TVW

Tijdschrift Vervoerswetenschap
Jaargang 57, nummer 1
Augustus 2021
pagina 83 - 101
ISSN: 1571-9227
www.vervoerswetenschap.nl

Allocatie van schaarse capaciteit bij gedeeltelijke lock-down

Erik Verhoef
VU Amsterdam, afdeling Ruimtelijke Economie[1]

De corona crisis heeft op vele fronten en zeker ook in het openbaar vervoer geleid tot grote uitdagingen in de allocatie van schaarse capaciteit. Lockdown maatregelen hebben effectieve capaciteiten tot soms een fractie van het gebruikelijke niveau, terwijl angst voor besmettingen het voor veel mensen onaantrekkelijk maakt om dichtgepakt op anderen langere tijd door te brengen. In dit artikel bestuderen we de vraag hoe in dergelijke gevallen de beperkte capaciteit zo goed mogelijk verdeeld kan worden. We beschouwen dit vanuit het perspectief van verschillende beoordelingscriteria: efficiëntie, effectiviteit, en maatschappelijke acceptatie. Voor de beoordeling van de efficiëntie van verschillende allocatieprincipes ontwikkelen we een conceptueel model. Effectiviteit en mogelijke acceptatie schatten we op kwalitatieve wijze in. We concluderen dat (1) verschillende methoden om schaarse capaciteit te rantsoeneren behoorlijk kunnen verschillen op criteria van efficiëntie, effectiviteit, acceptatie; (2) welvaartsverliezen door inefficiënte rantsoenering groter worden naarmate de vraag inelastischer is en de rantsoenering strikter; (3) verhandelbare of verruilbare rechten een voorbeeld zouden kunnen zijn van een instrument dat efficiënt en effectief is, en mogelijk voldoende draagvlak zou kunnen kennen.

Trefwoorden: allocatie, capaciteit, COVID-19, mobiliteit, prijsbeleid, verhandelbare rechten

---

[1] VU Amsterdam, afdeling Ruimtelijke Economie, E: e.t.verhoef@vu.nl

1. Inleiding

De corona crisis heeft een klassiek economisch probleem weer prominent op de kaart gezet: hoe kan maatschappelijke schaarste zo goed mogelijk gealloceerd worden?[2] Het vraagstuk speelt op vele fronten. In de meest donkere scenario's, die tot op heden gelukkig nooit werkelijkheid zijn geworden, moet zelfs de zware ethische kwestie over het toewijzen van IC-capaciteit wanneer de grens bereikt is onder ogen worden gezien. Maar op een groot aantal, ethisch minder zwaar beladen maar daarmee niet direct gemakkelijk op te lossen vraagstukken en beleidsterreinen, speelt ook de fundamentele vraag hoe om te gaan met plotselinge fluctuaties in schaarste ten gevolge van lock-down maatregelen en versoepelingen daarvan. Concrete voorbeelden zijn openbaar vervoer, zitplaatsen bij restaurants of in theaters, de bezetting van werkplekken, toegang tot winkelfaciliteiten, aanwezigheid bij (sport-)evenementen, en ga zo maar door.

Bezettingsgrenzen die soms ver onder de reguliere fysieke capaciteit liggen, zoals 40% zitplaatsbezetting in OV dat normaalgesproken in de spits overbezet is met meer reizigers dan zitplaatsen, roepen onherroepelijk de vraag op wie wél en wie níet op een bepaald moment "kan" of "mag". Het is ongetwijfeld mede vanwege de volatiliteit van capaciteit, en de bijzondere omstandigheden van een pandemie, dat deze vragen zo scherp in beeld komen. Immers, de allocatie van goederen en diensten over burgers is een vraagstuk dat in bijna al onze consumptieve activiteiten speelt, met uitzondering van de consumptie van puur publieke goederen waarvoor consumptie in economische jargon "niet rivaliserend" en "niet uitsluitbaar" is.[3] In moderne kapitalistische samenlevingen als de onze is "de markt" de gebruikelijke manier om die allocatie tot stand te laten komen. Markten werken daarbij gewoonlijk niet perfect, en het bijsturen van dergelijk marktfalen is een belangrijke economische taak voor de overheid. Hierbij gaat het bijvoorbeeld om het beperken of voorkomen van milieuschade en andere maatschappelijke risico's, van marktmacht, van uitbuiting of ander onethisch gedrag, enzovoort. Tegelijkertijd wordt de aldus gecontroleerde markt waarin consumenten keuzevrijheid behouden en ondernemers kunnen blijven ondernemen doorgaans wel als een wenselijker allocatiemechanisme gezien dan alternatieven in de richting van staatsgeleide economieën.

Het is desalniettemin volstrekt begrijpelijk dat voor het alloceren van plotselinge schaarste ten gevolge van coronamaatregelen niet direct op het marktmechanisme wordt vertrouwd. Immers, (veel) grotere schaarste zou zich vertalen in (veel) hogere prijzen; bijvoorbeeld voor het OV, een plaatsje op het terras, of in het theater. Daarmee komen verdelingsvraagstukken prominent in beeld: het zal voor velen als ronduit onwenselijk en/of onacceptabel worden gezien wanneer door coronamaatregelen met name alleen de hoogste inkomensgroepen toegang zouden blijven houden tot de betreffende diensten en producten. Juist vanwege de onvoorspelbaarheid van duur en mate van schaarste lijkt de gewoonlijk op welvaartsgronden te prefereren uitweg uit dilemma's tussen efficiëntie van prijzen en eerlijkheid van verdeling, namelijk het systematisch baseren van prijzen op schaarste en het tegelijkertijd veiligstellen van een acceptabele verdeling op basis van direct inkomensbeleid en toegang tot onder meer onderwijs en arbeidsmarkten, praktisch onuitvoerbaar. Tegelijkertijd lijkt een succesvol beroep op zelfbeperking in consumptief gedrag wanneer capaciteitsgrenzen overschreden dreigen te worden slechts een beperkte houdbaarheid te hebben. Dit hebben we bijvoorbeeld al kunnen zien bij het veel minder strikt naleven van "verzoeken-met-

[2] Welbeschouwd gaat het hier in feite om wat vaak als het centrale kenobject van de economische wetenschappen wordt gezien: het bestuderen van individuele en collectieve keuzevraagstukken in condities van schaarste.

[3] Bijvoorbeeld: bescherming tegen overstromingen door dijken voor de ene persoon gaat niet ten kosten van de bescherming van de andere persoon (niet rivaliserend), en het is niet mogelijk om door het vragen van een prijs voor bescherming deze alleen te bieden aan degenen die de prijs daadwerkelijk willen betalen (niet uitsluitbaat). Voor private goederen is geen van beiden het geval: het broodje dat ik eet kan niet door iemand anders genuttigd worden, en ik kies zelf op basis van de prijs of ik het broodje inderdaad wil kopen.

klem" vanuit overheden in de latere coronagolven, in vergelijking met de eerste.

Dit alles roept de vraag op, op welke wijzen we plotselinge schaarste ten gevolge van lock-down maatregelen zouden kunnen alloceren, en hoe deze zich verhouden in termen van welvaart, en in termen van rechtvaardigheid. Dat is de vraag waar ik in dit artikel licht op wil werpen; concreet door voor een aantal mogelijke allocatiemechanismen te bespreken hoe deze zullen presteren op verschillende relevante criteria: efficiëntie en maatschappelijke welvaart; effectiviteit; en acceptatie. Deze vragen zijn van onmiddellijk belang als we in de komende tijd geleidelijk aan, met het uitvoeren van het vaccinatieprogramma, kunstmatig ingestelde schaarste op diverse markten in stappen weer kunnen opheffen, maar ook weer moeten aanscherpen wanneer nieuwe mutaties van het virus zich openbaren en snel verspreiden. De vraag is ook op langere termijn van belang, niet alleen voor het geval we getroffen gaan worden door een nieuwe pandemie, maar ook wanneer we bijvoorbeeld vanwege klimaatdoelstellingen consumptie op bepaalde markten – denk bijvoorbeeld aan vliegreizen of autogebruik – zouden willen (blijven) beperken.

De vraag klinkt eenvoudig maar zou voor gedegen beantwoording tenminste een serie proefschriften verdienen, die bijvoorbeeld in detail zouden kijken naar uitwerkingen voor specifieke markten en via gericht enquête onderzoek de gepercipieerde rechtvaardigheid van concrete beleidsopties in specifieke contexten in beeld zouden kunnen brengen, en via feitelijke gedragsexperimenten de gedragsreacties van de betreffende consumenten in beeld zouden kunnen proberen brengen. De situatie is echter dat dit paper slechts een bescheiden onderdeel is, van een economisch sub-programma, in een kortlopend project. We zullen daarom ook voor een bescheiden opzet kiezen.

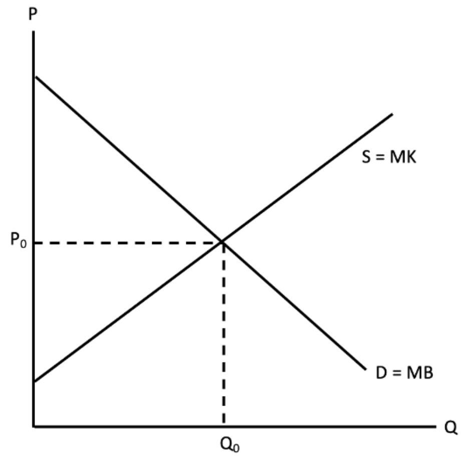
We beginnen in paragraaf 2 met een korte uiteenzetting over hoe normaalgesproken markten reageren op veranderende schaarsteverhoudingen; onder welke omstandigheden en in welke zin dit maatschappelijk welvaarts-maximaal en efficiënt is; en hoe de relatie met eerlijkheid en rechtvaardigheid ligt. Deze benchmark helpt bij het positioneren van de alternatieve allocatiemechanismen die we zullen beschouwen. Paragraaf 3 presenteert vervolgens een stilistische analyse die op conceptueel niveau illustreert hoe het gebruik van marktwerking, ook in situaties van snel veranderende schaarsteverhoudingen, allocatievraagstukken maatschappelijk gezien efficiënter, en daarmee met hogere welvaart tot gevolg, adresseert. In paragraaf 4 bespreken we aan de hand daarvan een aantal mogelijke manieren om snel veranderende schaarste, bij gedeeltelijke lock-down en versoepeling, te alloceren. Paragraaf 5 geeft de conclusies.

2. Schaarste en prijsmechanisme: de basis

Het is goed om eerst in zijn algemeenheid kort stil te staan bij enkele fundamentele aspecten van de werking van het prijsmechanisme voor het alloceren van schaarste. Natuurlijk valt daar heel veel over te zeggen (zie micro-economische tekstboeken zoals dat van Varian, 2010; of Mas-Colell, 1995), en kunnen voorbeelden zo moeilijk gemaakt worden als we maar willen, maar belangrijke inzichten kunnen al verkregen worden door te kijken naar het eenvoudigst denkbare geval: het tekstboekvoorbeeld waar er sprake is van één geïsoleerde markt voor één product, waarop vragers en aanbieders zonder marktmacht, met volledige informatie, en in afwezigheid van externe effecten zoals milieuvervuiling, actief zijn. Aan het eind van de paragraaf bespreken we waarom dergelijke complicaties, relevant in de praktijk, de basisinzichten uit deze paragraaf niet zullen laten veranderen maar eerder juist aan zeggingskracht laten winnen omdat voor het bijsturen van dergelijke vormen van marktfalen deze inzichten juist heel behulpzaam zijn.

Het allocatie-vraagstuk in deze setting omvat de volgende vragen: welke productiehoeveelheid dient bereikt te worden, welke consumenten zouden daarmee bediend moeten worden, welke producenten zouden de productie voor hun rekening moeten nemen, en welke prijs hoort daarbij? En, zijn dat ook de prijs, de hoeveelheid, en de verdeling over consumenten en producenten die in een vrije markt tot stand komt?

Deze vragen laat zich beantwoorden aan de hand van een gebruikelijk marktdiagram, zoals in Figuur 1 geschetst. De dalende lijn representeert de inverse vraagfunctie[4], die de relatie geeft tussen prijs P en gevraagde hoeveelheid Q. Wat in dit verband van belang is om te benadrukken, is dat de functie daarmee tegelijkertijd ook de marginale baten, MB, geeft. De reden dat mogelijke afzet rechts van het geschetste evenwicht Q_0 niet verkocht wordt bij een prijs P_0 is dat de baten die de betreffende consumenten toekennen aan die eenheden, blijkbaar kleiner zijn dan de baten die ze toekennen aan het in geld aanhouden van P_0 en zodoende ergens anders voor te kunnen aanwenden. Omdat de vraagfunctie D daarmee de marginale baten MB geeft, representeert het oppervlak onder de vraagfunctie, tussen 0 en een bepaalde Q, de totale baten die met de consumptie van die hoeveelheid Q samenhangen.



Figuur 1: Een standaard marktdiagram

Zonder marktmacht kunnen we ervan uitgaan dat bedrijven goederen blijven aanbieden zolang de kosten van een extra eenheid, de marginale kosten MK, niet hoger zijn geworden dan de prijs. De aanbodfunctie S, die het verband geeft tussen prijs en aangeboden hoeveelheid, valt daarom samen met de MK functie. Het marktevenwicht dat tot stand komt is bij hoeveelheid Q_0 en prijs P_0.

Het is hiermee ook inzichtelijk waarom het gebruik van het prijsmechanisme in dit voorbeeld (zonder marktfalen) tot een efficiënte allocatie leidt, dus: die allocatie die de maatschappelijke welvaart maximaliseert. In toegepast onderzoek wordt welvaart daarbij doorgaans geoperationaliseerd als maatschappelijk surplus; de som van consumenten surplus, producenten surplus, en overheidssurplus. Dit is minder restrictief dan het misschien lijkt: voor alle zogeheten "Paretiaanse" maatschappelijke welvaartsfuncties, een brede klasse welvaartsfuncties die gemeen hebben dat de maatschappelijke welvaart stijgt wanneer niemand er op achter uitgaat en ten minste een individu er op vooruitgaat, geldt dat maximalisatie van de welvaartsfunctie ook betekent dat het maatschappelijk surplus gemaximaliseerd is. Dit omvat daarmee een brede klasse aan welvaartsfuncties, inclusief het Rawlsiaanse principe dat we de maatschappelijke welvaart meten als het welbevinden van de individu die het minst goed af is in de maatschappij. Zolang men dus van mening is dat een maatschappij beter af is wanneer tenminste een individu beter af is en niemand slechter af is, is maximalisatie van maatschappelijk surplus een noodzakelijke voorwaarde voor welvaartsmaximalisatie. Zie ook Atkinson and Stiglitz (2015).

[4] De reguliere vraagfunctie geeft gevraagde hoeveelheid als functie van de prijs; omdat in Figuur 1 de assen zijn gewisseld noemen we het de inverse vraagfunctie, die voor elke hoeveelheid geeft bij welke prijs die hoeveelheid gevraagd zou worden. Deze weergave is dusdanig gebruikelijk dat in het dagelijks spraakgebruik het bijvoeglijk naamwoord "inverse" nogal eens achterwege wordt gelaten.

Terug naar Figuur 1. Doordat in het evenwicht de marginale baten en marginale kosten aan elkaar gelijk zijn, is in de eerste plaats de omvang van de productie en consumptie optimaal. Bij een verkleining van de productie ten opzichte van het evenwicht ontstaat er een steeds grotere wig tussen MB en MK, die aangeeft dat het dan weer vergroten van productie meer baten oplevert dan kosten. Het omgekeerde is het geval rechts van Q_0: de MK is groter dan de MB en het verlagen van de productie bespaart meer kosten dan dat het in termen van gederfde baten kost.

Minstens zo belangrijk, maar minder vaak benoemd, en zeer relevant voor het onderwerp van dit paper: niet alleen de *omvang* van de productie en consumptie is optimaal, maar ook de *verdeling* daarvan over consumenten en producenten. We zagen al dat vanwege de prijs P_0 consumenten alleen maar eenheden zullen consumeren die een hogere bate vertegenwoordigen. Eenheden die zich rechts van Q_0 op de horizontale as bevinden, worden niet geconsumeerd, omdat door het prijsmechanisme de afweging bij de consument zélf ligt. Tegelijkertijd, de productie wordt zodanig over aanbieders verdeeld dat voor geen van hen geldt dat er een eenheid met hogere MK dan P_0 zal worden voortgebracht. Het is dus niet alleen de *omvang* maar vooral ook juist ook de *samenstelling* van productie en consumptie, ofwel de verdeling daarvan over producenten en consumenten, die optimaal is. Een overheid die in een centraal geleide economie van bovenaf de allocatie bepaalt zal, zelfs als Q_0 toevallig goed gekozen zou zijn, naar alle waarschijnlijkheid nooit dezelfde allocatie bereiken van productie over aanbieders, en consumptie over huishoudens. Het prijsmechanisme doet dit, doordat de keuze van productie en consumptie "gedecentraliseerd" wordt door de prijs: er is zelf-selectie doordat degene die de relevante informatie heeft (de individuele consument en producent) ook de beslissing neemt om wel of niet, en wanneer, te consumeren of produceren. Het is gebruikelijk om in educatieve teksten alleen de optimale *omvang* te benoemen voor de uitleg van efficiëntie van markten; de optimale *verdeling* is echter minstens zo belangrijk om te benadrukken.

Natuurlijk heeft dit inzicht grote consequenties voor de efficiëntie van allocatiemechanismen in tijden van corona lock-downs. Als bijvoorbeeld de overheid bepaalt wie beperkte OV- of theater-capaciteit op een bepaalde dag mag gebruiken is de kans levensgroot dat het resultaat zal zijn een mix van mensen met relatief hoge baten en mensen met relatief lage baten voor gebruik specifiek op die dag. Er is dan nog wel zelfselectie van consumenten op basis van marginale baten die tenminste boven de gehanteerde marktprijs liggen, maar de extra restrictie wordt niet op efficiënte wijze gealloceerd. Allocatiemechanismen die het prijsmechanisme benutten verdienen om die reden de voorkeur.

Het geschetste mechanisme is niet alleen relevant op geaggregeerd niveau, dus voor volledige markten, maar ook op individueel niveau; dus, wanneer D in Figuur 1 de vraag van een enkele individu zou weergeven. Bijvoorbeeld: als het beleid tot gevolg heeft dat een bepaalde reiziger nog maar 2 dagen per week met het OV mag reizen, zal het verlies aan baten voor die persoon en daarmee ook voor de maatschappij als geheel kleiner zijn, en het beleid maatschappelijk gezien dus efficiënter, als die individu zelf kan kiezen welke 2 dagen in de week hij dan reist, in vergelijking met de situatie waarin de overheid dat voor hem besluit. De individu kiest zelf de twee dagen waar de baten het hoogst zijn, zogezegd links op zijn inverse vraagfunctie voor treinreizen.

Kijken we op meer geaggregeerd niveau dan blijven de afwegen rond efficiëntie en maatschappelijke welvaart vergelijkbaar, maar komen ook vraagstukken rond verdeling en (gepercipieerde) eerlijkheid nadrukkelijk in beeld. De hoogst gewaardeerde consumptie, links op de inverse vraagfunctie, zal vaak de bijbehorende hoge betalingsbereidheid kennen omdat de consument achter die vraag een hoog inkomen heeft. Toch is het niet zo dat het beschouwen van baten, kosten en efficiëntie betekent dat verdelingsvraagstukken in beleid buiten beeld zouden vallen. In de eerste plaats valt op te merken dat, zoals hierboven uitgelegd, er ook binnen de consumptiebundel voor één individu verschil zal zijn in baten die aan de verschillende eenheden worden toegekend. Daar speelt het verdelingsvraagstuk niet, en is de efficiëntie van allocatie via prijzen toch al relevant. Een tweede observatie is dat de efficiëntiemaatstaf juist bedoeld is om onafhankelijk van politiek bepaalde inkomens(her-)verdeling gebruikt te kunnen worden. Met andere woorden: de efficiëntiemaatstaf geeft niet een impliciet oordeel over de bestaande inkomensverdeling – inclusief de herverdeling van inkomens die door ander beleid al bereikt is – maar neemt die als gegeven door de daaruit resulterende effecten op betalingsbereidheid onverdund mee te nemen. Daarbij dient ook opgemerkt te worden dat het bereiken van bepaalde doelen ten aanzien van verdeling, gemotiveerd uit rechtvaardigheidsoverwegingen, doorgaans juist op efficiëntere wijze gerealiseerd kan worden (met name door direct inkomensbeleid) dan door inefficiënte allocaties in markten zoals in Figuur 1 te introduceren of te laten voortbestaan, door daar niet een maximalisatie van baten na te streven. Met andere woorden, ongemak met de bestaande inkomensverdeling, die leidt tot een bepaald patroon van baten over inkomensklassen doordat de waarde die verschillende individuen aan een extra Euro toekennen niet gelijk is, zou zich moeten vertalen in gericht inkomensbeleid en, voor de langere termijn, gericht beleid om toegang tot onder meer scholing en arbeidsmarkten te verzekeren. Het introduceren van additionele verstoringen elders in de economie, bijvoorbeeld door de op zichzelf efficiënte werking van het prijsmechanisme in een markt als in Figuur 1 te verstoren, vormt een maatschappelijk kostbaarder manier om herverdelingsdoelen te bereiken.

Tot slot komen we nog even terug op de aanname bij Figuur 1 dat er geen andere vormen van marktfalen zijn. De reden dat deze aanname niet cruciaal is voor de eerdere uiteenzetting is dat dezelfde benadering, met dezelfde overwegingen rond optimaliteit van prijzen en hoeveelheden, relevant blijft wanneer we marktfalen in de beschouwing betrekken. Een raamwerk als in Figuur 1, maar dan verrijkt met de waarde van bijvoorbeeld milieuvervuiling als het om externe effecten gaat, of een discrepantie tussen marktprijzen en marginale kosten als het om marktmacht gaat, blijft relevant als we in een dergelijke situatie geïnteresseerd zijn in bijvoorbeeld het bepalen van het verschil tussen marktevenwicht en maatschappelijk optimum, en de meest efficiënte manier om zo'n optimum te bereiken. Beleidsconclusies kennen dan ook vaak parallellen met wat hierboven betoogd werd, bijvoorbeeld wanneer het gaat om de efficiëntie van prijsbeleid bij het beperken van externe effecten zoals milieuvervuiling of verkeerscongestie.

Om een voorbeeld te noemen: met eenzelfde raamwerk is uit te leggen waarom een beleid met nummerborden om stedelijk autoverkeer te beperken, zoals in Athene of Beijing waar het eindcijfer van het nummerbord bepaalt op welke dagen de auto wel of niet gebruikt mag worden, minder efficiënt is dan een beleid dat gebruik maakt van road pricing, zelfs als de twee instrumenten tot eenzelfde afname van autoverkeer leiden. Het nummerbord maakt namelijk bepaalde verplaatsingen met hogere baten, dus links op de inverse vraagfunctie, onmogelijk; en staat juist andere ritten rechts op die functie toe. De uitkomst is minder efficiënt, en de maatschappelijke welvaart is lager, dan met prijsbeleid.

3. Allocatie bij plotselinge veranderingen in schaarste: conceptuele analyse

Zoals we hebben ondervonden kan vanwege Covid-maatregelen de beschikbare of toelaatbare capaciteit van bepaalde voorzieningen plotseling lager worden dan wat deze voorheen was. Wanneer de prijs niet tegelijkertijd stijgt zodat aan de bijbehorende maximale bezetting wordt voldaan, zou zonder verder ingrijpen een te hoge vraag bestaan en dienen aanvullende maatregelen genomen te worden om de vraag te rantsoeneren. Dat kunnen verschillende maatregelen zijn, zoals we in paragraaf 4 zullen bespreken, die op verschillende mogelijke criteria kunnen worden vergeleken alvorens tot een keuze te komen. Een veel gebruikte vierdeling daarbij is tussen criteria ten aanzien van:

1. Efficiëntie: in welke mate wordt een doel tegen zo laag mogelijke netto maatschappelijke kosten[5] bereikt, en in welke mate is het doel op zichzelf maatschappelijk optimaal gekozen?
2. Effectiviteit: in welke mate wordt het gestelde doel überhaupt bereikt?
3. Eerlijkheid: in welke mate zijn de principes die aan het ontwerp van de maatregel ten grondslag liggen, en is de verdeling van effecten die de maatregel tot stand brengt, eerlijk?
4. Acceptatie: in welke mate heeft het instrument draagvlak bij bevolking en/of beleidsmakers?

Van deze criteria is eerlijkheid waarschijnlijk het moeilijkst om eenduidig te operationaliseren in een wetenschappelijke setting, juist omdat aan de operationalisering niet wetenschappelijk-falsifieerbare overwegingen ten grondslag liggen. Bijvoorbeeld: het valt niet wetenschappelijk te weerleggen dat het "polluter pays principle" eerlijk is omdat het de rekening van milieuvervuiling neerlegt bij degenen die het veroorzaken; het is ook niet wetenschappelijk te weerleggen dat het oneerlijk is zolang de inkomensverdeling oneerlijk wordt gevonden, omdat koopkracht om te vervuilen dan op oneerlijke wijze varieert over groepen en individuen. Ook kan men vervolgens van mening blijven verschillen over de vraag of als het polluter pays principle oneerlijk uitpakt, dat een reden zou moeten zijn om met gericht inkomensbeleid de gewenste inkomensverdeling te implementeren maar het principe wel te implementeren, dan wel het principe niet te implementeren maar verder ook geen extra actie op het bereiken van een meer wenselijke inkomensverdeling uit te voeren. Wetenschap kan natuurlijk wel behulpzaam zijn bij het in kaart brengen van mogelijke definities van eerlijkheid maar niet dicteren welke definitie de juiste is, en bij zoeken naar het antwoord op de vraag of iets eerlijk wordt gevonden; door bepaalde individuen of groepen, onafhankelijk van hoe zij zelf "eerlijkheid" precies definiëren. Dit is een van de belangrijke determinanten van "acceptatie". In dit artikel zal ik niet trachten iets te zeggen over de objectieve eerlijkheid van bepaalde maatregelen, om de redenen zojuist gegeven, maar wel een inschatting trachten te geven van de acceptatie, daarbij mede anticiperend op de te verwachten oordelen over de eerlijkheid van de principes onder bepaalde maatregelen, en de te verwachten verdelingseffecten.

In deze paragraaf gaan we dieper in op het bepalen van de efficiëntie van rantsoeneringsmaatregelen. Effectiviteit en acceptatie zullen in kwalitatieve zin in paragraaf 4 nader aan de orde komen bij het bespreken van een aantal mogelijke maatregelen. Voor efficiëntie geldt dat ook, maar daarbij zullen we ons mede baseren op inzichten uit de conceptuele modellering die we in deze paragraaf bespreken.

De gegeven definitie van efficiëntie impliceert dat deze weergeeft in welke mate maatschappelijke welvaart door een maatregel verhoogd wordt. Bij het stellen en bereiken van het optimale maatschappelijke doel (bijvoorbeeld, hoeveelheid Q_0 in Figuur 1) tegen zo laag mogelijke maatschappelijke netto maatschappelijke kosten (dus, het realiseren van de maximale maatschappelijke baten bij consumptie Q_0, ofwel het oppervlak onder D links van Q_0, en de minimale maatschappelijke kosten voor een productie Q_0, ofwel het oppervlak onder MK links van Q_0) is de maatschappelijke welvaart gemaximaliseerd.

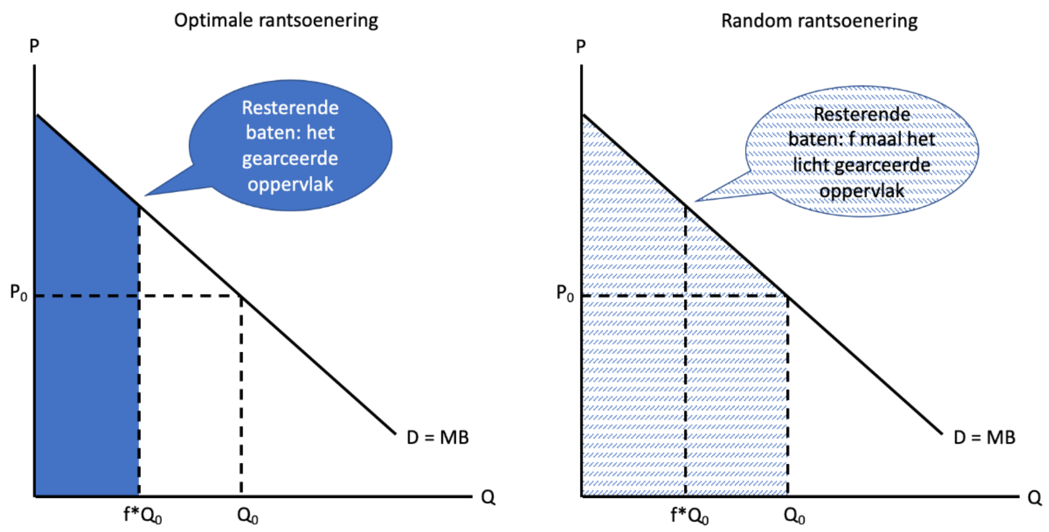
In deze paragraaf richten we ons op de situatie waar het maatschappelijk optimale doel buiten ons analysekader is bepaald. Bij lock-down maatregelen is dat veelal het geval. Maximale bezettingsgraden van bijvoorbeeld het OV, of van terrassen of theaters, worden op basis van medische en epidemiologische overwegingen bepaald. Natuurlijk kunnen daarin fouten gemaakt worden, maar de vraag die we ons stellen is deze: zodra het doel is vastgesteld, hoe groot zijn de netto maatschappelijke kosten om de beoogde rantsoenering te bereiken?

Uit de uiteenzetting in de vorige paragraaf is duidelijk geworden dat dat afhangt van de mate waarin na de rantsoenering de consumptie beperkt wordt tot niet alleen de gewenste omvang,

[5] Hiermee wordt bedoeld: het saldo van maatschappelijke kosten en maatschappelijke baten.

maar daarbij ook tot eenheden die de hoogste baten kennen; dus, het meest linkse gedeelte van de vraagfunctie. Vanuit de theorie is dat een helder inzicht, maar in de praktijk is de eerste belangrijke vraag: hoe groot is het belang dat we daaraan toe moeten kennen? Hoe groot zijn de welvaartsverliezen als we er niet in slagen de optimale samenstelling van consumptie te realiseren met een bepaald rantsoeneringsinstrument? En, wat zijn de determinanten van die welvaartsverliezen; met andere woorden: welke factoren maken die verliezen kleiner of groter, en daarmee ook het belang om rantsoenerings¬instrumenten vanuit dat perspectief te optimaliseren?

Dat is de vraag die we in deze paragraaf willen beantwoorden. We ontwikkelen daarvoor een stilistisch model waarin we voor twee typen vraagfunctie, namelijk de iso-elastische en de lineaire, bepalen hoe groot de relatieve verliezen aan baten zijn als rantsoenering tot een bepaalde hoeveelheid niet in de meest efficiënte manier tot stand komt, dus door het meest linkse gedeelte van de vraagfunctie te behouden, maar in plaats daarvan een volkomen random verdeling over de initiële consumptie kent. Dat laatste is natuurlijk niet de slechtst denkbare uitkomst – dat zou zijn als juist alleen de eenheden met de laagste baten zouden overblijven, rechts op de vraagfunctie – maar het is wel een relevante benchmark. Het geeft de statistische verwachte waarde van de baten wanneer rantsoenering geen systematische relatie met individuele baten kent.



Figuur 2: Overgebleven baten bij optimale rantsoenering (links) versus random rantsoenering (rechts)

Figuur 2 toont dit vraagstuk in de context van de markt die we eerder bij Figuur 1 zagen. Stel voor dat het marktevenwicht met een hoeveelheid Q0 excessief is vanuit het oogpunt van besmettingsrisico's, en dat de maximaal toegestane vraag een bepaalde fractie f daarvan is. Bijvoorbeeld, voor OV werd die maximaal veilige bezetting op 40% van de zitplaatsen gezet (f zou dan in de praktijk lager kunnen zijn dan 0.4, omdat de spitsbezetting van het OV natuurlijk hoger dan 100% van de zitplaatsen kan zijn).[6]

Het linker paneel toont de baten van consumptie die in stand blijven wanneer rantsoenering op optimale, efficiënte wijze plaatsvindt; dus, zodanig dat het meest linkse gedeelte van de

[6] Merk op dat aan het tweede deel van de definitie van efficiëntie, namelijk dat ook Q0 optimaal gekozen moet zijn, voldaan zou zijn als het verschil in MB bij f^*Q_0 versus bij Q_0 gelijk is aan de marginale externe kosten, inclusief de waarde van gezondheidsrisico's die samenhangen met besmettingen. Het bepalen van deze marginale externe kosten is een complex vraagstuk, niet in de laatste plaats omdat een besmetting vervolgens kan leiden tot een keten van nieuwe besmettingen. Ook daarom nemen we de keuze van f als gegeven in deze studie. Overigens kan dat consistent zijn met de situatie waarin het risico op besmettingen bij een bezetting onder f verwaarloosbaar klein is, en vanaf f zodanig snel oploopt dat de marginale externe kosten functie vrijwel verticaal loopt bij f.

vraagfunctie met de hoogste baten bediend blijft worden. Rechts is schematisch in beeld gebracht wat de baten zijn bij random rantsoenering; dus, wanneer bij een onveranderde prijs P_0 willekeurige rantsoenering, zonder relatie met baten, garandeert dat slechts een fractie f van de daarbij horende hoeveelheid Q_0 wordt toegestaan. Bij een eenmalige markt zou elke eenheid een even grote kans f hebben om te worden toegestaan; bij een zich herhalende markt (bijvoorbeeld woon-werkverkeer) zou elke eenheid op een fractie f van de dagen worden toegestaan. In beide gevallen is de statistisch verwachte waarde van de baten een fractie f van de baten die bij het volledig toestaan van Q0 gerealiseerd worden.

Het verschil tussen de baten in beide panelen geeft de maatschappelijke welvaartswinst van het realiseren van een optimale rantsoenering versus een random rantsoenering. De hoogte van dat verschil hangt uiteraard onder meer af van de omvang van de markt, de oorspronkelijke prijs P_0 en de oorspronkelijke hoeveelheid Q_0, en het is daarom ondoenlijk om daar in zijn algemeenheid iets over te zeggen. Echter, de ratio's van de baten zijn ook bijzonder informatief ("welk percentage van de baten blijft behouden bij random rantsoenering in vergelijking met optimale rantsoenering, en waar hangt die verhouding van af?") en zoals zal blijken kunnen we die ratio's bepalen onafhankelijk van Q_0 en P_0. Preciezer gezegd: we kunnen die ratio's uitdrukken in universele parameters: de elasticiteit van de vraag (ε)[7] en de fractie f. We zullen dat doen voor twee mogelijke functionele vormen van de vraagfunctie: de lineaire, en de iso-elastische (waarvoor de elasticiteit ε constant is).

3.1 Lineaire vraag

De lineaire inverse vraagfunctie laat zich als volgt schrijven:

$$P = a - b \cdot Q \quad (1)$$

waarbij a en b positieve parameters zijn. De helling van de functie is constant, maar de elasticiteit daarmee juist niet. In het oorspronkelijke evenwicht is de elasticiteit:

$$\varepsilon_0 = -\frac{1}{b} \cdot \frac{P_0}{Q_0} \quad (2)$$

Zowel a als b laten zich uitdrukken in de elasticiteiten en hoeveelheden in het oorspronkelijke evenwicht:

$$a = P_0 - \frac{P_0}{\varepsilon_0} \quad (3a)$$

$$a = P_0 - \frac{P_0}{\varepsilon_0} \quad (3a)$$

De totale baten die samenhangen met de consumptie van een bepaalde hoeveelheid Q zijn gegeven als het relevante oppervlak onder de inverse vraagfunctie, en laten zich dus schrijven als:

$$B = a \cdot Q - \tfrac{1}{2} \cdot b \cdot Q^2 \quad (4)$$

Dat betekent dat we de baten bij, respectievelijk, het oorspronkelijke evenwicht (B_0), onder optimale rantsoenering (B^O) en onder random rantsoenering (B^R) als volgt kunnen schrijven:

$$B_0 = a \cdot Q_0 - \tfrac{1}{2} \cdot b \cdot {Q_0}^2 \quad (5a)$$
$$B^O = f \cdot a \cdot Q_0 - f^2 \cdot \tfrac{1}{2} \cdot b \cdot {Q_0}^2 \quad (5b)$$

---

[7] De elasticiteit van de vraag geeft de procentuele verandering van de gevraagde hoeveelheid na een 1-procents verandering van de prijs, is daarmee een maatstaf voor de prijsgevoeligheid, en heeft onder meer als voordeel dat deze geen eenheden kent en onafhankelijk is van de schaal van de markt en de gekozen eenheid van prijs. Dat is niet het geval voor de helling van de inverse vraagfunctie.

$$B^R = f \cdot a \cdot Q_0 - f \cdot \frac{1}{2} \cdot b \cdot Q_0{}^2 \quad (5c)$$

We vinden de ratio van baten door (3a) en (3b) in te vullen in (5b) en (5c), de teller en noemer vervolgens te delen door $P_0 \cdot Q_0 \cdot f$, hetgeen uiteindelijk leidt tot:

$$\rho \equiv \frac{B^R}{B^O} = \frac{-\varepsilon + \frac{1}{2}}{-\varepsilon + \left(1 - \frac{f}{2}\right)} \quad (6)$$

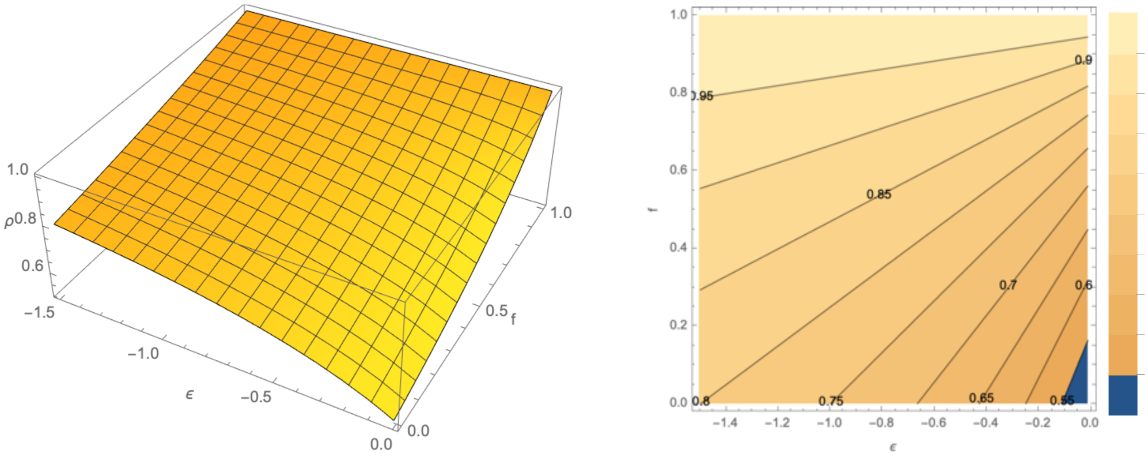
Merk allereerst op dat ρ zodanig is gedefinieerd dat deze altijd tussen 0 en 1 valt. Als random rantsoenering net zo goed is als optimale rantsoenering is de waarde 1; hoe groter echter de welvaartsverliezen, hoe dichter bij 0 ρ uitkomt. Ten tweede laat vergelijking (6) inderdaad zien dat de ratio van baten ρ niet afhangt van de initiële prijs en hoeveelheid. Dat is ook plausibel; immers, als we de eenheden van prijs en hoeveelheid zouden veranderen (van Euro's naar centen; van kilo's naar tonnen) of de tijdsschaal zouden veranderen (een maand in plaats van een week beschouwen) zou de ratio niet mogen veranderen.

De vraagelasticiteit doet er wel degelijk toe. Deze is op zichzelf negatief zodat $-\varepsilon$ positief is, en wanneer de vraag volkomen ("perfect") elastisch wordt, zodat de inverse vraagfunctie in Figuren 1 en 2 horizontaal wordt, gaat $-\varepsilon$ naar oneindig en ρ daarmee naar 1. Random rantsoenering wordt dan even efficiënt als optimale rantsoenering. Een moment van reflectie laat zien waarom. Bij perfect elastische vraag zijn de marginale baten constant. Elke eenheid die geconsumeerd wordt vertegenwoordigt evenveel baten, en het loten om consumptie is even efficiënt als het rantsoeneren via prijzen. Hoe minder elastisch de vraag, hoe groter de verschillen in baten; in Figuur 1 en 2 kan dit gevisualiseerd worden door in het oorspronkelijke evenwicht de helling van de inverse vraagfunctie steiler te maken. Het wordt dan steeds belangrijker om de juiste eenheden te selecteren bij rantsoenering, en dat vertaalt zich in een lagere waarde van ρ. In de noemer van (6) is te zien dat dit effect sterker wordt naarmate rantsoenering strikter is en f dus kleiner. Ook hier is de intuïtie aansprekend. Wanneer f richting 0 gaat en de rantsoenering dus heel strikt is, wordt de gemiddelde waarde van de baten van het optimale geval overblijvende eenheden steeds groter, terwijl de gemiddelde waarde bij random rantsoenering niet verandert (Figuur 2 kan helpen bij het doorzien van die intuïtie). De verhouding van overgebleven baten ρ wordt daarmee steeds kleiner, en rantsoenering op basis van baten in relatieve zin steeds belangrijker. De laagst mogelijke waarde van ρ bij een lineaire vraagfunctie bedraagt ½; dat ρ niet naar nul kan gaan heeft te maken met de geometrie bij een lineaire vraagfunctie, waardoor de baten van de meest gewaardeerde eenheid een eindige waarde kennen (gegeven door de parameter a in de inverse vraagfunctie (1)).

Het grote voordeel van een relatieve uitdrukking als in vergelijking (6) is dat het ons vrij makkelijk in staat stelt om een indruk te geven van het relatieve belang van het gebruiken van rantsoeneringsinstrumenten die op selectie naar baten zijn gebaseerd. Als we een inschatting hebben van de vraagelasticiteit ε en weten hoe strikt de rantsoenering f is, hebben we op basis van vergelijking (6) al een lineaire benadering van dat relatieve belang. Hierna zullen we nog zien dat we een vergelijkbaar flexibele uitdrukking krijgen voor het relatieve belang ρ bij een niet-lineaire, iso-elastische vraagfunctie. Wanneer we dus wel informatie hebben over de initiële vraagelasticiteit maar geen informatie over de vorm van de vraagfunctie, kunnen we toch een indruk voor ranges krijgen door naar beiden te kijken. En, we weten dat dat relatieve belang toeneemt naarmate de vraag minder elastisch wordt, en naarmate de rantsoenering strikter. Figuur 3 laat dit zien door op twee manieren dezelfde informatie te plotten: hoe ρ volgens vergelijking (6) afhangt van combinaties van f en ε (in het initiële evenwicht).

Figuur 3 bevestigt het kwalitatieve beeld zoals hierboven besproken, en helpt daarnaast om getallen te plakken op concrete casussen. Nemen we openbaar vervoer per trein als voorbeeld. Volgens een recente studie van het KIM (2018) is de vraagelasticiteit in het personenverkeer per trein in de orde van grootte van -0.45. Bij een restrictie naar 40% van de oorspronkelijke vraag zou volgens vergelijking (6), en afleesbaar uit Figuur 3, de waarde van ρ gelijk zijn aan 0.76. Dus, 76%

van de baten die met optimale rantsoenering wordt overgehouden, blijft behouden met random rantsoenering; oftewel, ongeveer een kwart gaat verloren als niet optimaal gerantsoeneerd wordt. Dat is een fors relatief verlies, waarvan we hieronder zullen zien dat het ook nog weer groter wordt als we uitgaan van een niet-lineaire, iso-elastische vraagfunctie.



Figuur 3: 3D-plot (links) en contour plot (rechts) van het verband tussen ε, f en ρ voor lineaire vraag

De absolute omvang van het verschil in baten kunnen we alleen bepalen als we informatie over initiële prijzen en hoeveelheden gebruiken. Binnen het bestek van deze studie was het niet mogelijk om dit voor feitelijke markten te onderzoeken, maar een schatting van de orde van grootte kan illustratief zijn. We nemen wederom het personenverkeer per trein. In 2019 was het aantal reizigerskilometers rond de 20 miljard (KIM, 2020). Uiteraard verschillen de prijzen die voor die kilometers gevraagd worden, maar laten we eens zien waar we op uitkomen als voor elk van die reizigerskilometers, vanwege het bestaan van abonnementen en kortingskaarten, gemiddeld de helft van het reguliere volle tarief van 17ct per kilometer gevraagd zou zijn; dus 8.5 ct per reizigerskilometer. Als we dan welvaartseffecten uitrekenen alsof de vraagfunctie, met de genoemde elasticiteit in het huidige evenwicht, lineair zou zijn, dan zouden de totale jaarlijkse baten voor reizigers in het oorspronkelijke evenwicht 3.59 miljard Euro bedragen, bij optimale rantsoenering 1.89 miljard, en bij random rantsoenering 1.44 miljard. Het verschil bedraagt 0.45 miljard Euro; een hoog bedrag, ook als we het afzetten tegen wat de opbrengsten zouden zijn bij een tarief van 8.5 ct per kilometer en een omvang van 0.4 x 20 miljard reizigerskilometers. Het verschil in baten is namelijk 2/3 van die tariefopbrengsten (die ratio, 2/3, blijft overigens hetzelfde als we uit zouden gaan van een tarief van 17 ct per reizigerskilometer voor dezelfde omvang van treinmobiliteit). We haasten ons te benadrukken dat dit echt alleen maar een indicatief sigarendoosje-berekening is, om een gevoel voor de orde van grootte te krijgen, maar die geeft wel degelijk aan dat het van belang is om goed na te denken over de vormgeving van rantsoeneringsregels, ook al bij de relatief conservatieve aanname van een lineaire vraagfunctie.

3.2 Iso-elastische vraag

De lineaire vraag die we hierboven beschouwen is maar één van verschillende mogelijke vormen die (inverse) vraagfuncties kunnen aannemen. Een andere veelgebruikte vorm is de iso-elastische vraagfunctie, waarvoor – zoals de naam al aangeeft – de aanname is dat elasticiteit constant is langs de gehele vraagfunctie. De vraagfunctie die daar bij hoort kent de volgende functionele vorm:

$$Q = A \cdot P^{\varepsilon} \quad (7)$$

waarbij A een parameter is die de schaal van de markt weergeeft en ε weer de elasticiteit is (merk op dat ε dus negatief is). Deze vertaalt zich in de volgende inverse vraagfunctie:

$$P = a \cdot Q^{e} \quad (8)$$

met:

$$a = A^{\frac{1}{-\varepsilon}} \quad (9a)$$

en:

$$e = \frac{1}{\varepsilon} \quad (9b)$$

De totale baten zijn weer te bepalen als de integraal van de inverse vraagfunctie:

$$B = \frac{1}{e+1} \cdot a \cdot Q^{e+1} + C \quad (10)$$

waarbij C een constante van integratie is die voor een elastische vraag met $\varepsilon < -1$ gelijk is aan nul, en in andere gevallen naar oneindig gaat, maar zolang we kijken naar verschillen in totale baten tussen marktevenwichten wegvalt. Als we voor elke relevante batenterm het verschil nemen in baten bij die uitkomst versus een heel kleine, maar toch nog positieve consumptie hoeveelheid, en de limiet nemen van de situatie waarin die kleine hoeveelheid oneindig dicht naar nul gaat, kunnen we dit probleem omzeilen. We schrijven die kleine hoeveelheid als een fractie φ van Q_0, en laten dus φ oneindig dicht naar 0 gaan.

Dit betekent dat we voor deze inverse vraagfunctie de baten bij, respectievelijk, het oorspronkelijke evenwicht (B_0), onder optimale rantsoenering (B^O) en onder random rantsoenering (B^R) als volgt kunnen schrijven:

$$B_0 = \lim_{\varphi \to 0} \frac{1}{e+1} \cdot a \cdot (1-\varphi) \cdot {Q_0}^{e+1} \quad (11a)$$

$$B^O = \lim_{\varphi \to 0} f^{e+1} \cdot \frac{1}{e+1} \cdot a \cdot \left(1 - \frac{\varphi}{f^{e+1}}\right) \cdot {Q_0}^{e+1} \quad (11b)$$

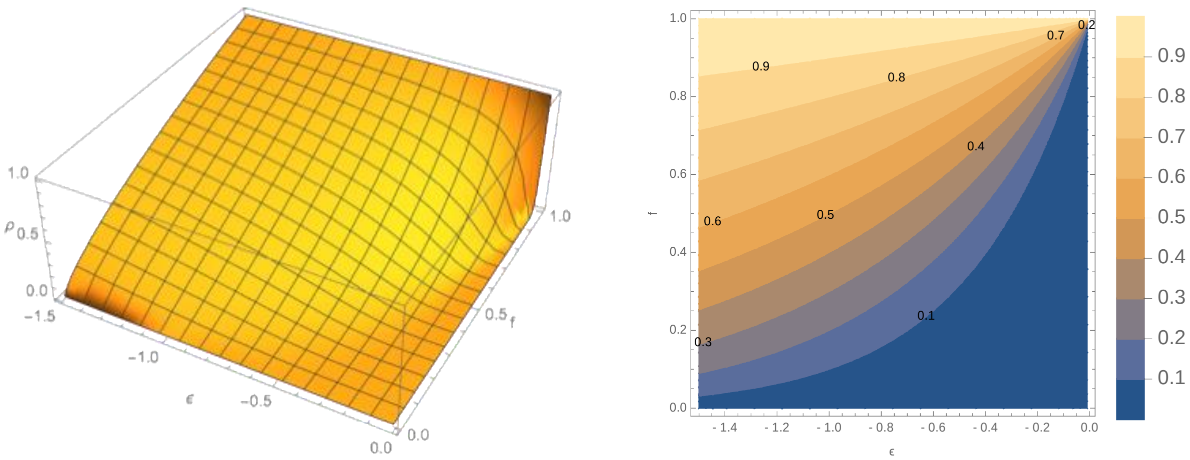
$$B^R = \lim_{\varphi \to 0} f \cdot \frac{1}{e+1} \cdot a \cdot (1-\varphi) \cdot {Q_0}^{e+1} \quad (11c)$$

We vinden de ratio van baten door (11c) te delen door (11b), de teller en noemer vervolgens te vereenvoudigen, hetgeen uiteindelijk leidt tot:

$$\rho \equiv \frac{B^R}{B^O} = \lim_{\varphi \to 0} \frac{f - f \cdot \varphi}{f^{e+1} - \varphi} = \frac{1}{f^{\frac{1}{\varepsilon}}} = f^{-\frac{1}{\varepsilon}} \quad (12)$$

Al met al een enigszins complexe afleiding wellicht, maar met uiteindelijk een elegant eindresultaat, dat ook weer een heldere interpretatie kent. Net als voor de lineaire vraagfunctie is de ratio van baten ρ hoger als de vraag elastischer wordt: als ε naar min oneindig gaat, gaat de macht naar 0, en ρ dus naar 1. Als ε daarentegen naar –0 gaat, richting perfect inelastische vraag, gaat de macht naar oneindig, en ρ dus naar 0 (bedenk dat f tussen 0 en 1 ligt). En, naarmate f dichter bij 1 ligt, is ρ hoger voor een gegeven ε. Dit zijn dezelfde kwalitatieve patronen als wat we vonden voor lineaire vraag.

Het is om die reden niet verbazingwekkend dat de kwalitatieve patronen die we vinden als we in Figuur 4 ρ plotten als functie van ε. en f, vergelijkbaar zijn met wat we in Figuur 3 zagen voor lineaire vraag. Het belangrijkste verschil is dat voor gegeven combinaties van ε. en f, we bij iso-elastische vraag lagere waarden voor ρ vinden. De intuïtie daarachter is dat, vanwege de convexe vorm van de inverse vraagfunctie, de baten van de eenheden die het meest links op de functie liggen, relatief zeer hoog zijn. Het verlies van dergelijke eenheden bij random rantsoenering brengt dus relatief grote welvaartsverliezen met zich mee, waardoor ρ relatief laag wordt. Bijgevolg is ρ voor de eerder genoemde combinatie van ε=-0.45 en f=0.4 bij iso-elastische vraag gelijk aan 0.13, waar het voor de lineaire vraag nog 0.76 was. Het is voor het bepalen van het maatschappelijk belang van optimale rantsoenering versus random rantsoenering dus niet alleen van belang om de vraagelasticiteit te kennen, maar ook de vorm van de vraagfunctie. Zonder kennis daarvan zou het gebruik van een lineaire vraagfunctie een meer conservatieve schatting van het relatieve belang van optimale in plaats van random rantsoenering geven dan een iso-elastische.



Figuur 4: 3D-plot (links) en contour plot (rechts) van het verband tussen ε, f en ρ voor iso-elastische vraag

4. Verschillende rantsoeneringsmethoden: een kwalitatieve verkenning

Gewapend met de kennis uit paragraaf 3 zullen we nu in deze paragraaf een aantal mogelijke instrumenten voor de rantsoenering van de vraag bespreken, en daarbij een kwalitatieve inschatting geven van de drie eerdergenoemde criteria:

1. Efficiëntie: in welke mate wordt de verlaagde capaciteitsgrens tegen zo laag mogelijke netto maatschappelijke kosten bereikt?
2. Effectiviteit: in welke mate wordt het gestelde doel überhaupt bereikt?
3. Acceptatie: in welke mate heeft het instrument draagvlak bij bevolking en/of beleidsmakers?

4.1 Laissez-passer

Een eerste manier om met nieuwe schaarste om te gaan is wat we "laissez passer" noemen: in feite niets anders doen dan te communiceren wat de feitelijke capaciteit is, en middels publieke berichtgeving een beroep op zelf-beperking te doen om bepaalde plekken op drukke tijden en plaatsen te voorkomen.

Efficiëntie De efficiëntie van deze benadering ligt waarschijnlijk tussen die van random rantsoenering en die van optimale rantsoenering. Aan de ene kant valt te verwachten dat mensen met uitzonderlijk hoge baten altijd wel tot de uiteindelijke consumenten zullen behoren. Tegelijkertijd zal het gehoor geven aan dergelijke oproepen ook afhangen van factoren die niet gecorreleerd zijn met deze baten, zoals volgzaamheid, sociaal inlevingsvermogen, bescheidenheid; maar ook: angst voor eigen besmettingsrisico. Dit heeft als gevolg dat de resulterende rantsoenering niet optimaal zal zijn. Bovendien kunnen op de langere termijn met name de eersten hiervan, door irritatie over het onbescheiden gedrag van anderen en gevoelens van oneerlijkheid, ondermijnd worden, waardoor de effectiviteit afneemt. Er is anekdotische informatie dat voor en in de tweede golf mensen op publieke plaatsen onvoorzichtiger waren dan ten tijde van de eerste golf, hetgeen de mogelijke relevantie lijkt te ondersteunen.

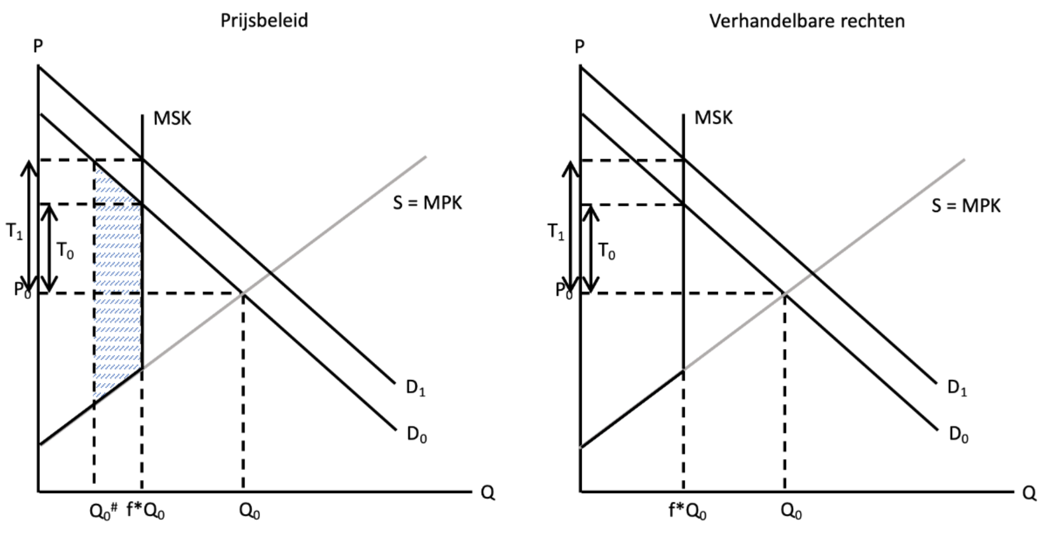
Effectiviteit De effectiviteit van deze benadering is niet gegarandeerd en er zijn voorbeelden waar het misging; bijvoorbeeld bepaalde natuurgebieden of winkelstraten. Zeker als besmettingsrisico's snel oplopen naarmate de normen in hogere mate worden overschreden, is dit een punt van zorg.

Acceptatie De acceptatie van deze benadering lijkt redelijk hoog, zeker in het begin van de pandemie. Er moet rekening mee worden gehouden dat de acceptatie op langere termijn zou kunnen afnemen wanneer er sterke gevoelens van oneerlijkheid staan, gevoed door de gedachten dat een bepaald deel van de mensen steeds maar weer plaats maakt voor anderen die zich minder van de gezamenlijke noodzaak aan lijken te trekken.

4.2 Prijsmechanisme

Een tweede manier om met plotselinge schaarste om te gaan is het inzetten van het prijsmechanisme; dus, prijzen zodanig aanpassen dat aan de nieuwe capaciteitslimieten voldaan zal worden.

Efficiëntie De efficiëntie van deze benadering zal hoog liggen; immers, zoals in het voorgaande is uiteengezet garandeert deze benadering dat alleen de eenheden die de hoogste baten vertegenwoordigen geconsumeerd zullen worden. Dat geldt in de eerste plaats voor de consumptiebundel van een individu: bij random toewijzing zal het vaak kunnen gebeuren dat iemand toegang tot bijvoorbeeld het OV krijgt op een dag waar dat minder nut oplevert, en niet op wat juist een belangrijke dag is. Het geldt ook voor consumptiepatronen over individuen: degene die redelijk makkelijk met de fiets gaat zal de tram laten gaan als deze te duur is; degene voor wie de tram heel belangrijk is blijft. Bij acceptatie hieronder zullen we bespreken dat dit alles uiteraard ook verdelingsconsequenties heeft die negatief beoordeeld zullen kunnen worden.



Figuur 5: Prijsmechanisme (links) versus verhandelbare rechten (rechts) bij fluctuerende vraag en starre instrumenten

Het is met opzet dat we niet zeggen dat het prijsmechanisme maximaal efficiënt zal zijn. In de praktijk zullen theoretisch optimale prijsinstrumenten moeilijk of onmogelijk te implementeren zijn, en zullen in plaats daarvan "second-best", niet optimale, prijsinstrumenten gebruikt worden. Het zal bijvoorbeeld onwaarschijnlijk zijn dat een regulerende prijs iedere dag kan worden aangepast op basis van de feitelijke drukte. Het linker paneel in Figuur 5 toont deze situatie. Stel dat de vraag D_0 uit de eerdere voorbeelden niet iedere dag hetzelfde is, maar dat er dagen met extra grote vraag kunnen zijn, weergegeven door D_1. De maximaal toegestane bezetting blijft echter hetzelfde, $f{\cdot}Q_0$. En, stel dat die doelstelling inderdaad optimaal gekozen is, zodat de marginale sociale kosten MSK zodanig snel stijgen bij $f{\cdot}Q_0$ dat de functie daar, zoals getekend, in feite verticaal loopt (MPK staat nu voor marginale private kosten, zoals ondervonden door de aanbieder). Als het instrument star is in de zin dat het niet over de dagen kan worden aangepast, zal – om te voorkomen dat op drukke dagen te veel consumptie plaatsvindt – de marktprijs P_0 moeten worden verhoogd met een toeslag T_1; hoger dan de toeslag T_0 die op de andere dagen voldoende zou zijn geweest. Het gevolg is dat op dagen waar D_0 geldt, de evenwichtshoeveelheid $Q_0^{\#}$ lager zal zijn dan het optimum, met op die dagen een welvaartsverlies ter grootte van het gearceerde gebied als gevolg.

Het voorbeeld laat zien dat, hoewel het gebruik van prijsprikkels een noodzakelijke voorwaarde is voor het bereiken van de meest efficiënte uitkomst, het nog geen voldoende voorwaarde is.

Effectiviteit De effectiviteit van prijsmaatregelen is in principe hoog: er is altijd een prijs die hoog genoeg is om de vraag ver genoeg terug te dringen. Toch zijn er kanttekeningen bij de effectiviteit te maken. De eerste heeft te maken met de analyse in Figuur 5: als de prijs het instrument is, en de hoeveelheid zich daarop aanpast, ontstaat er onzekerheid over de werkelijke effectiviteit wanneer de vraagfunctie niet met zekerheid gekend is. Indien, in dat eerdere voorbeeld, een nog grotere schok optreedt dan de verschuiving naar D_1, wordt de kritische bezettingsgrens alsnog gepasseerd. De tweede kanttekening is dat de vereiste hoogte niet acceptabel geacht kan worden, door de bevolking en/of door de beleidsmakers. De potentiele effectiviteit kan dan niet volledig benut worden.

Acceptatie De acceptatie van prijsmaatregelen om (spits)drukte te beperken is in reguliere omstandigheden beperkt. Pogingen om congestieheffingen op de weg of in de trein in te voeren stuiten doorgaans op te grote weerstand om uiteindelijk ingevoerd te kunnen worden. In het geval van Covid maatregelen zou de acceptatie hoger kunnen zijn wanneer mensen van mening zijn dat invoering echt noodzakelijk is voor volksgezondheid, en tijdelijk zou zijn - voor de duur van de pandemie. Waarschijnlijker is echter dat men ook dan een voorkeur zou hebben voor anderssoortige instrumenten. Waarschijnlijke argumenten zullen zijn dat (1) met prijsbeleid schaarse capaciteit voornamelijk bij de hoogste inkomensgroepen terecht zal komen, hetgeen oneerlijk gevonden kan worden; (2) de financiële zorgen om de pandemie al groot genoeg zijn, steunpakketten meer voor de hand liggen, en prijsverhogingen dus onwelkom zijn; en (3) het zeer onsympathiek overkomt wanneer aanbieders "profiteren" van een pandemie door prijzen te verhogen. Deze argumenten zouden geadresseerd kunnen worden door de verhoogde opbrengsten op slimme manieren terug te sluizen naar de consumenten, maar het lijkt erg optimistisch om te denken dat daarmee de acceptatie op voldoende peil gebracht zou kunnen worden.

4.3 Verhandelbare of verruilbare rechten

Een ander mogelijk instrument is het gebruik van verhandelbare rechten. Dit type instrument wordt al langer gebruikt in toepassingen rond milieuvervuiling, met als in Europa als waarschijnlijk bekendste toepassing de verhandelbare CO_2-rechten. Terwijl dit een toepassing is waarbij bedrijven handelen in rechten, is meer recentelijk ook experimenteel onderzoek gedaan naar systemen waarbij individuen in rechten handelen, met name binnen het mobiliteitsdomein. Het algemene idee achter verhandelbare rechten is dat een beperkt aantal rechten op de markt wordt gebracht, corresponderend met de doelstelling voor maximale consumptie; in het eerdere voorbeeld zou dat $f{\cdot}Q_0$ zijn. Door handel komen de rechten uiteindelijk terecht bij, en worden dus gebruikt door, degenen die de hoogste betalingsbereidheid kennen, en worden ingezet voor consumptie van de meest links gelegen eenheden langs de inverse vraagfunctie. De marktprijs voor de rechten wordt gelijk aan de toeslag T die bij dezelfde optimale hoeveelheid hoort, en in theorie is de efficiëntie daarmee maximaal, net als voor prijsbeleid. Wanneer de rechten van tevoren om niet worden uitgedeeld ("grandfathering") is de uitwerking voor de markt als geheel budget neutraal, terwijl bij prijsbeleid voor iedereen de prijs stijgt: degene die minder rechten gebruiken dan ze ontvangen kunnen een deel verkopen en gaan er financieel op vooruit; degene die meer rechten gebruiken dan ze ontvangen moeten bijkopen en gaan er financieel op achteruit, maar natuurlijk in mindere mate dan bij prijsbeleid wanneer ze een aantal rechten al voor niets hebben ontvangen. De verwachting is dat de acceptatie daarom over het algemeen hoger zal zijn.

Efficiëntie De efficiëntie van het instrument, dat in theorie in gedragssturing dezelfde effecten zou hebben als prijsbeleid, is ook weer hoog. In theorie zou de efficiëntie net als voor prijzen maximaal zijn, maar in de praktijk vereist een hoge efficiëntie dat transactiekosten voor het handelen in rechten laag gehouden kunnen worden, en zal de efficiëntie net als bij prijsbeleid lager zijn bij second-best oplossingen. Interessant genoeg kunnen effecten daarvan op prijsbeleid anders uitpakken dan bij verhandelbare rechten. Voor het eerder genoemde voorbeeld waarbij het instrument star is terwijl de vraag varieert over de dagen zouden verhandelbare rechten efficiënter kunnen uitpakken dan prijzen, met name wanneer het aantal rechten per dag hetzelfde is omdat de toegestane capaciteit hetzelfde is. Dit staat weergegeven in het rechter diagram in Figuur 5. De prijs van rechten zou op de dag waarop D_1 geldt uit zichzelf stijgen naar T_1 en het welvaartsverlies dat in de linker diagram voor prijsbeleid wordt getoond, zou op de dagen dat D_0 geldt niet optreden omdat dan de prijs vanzelf op T_0 uitkomt.

Een mogelijke achilleshiel zouden de transactiekosten kunnen zijn. Via online handel zouden deze beperkt kunnen blijven – zo hebben eerdere experimenten ook laten zien – maar hier ligt in ieder geval een belangrijk aandachtspunt.

Effectiviteit De effectiviteit van verhandelbare rechten is in principe hoog, ervan uitgaande dat ontduiking niet mogelijk is (maar dat geldt natuurlijk voor alle maatregelen). Doordat het instrument direct aangerijpt bij de hoeveelheid is de effectiviteit ook directer te sturen dan bij prijsbeleid. Waar bij prijzen de reguleerder de prijs of toeslag kiest en marktmachten bepalen welke hoeveelheid daar bij zal horen, wordt bij verhandelbare juist de hoeveelheid gekozen en bepalen marktmachten vervolgens de prijs. Dat lijkt in de context van snel oplopende besmettingsrisico's bij hogere bezetting, zoals in Figuur 5 geschetst met de verticaal stijgende MSK, te prefereren.

Acceptatie Verwacht mag worden dat de acceptatie van verhandelbare rechten ten opzichte van die van prijsbeleid positief beïnvloed zal worden doordat de drie eerder bij prijsbeleid genoemde bezwaren verzacht of zelfs weggenomen worden. (1) Lage inkomens kunnen de door hen ontvangen rechten gebruiken voor consumptie en hoeven tot aan de ontvangen rechten dus niet op te bieden tegen de betalingsbereidheid van hogere inkomensgroepen. (2) Doordat het instrument op geaggregeerd niveau budget-neutraal is, is het effect op bestedingsruimte voor hen die per saldo rechten verkopen zelfs positief, en voor hen die bijkopen weliswaar negatief maar minder dan bij prijzen. (3) De aanbieder(s) profiteren niet via hogere prijzen van extra schaarste door lock-down maatregelen: de schaarste wordt onder consumenten zelf verhandeld.

Een nog lopend NWO onderzoeksproject, U-Smile, heeft de acceptatie van verhandelbare rechten in de context van mobiliteitstoepassingen onderzocht (zie o.m. Brands et al, 2021). Hieruit kwam onder meer naar voren dat de acceptatie afhangt van de gepercipieerde eerlijkheid van de verdeling van de rechten, en hier zal een politieke keuze liggen. Verder kwam naar voren dat deelnemers aan experimenten verhandelbare rechten eenvoudig in gebruik vonden, aantrekkelijker dan prijsbeleid, en dat speculatiegedrag en fraude goed te voorkomen zijn; terwijl respondenten zonder eigen ervaring in experimenten daar juist meer zorgen over hebben. Goede voorlichting en demonstratieprojecten zouden daarbij kunnen helpen.

Het kan aantrekkelijk zijn om eenvoudiger varianten dan volledig verhandelbare rechten te overwegen, zeker in een beginfase. Te denken valt aan de volgende mogelijkheden:

1. Ruil in plaats van handel. Wanneer het implementeren van handel als te complex wordt gezien of toch zorgen om verdelingseffecten blijven bestaan zou een variant kunnen zijn dat in elk geval rechten geruild kunnen worden. Een voorbeeld is wanneer rechten aan een specifieke weekdag gekoppeld zijn (denk aan treingebruik in woon-werkverkeer) en deelnemers aan het systeem minder dan 5 rechten per week ontvangen. Wanneer voor individuen de baten sterk over de dagen verschillen maar niet voor iedereen op dezelfde wijze, en de capaciteitsgrenzen per dag bewaakt moeten worden (een legere bezetting op woensdag compenseert niet voor de overbezetting op dinsdag) zodat per dag een maximum aantal rechten beschikbaar wordt gemaakt, kunnen deze verruilbaar gemaakt worden zodat iedereen zoveel mogelijk op de meest gewenste dagen kan reizen.
2. Rechten eerst zonder ruil. Om noodzaak voor mogelijkheid tot ruil en of handel te verkennen zou in eerste instantie ook gewerkt kunnen worden met toegekende rechten (al dan niet na loting na opgave van gewenste reserveringen) die op persoon staan en blijven. Evaluatie enquêtes na enige gewenningstijd kunnen dan aangeven of daarbij de behoefte wordt gevoeld om te kunnen ruilen en of handelen. Bij voldoende interesse kan dit mogelijk gemaakt worden, waarbij men er altijd voor kan blijven kiezen om eenvoudigweg geen gebruik te maken van ruil of handel.

4.4 Reserveringen: wie het eerst komt, wie het eerst maalt

Bij reserveringen voor schaarse capaciteit geldt in feite als rantsoeneringsprincipe: wie het eerst komt, wie het eerst maalt. In de praktijk wordt dit mechanisme vaak toegepast; eigenlijk bij alle producten en diensten waarvoor geldt dat reserveringen gemaakt kunnen worden en het risico bestaat dat men bij te laat reserveren achter het net vist.

Efficiëntie Doordat bij reserveringen niet direct gedifferentieerd wordt op basis van feitelijke betalingsbereidheid is de efficiëntie naar verwachting lager dan bij beprijzen of verhandelbare rechten. De efficiëntie zal wel hoger liggen dan bij random rantsoenering wanneer mensen met een hogere betalingsbereidheid een sterkere prikkel voelen om (eerder) te reserveren. In welke mate dat het geval is zal van situatie tot situatie verschillen. Daarnaast kan het strategisch reserveren (meer dan nodig en/of gepland, om alle opties open te houden), wanneer dat mogelijk is, tot verdere inefficiënties leiden.

Effectiviteit De effectiviteit van reserveringen is in principe net zo hoog als van verhandelbare rechten. Als er niet gesjoemeld kan worden zullen er nooit meer consumenten zijn dan de ruimte die van tevoren beschikbaar is gesteld.

Acceptatie De acceptatie zal relatief hoog zijn vanwege bekendheid met het principe, en de relatieve eenvoud. De acceptatie kan bij langer gebruik ondermijnd worden als strategisch en oneigenlijk gedrag van anderen een probleem gaan vormen, en als het wie-het-eerst-komt-principe tot problemen leidt wanneer voor sommige belangrijke deelgroepen pas laat de informatie beschikbaar komt die bepaalt voor welke dag ze zouden willen reserveren.

4.5 "Alles of niets" over groepen

Een laatste rantsoeneringsprincipe is om toegang alleen voor bepaalde groepen te bieden, mogelijk in bepaalde tijdvakken.

Efficiëntie De efficiëntie van deze benadering is naar verwachting lager dan bij optimale rantsoenering, maar hoger dan bij random rantsoenering. Deze rantsoenering houdt er geen rekening mee dat niet iedereen in de doelgroepen voor alle eenheden hoge baten kennen, en dat niet iedereen die buiten de doelgroepen valt lage baten zou kennen. De efficiëntie wordt hoger naarmate de baten sterker tussen groepen verschillen, en tegelijkertijd minder binnen groepen.

Daarnaast wordt de efficiëntie hoger als de tijdsdimensie slim benut wordt en tijdssubstitutie relatief gemakkelijk is; bijvoorbeeld, als bepaalde groepen exclusieve toegang op bepaalde uren hebben, het voor anderen relatief gemakkelijk is de uren aan te passen, en er baten zijn van het scheiden van groepen in de tijd. Speciale uren voor ouderen in supermarkten zou zo'n voorbeeld kunnen zijn.
Effectiviteit Bij een juiste keuze van de doelgroepen en, uiteraard, een goede controle, zou de effectiviteit van deze benadering net zo hoog als van optimale rantsoenering moeten zijn. Het is zelfs denkbaar dat de effectiviteit te hoog is, hetgeen weer ten koste van de efficiëntie zou gaan; met name wanneer om zeker te zijn van het voldoen aan de grenswaarde doelgroepen zodanig strikt gedefinieerd worden dat minder mensen dat optimaal gebruik maken van de mogelijkheid.

Acceptatie Het is lastig om de acceptatie van tevoren in te schatten; veel zal afhangen van hoe de doelgroepen gekozen worden, en of er buiten die doelgroepen veel potentiële consumenten zouden zijn dei ook hoge baten of belang bij gebruik zouden hebben. Het zou goed zo kunnen zijn dat het principe op zichzelf een hoge acceptatie kent, maar de concrete uitwerking op grote acceptatieproblemen stuit. Eerdere discussies over welke beroepen essentieel zijn, welke topsporten doorgang moeten vinden, welke grote bijeenkomsten doorgang moeten vinden (kerkdiensten), geven aan dat acceptatie van de concrete invulling van op zichzelf acceptabel geachte instrumenten toch problematisch kan zijn.

5. Conclusie

De belangrijkste conclusies laten zich als volgt samenvatten:

Rantsoenering is een belangrijk vraagstuk en "er valt iets te kiezen": verschillende methoden kunnen behoorlijk verschillen op criteria van efficiëntie, effectiviteit, acceptatie.

Efficiëntie van rantsoenering betreft de welvaartsverliezen die met niet optimale rantsoenering gepaard gaan. Deze welvaartsverliezen worden groter naarmate de vraag over de relevante range inelastischer is, en naarmate de rantsoenering strikter is.

Het kan gaan over "big potatoes": bijvoorbeeld rantsoenering in het nationale spoorvervoer kan een additioneel welvaartsverlies kennen als het random plaatsvindt ten opzichte van optimaal in de orde van grootte van 0.5 miljard per jaar, zoals het voorbeeld met lineaire vraag liet zien.

Zeker als rantsoenering langer gaat duren en mogelijk terugkerend is bij eventuele nieuwe pandemieën of uitbraken wordt het belangrijk om te zoeken naar rantsoeneringsmethoden die niet alleen gegarandeerd effectief zijn, maar ook de welvaartsverliezen beperken en toch voldoende acceptabel zijn. Verhandelbare of verruilbare rechten zouden daar een voorbeeld voor kunnen zijn.

Referenties

Atkinson, A.B. & J.E. Stiglitz (2015). *Lectures on Public Economics* (updated edition). Princeton University Press.

Brands, D., E.T. Verhoef & J.S.A. Knockaert (2021). Pcoins for parking: a field experiment with tradable mobility permits. Tinbergen Institute Discussion Paper 21-029-viii, Tinbergen Institute Amsterdam/Rotterdam.

KIM (2020). *Mobiliteitbeeld 2019*. Kennisinstituut voor Mobiliteitsbeleid, Den Haag.

Mas-Colell A., M.D. Whinston & J.R. Green (1995). *Microeconomic Theory*. New York Oxford University Press.

Varian, H.R. (2010). *Intermediate Microeconomics: A Modern Approach* (8th ed.). W. W. Norton & Company, New York.