial T^2 / \partial z}{\partial T^2 / \partial y_\ell^2} \leq 0 \quad \perp \quad z \geq 0, \quad (8)$$

waarbij $a \geq 0 \perp b \leq 0$ staat voor $a \geq 0$, $b \leq 0$ en $a^\top b = 0$. Conditie (7) zegt dat in een efficiënte allocatie de marginale substitutieverhouding van het complementaire private goed van consument h gelijk is aan de marginale transformatieverhouding ervan als de beschikbaarheidsrestrictie $s^h \leq q^h(z)$ niet bindend is. Als deze restrictie wel bindend is, is de marginale substitutieverhouding minstens gelijk aan de marginale transformatieverhouding, hetgeen wil zeggen dat de consument voor het gebruik van meer infrastructuur bereid is een opslag te betalen boven op de marginale kosten van het complementaire private goed. Conditie (8) zegt dat de som van alle opslagen maal de marginale toename van de beschikbaarheid gelijk is aan de marginale transformatieverhouding van de infrastructuur als er alleen (meer) infrastructuur geproduceerd wordt, met andere woorden het is alleen efficiënt om het wegnemen uit te breiden als consumenten bereid zijn voor het gebruik van het wegnemen een opslag te betalen boven op de benzinekosten. Omdat conditie (7) laat

zien dat een consument hiertoe alleen bereid is als de beschikbaarheidsrestrictie bindend is, volgt hieruit dat $z > 0$ impliceert dat $s^h = q^h(z)$ voor minstens één consument h . Immers, als $s^h < q^h(z)$ voor alle h , dan zijn alle opslagen nul, zodat aan de linkerkant van (8) wordt voldaan met strikte ongelijkheid en dus moet gelden dat $z = 0$.

De efficiency condities laten zien dat in dit model van een pp-dienst de kosten van de infrastructuur moeten worden opgebracht uit de opslagen op de prijs van het complementaire private goed. Deze observatie maakt duidelijk dat dit model met een pp-dienst belangrijk afwijkt van de standaard publieke goederen literatuur. Waar in deze literatuur het cruciale probleem is dat er geen informatie bekend is over de waardering van de gebruikers voor het publieke goed, en het daarmee moeilijk, zo niet onmogelijk is om rechtvaardige en efficiënte individuele bijdragen in de bekostiging te bepalen, laat dit model met een pp-dienst zien dat de individuele vraag naar infrastructuur kan worden waargenomen door middel van de consumptie van het complementaire goed. Deze informatie kan in principe worden gebruikt om de individuele bijdragen in de bekostiging te bepalen. In de volgende paragraaf wordt hiervoor een institutionele structuur voorgesteld binnen het kader van het algemeen evenwichtsconcept.

2 Een publiek-privaat samenwerkingsevenwicht

Om de in de vorige paragraaf gemodelleerde economie met een publiek-private dienst binnen het competitief kader van een algemeen evenwichtsmodel te brengen, wordt naast de consumenten en de twee producenten een derde institutionele economische agent geïntroduceerd, namelijk een publiek-privaat samenwerkingsverband (pps). Zoals gebruikelijk in een competitieve omgeving wordt verondersteld dat de twee producenten winstmaximaliserend zijn en de consumenten nutsmaximaliserend. Het pps is een principaal-agent verband, dat de publieke-private dienst organiseert en exploiteert. De principaal is de *publieke autoriteit voor infrastructuur*, de agent is de *private infrastructuur onderneming*. De publieke autoriteit bepaalt de optimale omvang van de infrastructuur, dat is die omvang waarbij wordt voldaan aan conditie (8). De private onderneming is winstmaximaliserend en sluit een contract met de autoriteit om de vastgestelde omvang te realiseren en te financieren. We zullen zien dat in evenwicht dit contract vanwege de concaviteit van de beschikbaarheidsfuncties nooit verliesgevend is.

Zoals gebruikelijk in een economie met privaat eigendom gaan alle winsten naar de consumentensector. In het vervolg is $\phi_j^h \geq 0$ het aandeel van consument h in de winst van onderneming j , $j = 0, 1, 2$, waarbij $j = 0$ de private infrastructuur onderneming van de pps is en $j = 1, 2$ de twee andere producenten zijn. We nemen aan dat alle winst wordt uitgekeerd en dus dat $\sum_{h=1}^m \phi_j^h = 1$, $j = 0, 1, 2$. Zij verder $\hat{p} = (p, p_v, p_z) \in \mathbb{R}_+^n \times \mathbb{R}_+ \times \mathbb{R}_+$ een vector van prijzen met $p = (p_1, \dots, p_n)^\top \in \mathbb{R}_+^n$ de vector van prijzen voor de zuiver private goederen $\ell = 1, \dots, n$, p_v de prijs van het complementaire private goed en p_z de prijs per eenheid infrastructuur. Daarnaast is z de omvang van de infrastructuur en $t = (t^1, \dots, t^m)^\top \in \mathbb{R}_+^m$ een vector van individuele ‘opslagen’ op de prijs van het complementaire private goed

met t^h de opslag voor consument h , $h = 1, \dots, m$. Tenslotte geven we met $I = (I^1, \dots, I^m)^\top \in \mathbb{R}_+^m$ de vector van inkomens aan.

We formuleren nu eerst het nutsmaximalisatie probleem van een consument h bij gegeven prijzen \hat{p} , opslag t^h , inkomen I^h en omvang van de infrastructuur z . De bestedingen van consument h zijn de uitgaven aan consumptie van de zuiver private goederen plus de kosten voor het gebruik van infrastructuur. Per eenheid van het complementaire private goed bestaan deze kosten uit de prijs p_v van het complementaire private goed, te betalen aan producent 1, plus de opslag t^h ‘gebruik infrastructuur’ te betalen aan de private infrastructuur onderneming 0. Rekening houdend met de beschikbaarheidsrestrictie (1) wordt het nutsmaximalisatieprobleem van consument h gegeven door

$$\max_{x^h \geq 0, s^h \geq 0} u^h(x^h, s^h) \quad \text{onder de restricties} \quad \begin{cases} p^\top x^h + (p_v + t^h)s^h \leq I^h, \\ s^h \leq q^h(z). \end{cases} \quad (9)$$

De oplossing $x^h \equiv x^h(\hat{p}, t^h, I^h, z) \in \mathbb{R}_+^n$ en $s^h \equiv s^h(\hat{p}, t^h, I^h, z) \geq 0$ geeft de vraag van de consument naar de zuiver private goederen respectievelijk het complementaire private goed als functie van de prijzen, de persoonlijke opslag, het eigen inkomen en de omvang van de infrastructuur.

Gegeven de prijsvector $\hat{p} = (p, p_v, p_z)$ wordt voor de producenten $j = 1, 2$ het winstmaximalisatie probleem gegeven door

$$\max_{y^j \geq 0, r^j \geq 0} p^\top y^j + p_{r^j} r^j \quad \text{onder de restrictie} \quad T^j(y^j, r^j) \leq 0, \quad (10)$$

waarbij $r^1 = v$ en $r^2 = z$. De oplossing $y^j \equiv y^j(\hat{p}) \in \mathbb{R}_+^n$ en $r^j \equiv r^j(\hat{p}) \geq 0$ geeft de vraag van producent j , $j = 1, 2$, naar inputs en zijn aanbod van output, respectievelijk het aanbod van het complementaire private goed van producent 1 en het aanbod van ‘infrastructuur’ van producent 2, als functie van alle prijzen. De bijbehorende winsten als functie van de prijzen worden gegeven door

$$\pi^j(\hat{p}) = p^\top y^j(\hat{p}) + p_{r^j} r^j(\hat{p}), \quad j = 1, 2. \quad (11)$$

Binnen de principaal-agent relatie van de pps speelt zowel de principaal als de agent een eigen rol. Hierbij bepaalt de publieke autoriteit (principaal) de omvang van de infrastructuur en de private onderneming (agent) de opslagen. Allereerst bekijken we de autoriteit. Deze stelt de omvang van de infrastructuur vast. Gegeven de prijs p_z , de opslagen t^h en de van z afhankelijke vraag $s^h = s^h(\hat{p}, t^h, I^h, z)$ van iedere consument h naar het complementaire private goed, wordt de optimale omvang bepaald door de complementariteitsconditie

$$\sum_{\{h | s^h = q^h(z)\}} t^h \frac{\partial q^h}{\partial z} \leq p_z \perp z \geq 0. \quad (12)$$

Zij $\zeta \equiv \zeta(\hat{p}, t, I)$ de oplossing van (12) als functie van de prijzen, de opslagen en de inkomens. Dan geldt $\zeta(\hat{p}, t, I) = 0$ als $\sum_{\{h | s^h = q^h(z)\}} t^h \frac{\partial q^h}{\partial z} \leq p_z$ en $z = 0$. Als $\sum_{\{h | s^h = q^h(z)\}} t^h \frac{\partial q^h}{\partial z} > p_z$ ien $z = 0$, dan heeft vergelijking (12) een unieke positieve

oplossing $\zeta(\widehat{p}, t, I)$, dit vanwege de concaviteit van de beschikbaarheidsfuncties q^h , waardoor de afgeleiden niet-stijgend zijn, en het feit dat het aantal consumenten h met $s^h(\widehat{p}, t^h, I^h, z) = q^h(z)$ dalend is in z .

De private infrastructuur onderneming (agent) in de principaal-agent relatie contracteert zich om de omvang van de infrastructuur te realiseren en te financieren. Hierbij krijgt de agent contractueel de vrijheid om de individuele opslagen vast te stellen onder de voorwaarde dat het tarief nul is voor elke consument waarvan de beschikbaarheidsrestrictie voor het gebruik van de infrastructuur niet bindend is. De kosten van de agent worden bepaald door de omvang z , die tegen prijs p_z per eenheid moet worden aangekocht van de infrastructuurproducent 2. De opbrengsten verkrijgt de agent uit de opslagen voor de consumenten. Gegeven een omvang z aan infrastructuur en de van de opslag afhankelijke vraag $s^h = s^h(\widehat{p}, t^h, I^h, z)$ van iedere consument h wordt, onder de tarieven vaststellende bevoegdheid van de agent, zijn winstmaximalisatieprobleem gegeven door

$$\max_{t^1, \dots, t^m \geq 0} \sum_{h=1}^m t^h s^h(\widehat{p}, t^h, I^h, z) - p_z z \quad (13)$$

onder de restrictie dat $t^h = 0$ als $s^h(\widehat{p}, 0, I^h, z) < q^h(z)$. Zij $\tau^h \equiv \tau^h(\widehat{p}, I^h, z)$, $h = 1, \dots, m$, de oplossing van (13) als functie van de prijzen \widehat{p} , het inkomen I^h en de omvang z aan infrastructuur. Dan geldt dat deze winstmaximaliserende opslagen τ^h voldoen aan

$$\tau^h = \max \{t^h \mid s^h(\widehat{p}, t^h, I^h, z) \leq q^h(z) \perp t^h \geq 0\}, \quad h = 1, \dots, m. \quad (14)$$

Als $s^h(\widehat{p}, t^h, I^h, z) < q^h(z)$ voor $t^h = 0$, dan volgt dat $\tau^h(\widehat{p}, I^h, z) = 0$, anders is $\tau^h(\widehat{p}, I^h, z)$ de maximale opslag waarbij de beschikbaarheidsrestrictie $q^h(z)$ nog net bindend is voor consument h . De bijbehorende winst van de agent is gelijk aan

$$\pi^0(\widehat{p}, I, z) = \sum_{h=1}^m \tau^h(\widehat{p}, I^h, z) s^h(\widehat{p}, \tau^h(\widehat{p}, I^h, z), I^h, z) - p_z z. \quad (15)$$

Bovenstaand hebben we de vraagfuncties x^h en s^h van de consumenten, de aanbod-, vraag- en winstfuncties y^j , r^j en π^j van de producenten $j = 1, 2$, de optimale omvangfunctie ζ van de principaal en de opslagfuncties τ^h en de winstfunctie π^0 van de agent afgeleid als functies van de prijzen $\widehat{p} = (p, p_v, p_z)$, de opslagen $t = (t^1, \dots, t^m)$, de inkomens $I = (I^1, \dots, I^m)$ en de omvang z aan infrastructuur. In evenwicht moet uiteraard gelden dat alle consumenten hun nut maximaliseren, alle producenten hun winst maximaliseren, aan alle inkomens- en marktcondities wordt voldaan en tevens dat de ‘ex-post’ door de principaal vastgestelde omvang ζ gelijk is aan de ‘ex-ante’ hoeveelheid z en de ‘ex-post’ door de agent vastgestelde opslagen τ^h gelijk zijn aan de ‘ex-ante’ opslagen t^h , $h = 1, \dots, m$. Voor ons model van een economie met één publiek-private dienst leidt dit tot de onderstaande definitie van een algemeen evenwicht: het publiek-privaat samenwerkingsevenwicht (ppse).

Definitie 1

Een publiek-privaat samenwerkingsevenwicht (ppse) is een allocatie $e =$

$((x^h, s^h)_{h=1}^m, (y^1, v), (y^2, z))$, een niet-negatieve prijsvector \widehat{p} , een niet-negatieve vector van inkomens I en een niet-negatieve vector van opslagen t , zodanig dat

- (i) voor alle h : $x^h = x^h(\widehat{p}, t^h, I^h, z)$, $s^h = s^h(\widehat{p}, t^h, I^h, z)$,
- (ii) $y^j = y^j(\widehat{p})$, $j = 1, 2$, $v = v(\widehat{p})$, $z = z(\widehat{p})$,
- (iii) voor alle h : $I^h = p^\top \omega^h + \phi_0^h \pi^0(\widehat{p}, I, z) + \sum_{j=1,2} \phi_j^h \pi^j(\widehat{p})$,
- (iv) $\sum_{h=1}^m x^h - y^1 - y^2 = \sum_{h=1}^m \omega^h$,
- (v) $\sum_{h=1}^m s^h = v$,
- (vi) $z = \zeta(\widehat{p}, t, I)$,
- (vii) voor alle h : $t^h = \tau^h(\widehat{p}, I^h, z)$.

De condities (i) en (ii) zeggen dat in de evenwichtsallocatie elke consument zijn nut maximaliseert onder de in (9) gegeven restricties en dat elke producent zijn winst maximaliseert, conditie (iii) zegt dat in een pps het inkomen van consument h gelijk is aan de waarde van zijn initieel bezit plus zijn aandelen in de winst van de drie ondernemingen, condities (iv) en (v) zijn de markevenwichtscondities voor respectievelijk de zuiver private goederen en het complementaire private goed, conditie (vi) zegt dat de evenwichtsomvang z gelijk is aan de door de principaal vastgestelde omvang van de infrastructuur en conditie (vii) zegt dat evenwichtsopslagen gelijk zijn aan de door de agent vastgestelde opslagen.

Merk op dat in evenwicht alle winsten niet negatief zijn. Voor de producenten $j = 1, 2$ volgt dit onmiddellijk uit de regulariteitsconditie $T^j(\underline{Q}) = 0$. Voor de infrastructuur onderneming geldt $\pi^0(\widehat{p}, I, z) \geq 0$ als $z = 0$. Als $z > 0$, dan geldt in evenwicht dat $t^h = 0$ als $s^h < q^h(z)$ en wordt aan de linkerterm van (12) voldaan met gelijkheid, zodat

$$\begin{aligned} \pi^0(\widehat{p}, I, z) &= \sum_{\{h|s^h=q^h(z)\}} t^h q^h(z) - p_z z \\ &= \sum_{\{h|s^h=q^h(z)\}} t^h (q^h(z) - \frac{\partial q^h}{\partial z} z) \\ &= \sum_{\{h|s^h=q^h(z)\}} t^h q^h(z) (1 - \epsilon^h(z)), \end{aligned}$$

waarbij voor alle h geldt dat consument h 's elasticiteit van de beschikbaarheidsfunctie $\epsilon^h(z) = \frac{\partial q^h}{\partial z} \cdot \frac{z}{q^h(z)} \leq 1$, aangezien q^h een niet-negatieve, stijgende, concave functie is. Onder deze regulariteitsvoorwaarde volgt dus dat $\pi^0(\widehat{p}, I, z) \geq 0$. In het bijzonder volgt dat $\pi^0 = \sum_{\{h|s^h=q^h(z)\}} t^h b^h$ als voor alle h de beschikbaarheidsfunctie wordt gegeven door $q^h(z) = a^h z + b^h$ met $a^h > 0$, $b^h > 0$. Een evenwichtssituatie resulteert dus in een niet-negatieve winst voor de agent in de pps, zodat de agent daadwerkelijk bereid zal zijn om zich te contracteren tot de realisatie en de financiering van de door de principaal vastgestelde omvang van de infrastructuur als de omvang positief is.

Onder de regulariteitscondities bestaat er een evenwicht. De evenwichtsallocatie is toegelaten en in het evenwicht zijn de consumenten en producenten nuts- en winstmaximaliserend. Samen met het feit dat de omvang van de infrastructuur voldoet aan conditie (12) en de opslagen aan conditie (14), volgt hieruit dat de evenwichtsallocatie aan alle in de vorige paragraaf gegeven eerste orde voorwaarden voor efficiëntie voldoet. We hebben dan ook de volgende stelling, voor een bewijs van deze stelling verwijzen we naar van der Laan, Ruys en Talman (2001).

Hoofdstelling 1 *Onder de regulariteitsaannames is er een pps in de economie met één privaat-publieke dienst. Elke evenwichtsallocatie is Pareto-efficiënt.*

In deze paragraaf hebben we een algemeen evenwicht gedefinieerd voor een economie met een publiek-private dienst. In dit evenwicht wordt de benodigde infrastructuur georganiseerd door een pps en privaat gefinancierd via het gebruik van de infrastructuur door middel van opslagen op de prijs van een complementair privaat goed, waarvan de vraag via de markt tot stand komt. Voor een goed functioneren van dit systeem is informatie over de individuele beschikbaarheidsfuncties nodig, immers deze informatie wordt door zowel de publieke autoriteit voor infrastructuur gebruikt om de optimale omvang van het wegennet vast te stellen, als door de private infrastructuur onderneming om de individuele opslagen vast te stellen. In de literatuur over publieke goederen is het verkrijgen van informatie over de individuele voorkeuren een nog nauwelijks opgelost probleem. Uiteraard is ook in dit model het verkrijgen van deze informatie het cruciale punt. Echter, met behulp van moderne informatie-technologie (zoals het werken met chips in auto's en via automatische registratie) is het mogelijk om een overzicht van het individuele gebruik van het wegennet te bereiken. Op basis van algemene kenmerken wat betreft gebruik (hoeveel, welke delen van het wegennet, uren waarop het wegennet het meest wordt gebruikt, type auto, etc.) is het vervolgens mogelijk de gebruikers in te delen in verschillende gebruikersklassen, en op basis van deze indeling de gebruikers te tarifieren. Op deze wijze kan een aanzet worden gemaakt tot de implementatie van een second-best systeem, waarvan de efficiëntie zal toenemen naarmate de classificatie kan worden verfijnd. Bij een volledig doorgevoerde classificatie wordt zelfs een first-best systeem bereikt.

3 Algemeen evenwichtsdenken in Nederland

Binnen de economische wetenschap is er brede overeenstemming over het feit dat de geschiedenis van het algemeen evenwichtsmodel begint met Walras (1874). De ontwikkeling sindsdien, met name nadat Debreu (1959) erin was geslaagd om een existentiebewijs te leveren, is veelvuldig beschreven, zie bijvoorbeeld Ruys (1985) in zijn (tweede) oratie, uitgesproken aan de Technische Hogeschool Eindhoven. Het feit dat ook Nederlandse economen hieraan veel hebben bijgedragen is met name te danken aan Pieter Ruys en Claus Weddepohl. In de jaren zeventig van de twintigste eeuw waren zij de eerste twee Nederlandse economen die zich intensief met dit onderwerp bezig hielden. Hun gezamenlijk onderzoek leidde tot originele resultaten met betrekking tot convexiteit en dualiteit, in 1979 bij Springer gepubliceerd in een

uitgebreid en belangwekkend artikel, en voor beiden een benoeming tot hoogleraar in de wiskundige economie aan respectievelijk de Katholieke Hogeschool Tilburg en de Universiteit van Amsterdam. Zonder twijfel kan worden gezegd dat zonder de voortrekkersrol van Pieter en Claus onderzoek op het gebied van algemeen evenwicht in Nederland niet of nauwelijks van de grond zou zijn gekomen. Inmiddels kan worden gesteld dat Nederland in ruime mate heeft bijgedragen aan de theorie van algemeen evenwicht. Ondermeer zijn hierbij theoretische resultaten bereikt met betrekking tot algemeen evenwicht in economieën met (i) publieke goederen, (ii) overlappende generaties, (iii) (on)uitputbare hulpbronnen en oneindig aantal goederen, (iv) prijsrigiditeiten, (v) toenemende meeropbrengsten, en (vi) onvolledige markten, terwijl ook op het gebied van prijsaanpassingsprocessen, zowel stochastisch als deterministisch en al dan niet in een dynamische context, en op het gebied van de berekening van evenwichten vooraanstaande bijdragen zijn geleverd.

Eind jaren tachtig waren de topjaren van het onderzoek in de algemeen evenwichtstheorie voorbij. Dit werd als een afgesloten hoofdstuk beschouwd. Pieter schreef in zijn oratie van 1985 al: ‘Als concept is het algemeen evenwichtsmodel afgerond en behoort het tot de gevestigde kennis van de economische wetenschap’. Had hij hierin gelijk? Ja en nee! Het algemeen evenwichtsmodel wordt aan alle economische faculteiten gedoceerd, in beperkte mate op bachelor/master niveau, maar prominent in de Ph.D opleidingen. In het economische onderzoek wordt veelvuldig gebruik gemaakt van het algemeen evenwichtsmodel, niet alleen aan universiteiten, maar zeker ook daarbuiten op onderzoeksafdelingen voor economisch beleid, zoals bij het CPB en de DNB. Inderdaad behoort het algemeen evenwichtsmodel tot de gevestigde kennis, ja dus. Echter, het onderzoek op het gebied van algemeen evenwicht is zeker nog niet afgerond. Juist in de tweede helft van de jaren negentig zien we, zeker ook binnen Nederland, een opleving van dit onderzoek, bijv. met betrekking tot onvolledige markten, dynamische prijsvormingsprocessen, oneindige horizon, financiële markten, core-equivalentie, netwerkeconomieën en zelfs met relatie tot evolutionaire processen. Het onderzoek gaat nog steeds door met veelbelovende resultaten.

Uiteraard is de wiskundige economie niet beperkt tot het algemeen evenwichtsmodel. Wel kunnen we stellen dat met dit model de wiskundige economie zich heeft kunnen vestigen als een aparte discipline binnen de economische wetenschappen. Ontwikkelde technieken en verkregen inzichten blijken ook bruikbaar in andere gebieden, zoals in de speltheorie. Pieter heeft dit in zijn werk veelvuldig laten zien, zeker ook bij de begeleiding van promovendi. De wiskundige economie, waaronder we ook de speltheorie mogen rekenen, houdt zich inmiddels bezig met een breed scala van onderwerpen, ook in Nederland. Wiskundige economie heeft dan ook zeker toekomst. Daarvoor is het wel nodig dat de wiskundig economen beseffen dat het daarbij moet gaan om het verkrijgen van meer inzicht in de werking van de economie, wiskunde is en blijft slechts een formele taal om deze werking te bestuderen.

Referenties

- Debreu, G. (1959), *Theory and Value*, John Wiley and Sons, New York.
- Laan, G. van der, P.H.M. Ruys en A.J.J. Talman (2001), “Optimal provision of infrastructure using public-private partnership contracts”, TI-Discussion Paper 2001-011/1, Tinbergen Instituut.
- Ruys, P.H.M. (1974), *Public Goods and Decentralization: The Duality Approach in the Theory of Value*, Dordrecht: Kluwer Academic Publishers.
- Ruys, P.H.M. (1981), *Rationeel gedrag*, Intreerede Katholieke Hogeschool Tilburg.
- Ruys, P.H.M. (1985), *Wiskundige economie en sociale technologie*, Intreerede Technische Universiteit Eindhoven.
- Ruys, P.H.M., en H.N. Weddepohl (1979), “Economic theory and duality”, in J. Kriens (editor), *Convex Analysis and Mathematical Economics*, Lecture Notes in Economics and Mathematical Systems 168, Springer, Berlijn, pp. 1–72.
- Walras, L. (1884), *Eléments d' économie politique pure*, L. Corbaz, Lausanne.

BIOGRAFIEËN VAN DE AUTEURS

René van den Brink promoveerde in 1995 aan de Katholieke Universiteit Brabant (KUB) op het onderwerp ‘Relational Power in Hierarchical Organizations’ met als promotor Pieter Ruys en als co-promotor Rob Gilles. Direct na zijn promotie was hij een jaar post-doc bij de vakgroep Econometrie van de Vrije Universiteit Amsterdam (VU). Vanaf 1997 was hij weer enkele jaren verbonden aan de KUB (o.a. als post-doc op het NWO-project ‘Oneerlijke Concurrentie en Overheidsbelang’). Gedurende dit verblijf aan de KUB heeft hij, behalve aan de KUB, gedoceerd aan o.a. de universiteiten van Nijmegen, Utrecht en Arlington (Texas, USA). Sinds september 2001 is hij wederom werkzaam bij de VU, nu als universitair docent Wiskundige Economie.

Edward Droste promoveerde in 1999 aan de Katholieke Universiteit Brabant met als promotor Pieter Ruys. De titel van zijn proefschrift is ‘Adaptive behavior in economic and social environments’. Na zijn promotie was hij als consultant werkzaam bij KPMG op het gebied van marktwerking en mededinging. Sinds 2002 werkt hij als adviseur Regulatory Affairs bij Essent.

Rob Gilles studeerde in mei 1986 *Cum Laude* af aan de Katholieke Hogeschool Tilburg als wiskundig econoom bij Pieter Ruys. Het begonnen onderzoek werd voortgezet in het NWO project “Competitive Equilibrium in an Economy with a Non-Profit Sector, and Comparison with Allocation Rules in the Health Sector”. Rob Gilles promoveerde in april 1990 *Cum Laude* op het proefschrift ‘Core and Equilibria of Socially Structured Economies: The Modelling of Social Constraints in Economic Behaviour’ met Pieter Ruys als zijn promotor. Na een jaar als universitair docent aan de KUB, is hij sinds 1991 verbonden aan de Virginia Polytechnic Institute and State University te Blacksburg, Virginia, USA. Daar was hij van 1998 tot 2001 Director of Graduate Studies.

Wim Hafkamp studeerde in oktober 1977 af aan de Katholieke Hogeschool Tilburg, als econometrist bij Pieter Ruys, en werd wetenschappelijk medewerker bij de vakgroep Wiskunde en Statistiek van de economische faculteit van de UvA. Hij promoveerde in 1983 aan de Vrije Universiteit op het proefschrift ‘Triple Layer Model: A National-Regional Environmental-Economic Model for The Netherlands’. Promotor was Peter Nijkamp, terwijl Pieter Ruys optrad als co-promotor. Wim Hafkamp ging vervolgens een jaartje naar Amerika, als visiting professor bij het Geography Department van Boston University. In 1984 stapte hij over van de UvA naar de VU, en werd daar hoofd van de Economisch-Technologische Afdeling van het Instituut voor Milieuvraagstukken. In 1988 verliet hij ‘academia’ om senior-adviseur te

worden bij de management consultants van Deloitte en Touche, waarna hij in 1992 met de hele milieugroep overstapte naar KPMG. Ondertussen werd hij in 1990 benoemd tot bijzonder hoogleraar Milieu- en Natuurvraagstukken aan de Katholieke Universiteit Brabant, een unieke leerstoel, ingesteld door een drietal Brabantse milieu-organisaties. In 1994 werd hij benoemd tot hoogleraar Milieukunde aan de Erasmus Universiteit Rotterdam, waar hij leiding ging geven aan het Erasmus Studiecentrum Milieukunde. Sinds december 2001 is hij daar decaan van de Faculteit Sociale Wetenschappen. Hij is tevens lid van de VROM-raad en de Raad voor Verkeer en Waterstaat.

Yuan Ju studied Money and Banking for a Bachelor's degree during 1993-1997 and Industrial Organization for a Master's degree during 1997-2000 at Shandong University, China. He followed the Graduate Program in Economics at CentER in 2000 as a guest student and started his Ph.D. research at CentER and the Department of Econometrics and Operations Research, Tilburg University, the Netherlands in 2001. He has made several academic visits in PURC, University of Florida, USA, and IDEI, Toulouse, France. His main research interests are Game Theory, Industrial Organization and Network Economics.

Jan Koning was tot 1984 als wetenschappelijk hoofdmedewerker verbonden aan de Katholieke Universiteit Brabant (KUB)/Vakgroep Geld-, Krediet- en Bankwezen. Hij promoveerde in 1982 op het onderwerp 'Kredietrantsoenering en Onevenwichtigheid' met als promotoren Prof. dr. H.W.J. Bosman en Prof. dr. P.H.M. Ruys. Vanaf 1992 is Jan Koning algemeen directeur van de GBF te Amsterdam.

Michael Kosfeld is Assistant Professor at the Institute for Empirical Research in Economics at the University of Zurich. He received his PhD at Tilburg University in 1999, where he wrote his dissertation on 'Individual Decision-Making and Social Interaction' supervised by Pieter Ruys and Dolf Talman. Before he moved to Zurich in 2000, he was a Marie Curie Research Fellow at the CentER for Economic Research in Tilburg. During the academic year 2001/2002 he was a Fellow at the Institute for Advanced Study in Berlin.

Gerard van der Laan studeerde in 1975 af als wiskundig ingenieur aan de Technische Universiteit Delft, en promoveerde in 1980 aan de Vrije Universiteit Amsterdam op het onderwerp 'Simplicial Fixed Point Algorithms'. Samen met Dolf Talman en anderen publiceerde hij vele artikelen over dit onderwerp in onder andere Mathematics of Operations Research, en Mathematical Programming. Daarnaast publiceerde hij vooral op het gebied van de algemeen evenwichtstheorie en de speltheorie. Hij is sinds 1975 verbonden aan de Vrije Universiteit, sinds 1988 als hoogleraar Wiskundige Economie.

Chris van Raalte promoveerde in 1996 aan de Katholieke Universiteit Brabant (KUB) op het onderwerp 'Market Formation and Market Selection' met als promotor Pieter Ruys en als co-promotor Rob Gilles. Tijdens zijn aanstelling aan de KUB nam hij deel aan het Graduate Program in Economics van CentER. Bovendien verbleef hij enige tijd aan Virginia Polytechnic Institute and State University te

Blacksburg, Virginia (USA). Sinds april 1998 is hij werkzaam als software engineer bij ICTRO (Ministerie van Justitie).

Radislav Semenov received his PhD at Tilburg University in 2000, where he wrote his dissertation on ‘Cross-Country Differences in Economic Governance: Culture as a Major Explanatory Factor’ under supervision of Pieter Ruys. Since November 1999 he is a post-doc at the Catholic University of Nijmegen.

Willy Spanjers, geboren in 1964, promoveerde in 1992 aan de Katholieke Universiteit Brabant (KUB) op het proefschrift ‘Price Setting Behavior in Hierarchically Structured Economies’ met Pieter Ruys als promotor en Rob Gilles als co-promotor. Na zijn promotie was hij tot eind 1994 werkzaam als medewerker aan het Institut für Mathematische Wirtschaftsforschung (IMW) aan de Universiteit van Bielefeld. Gedurende deze tijd doceerde hij tevens aan de Faculteit voor Wiskunde en Informatica van de Universiteit van Leipzig. Vanaf 1995 was hij als Assistent verbonden aan de leerstoel Theoretische Economie van de Universiteit van het Saarland. Deze verleende hem in 2000 de *venia legendi* in de Algemene Economie op basis van zijn Habilitationsschrift *Liquidittversorgung unter Ungewiheit: Ein theoretischer Vergleich von Banken und Mrkten*. Gedurende 1999/2000 was hij werkzaam als lecturer aan de Universiteit van Birmingham, waarna hij in 2002 naar de Universiteit van het Saarland terugkeerde als waarnemend hoogleraar Theoretische Economie. Gedurende deze tijd doceerde hij o.a. als Commerzbank-Gastprofessor voor Monetaire en Financiële Economie aan de Technische Universiteit Chemnitz, die hem in 2001 de *venia legendi* in de Algemene Economie verleende. Aansluitend werd hij in 2002 waarnemend hoogleraar Macroeconomie aan de TU Chemnitz, waarna hij een half jaar gastdocent aan de Universiteit van Birmingham was. Sinds maart 2003 is hij weer aan de TU Chemnitz verbonden, nu als waarnemend hoogleraar Internationale Economie.

Ton Storcken promoveerde in 1989 aan de Katholieke Universiteit Brabant op het proefschrift met de titel ‘Possibility Theorems for Social Welfare Functions’. Pieter Ruys, Stef Tijs en Harrie de Swart waren zijn promotoren. Van 1984 tot 1992 was hij onderzoeker aan die zelfde universiteit: de eerste vier jaar bij NWO, de laatste vier jaar bij het Samenwerkings Orgaan Brabantse Universiteiten. Vanaf 1992 is hij werkzaam aan de Universiteit Maastricht in het begin als universitair docent en inmiddels als universitair hoofddocent.

Dolf Talman studeerde econometrie aan de Rijksuniversiteit Groningen en promoveerde op 29 februari 1980 in de wiskunde en natuurwetenschappen aan de Vrije Universiteit te Amsterdam op het onderwerp ‘Variable Dimension Fixed Point Algorithms and Triangulations’. Na zijn promotie was hij een jaar visting lecturer bij de School of Organization and Management en de Cowles Foundation van Yale University in New Haven. Vanaf juni 1981 is hij werkzaam bij de vakgroep Econometrie en Operations Research van de Universiteit van Tilburg, eerst als medewerker en universitair hoofddocent en vanaf 1991 als hoogleraar Speltheorie en Evenwichtsprogrammeren. Behalve aan Yale en de Universiteit van Tilburg, heeft hij gedoceerd aan onder andere de universiteiten van Utrecht, Tsukuba en Kanton.

Claus Weddepohl was van 1962 tot 1980 verbonden aan de Katholieke Hogeschool Tilburg (tegenwoordig Universiteit van Tilburg) als assistent respectievelijk (hoofd)-medewerker. In 1971 promoveerde hij in Tilburg op het proefschrift "Axiomatic choice models and duality". In het studiejaar 1974-'75 was hij als research fellow verbonden aan het CORE (Center for Operations Research and Econometrics) te Leuven. Van 1980 tot 2002 was hij hoogleraar Wiskundige Economie aan de Universiteit van Amsterdam.

Na zijn afstuderen als wiskundig econoom aan de Universiteit van Tilburg in 1974 werkte **Cees Withagen** als wetenschappelijk medewerker achtereenvolgens aan de TU Delft, de Universiteit van Amsterdam en de TU Eindhoven. In 1984 promoveerde hij bij Pieter Ruys en Claus Weddepohl aan de Universiteit van Tilburg op een proefschrift over internationale handel en niet-vernieuwbare natuurlijke hulpbronnen. Inmiddels is hij sinds een aantal jaren als hoogleraar milieueconomie verbonden aan de Vrije Universiteit te Amsterdam en de Universiteit van Tilburg. Zijn belangstelling in het onderzoek gaat vooral uit naar duurzame ontwikkeling (mede in een evolutionair perspectief) en de relatie tussen milieubeleid en internationaal concurrentievermogen.