



DURF TE METEN

Eindrapport Expertwerkgroep Effectmeting

november 2012

In Memoriam Jules Theeuwes 1944-2012

Jules Theeuwes heeft als voorzitter van de werkgroep een belangrijke rol gespeeld bij de totstandkoming van dit rapport. Helaas heeft Jules de definitieve afronding van dit rapport niet meer mogen meemaken. Ondanks zijn ziekte heeft Jules nog tot op het laatste moment zijn betrokkenheid bij dit rapport laten merken. We hopen daarom dat Jules zijn gedachtegoed doorklinkt in de uiteindelijke versie van het rapport.

Samenstelling werkgroep

Wolter Hassink (UU)
Bas van der Klaauw (VU)
Marco van Maasacker (Algemene Rekenkamer)¹
Wim Schaasberg (CBS)
Bas Straathof (CPB)
Jules Theeuwes (SEO, voorzitter) †
Thomas Dirkmaat (Ministerie EZ; AEP, secretaris)
Tim Gelissen (Ministerie EZ; AEP, secretaris)
Jeroen Heijs (Ministerie EZ; AgNL)
Luuk Klomp (Ministerie EZ, DG B&I)

De werkgroep wil de volgende mensen bedanken voor hun bereidheid om hun kennis en ervaringen met de werkgroep te delen:

Godelieve van der Beken, René den Boer, Pieter de Bruijn, Jacqueline Coenraads, Jan Dixel, Jeroen Doorman, Anja Hezemans, Jan Hendrik Hoekstra, Edi Huster, Martijn Janmaat, Roe van Kasteren, Sander Kes, Mineke de Lange, Pieter Moerman, Aisia Okma, Marja Peperkamp, Theo Roelandt, Karen de Ruijter, Robert Schaart, Paul Silvertant, Gideon van der Staaïj, Koos Straver, Marjolijn van Valkenhoef, Suzanne Verboon, Dmitri Verhoeven en Upton van der Vliet.

Een speciaal woord van dank aan Piet Donselaar voor zowel zijn inhoudelijke bijdrage als zijn redactionele werk.

Ten slotte ook een woord van dank aan de volgende personen die hun expertise inbrachten als waarnemers voor enkele leden van de werkgroep:

Peter Kempkes (Algemene Rekenkamer), Ruurd Schoonhoven (CBS), Geert Thijssen (AgNL) en Henry van der Wiel (CPB).

¹ De vertegenwoordiger van de Algemene Rekenkamer is toevoerder bij de werkgroep. Hij geeft reflectie op de plannen van de werkgroep, gebaseerd op de (uitwerking van de) aanbevelingen die de Algemene Rekenkamer in haar onderzoek Innovatiebeleid heeft gedaan. Het betreft een ambtelijke vertegenwoordiging van de Rekenkamer. Het College van de Algemene Rekenkamer is niet gebonden aan deze vertegenwoordiging en kan bij toekomstige oordeelsvorming anders oordelen.

Managementsamenvatting

Met het bedrijvenbeleid probeert de overheid het ondernemersklimaat in Nederland te verbeteren. Dit beleid wordt gelegitimeerd door de aanwezigheid van marktfalen, zoals externe effecten en informatie- en coördinatiegebreken. Het is daarbij belangrijk dat de kosten van overheidsingrijpen niet groter zijn dan de kosten van het marktfalen. De overheid heeft een vrij omvangrijk budget beschikbaar voor verschillende beleidsmaatregelen. Bovendien zijn er veel bedrijven die in aanraking komen met het bedrijvenbeleid. Het is daarom van het grootste belang de effectiviteit van de verschillende beleidsmaatregelen zo goed mogelijk in kaart te brengen.

De hoofdpodracht van de Expertwerkgroep Effectmeting is het gestructureerd in beeld brengen van evaluatiemogelijkheden om de directe effecten (output) te meten van een aantal beleidsmaatregelen van het ministerie van Economische Zaken. Het gaat daarbij om de vraag welk deel van de geobserveerde output is toe te schrijven aan de inzet van een beleidsinstrument. Bijvoorbeeld de hoeveelheid additionele R&D die bij bedrijven wordt uitgelokt met de Research & Development Aftrek (RDA) of de hoeveelheid extra uitstroom van bètatechnici op MBO-niveau richting topsectoren die plaatsvindt via de Centra voor Innovatief Vakmanschap (CIV). Evaluatie van de effectiviteit op outputniveau kan gezien worden als de eerste stap bij het evalueren van de uiteindelijke maatschappelijke en macro-economische effecten van een instrument, de effecten op outcome-niveau. Het meten van outcome-effecten van beleidsinstrumenten is echter erg lastig vanwege de invloed van exogene factoren, zoals de conjunctuur en internationale economische ontwikkelingen.

Het grote probleem bij empirische beleidsevaluaties is dat instrumenten vaak selectief worden ingezet; hierdoor is de groep van bedrijven die van een instrument gebruik maken niet vergelijkbaar met de bedrijven die dat niet doen. Een verschil in uitkomsten tussen beide groepen kan daarom zowel het gevolg zijn van de verschillen in bedrijfskarakteristieken als van de beleidsinterventie. Dit is problematisch, te meer omdat het verschil vaak wordt veroorzaakt door niet-geobserveerde kenmerken van bedrijven, waarvoor moeilijk valt te corrigeren. Stel dat vooral bedrijven met een goed management een subsidie aanvragen. Wanneer aanvragers met niet-aanvragers worden vergeleken, kan het gevonden effect zowel door de subsidie zijn ontstaan als door verschillen in managementkwaliteit. In de literatuur wordt dit zelfselectie genoemd. Om dit probleem te ondervangen zou men idealiter, net als bij het testen van medicijnen, een experiment met willekeurige toewijzing willen opzetten. Wanneer zo'n sociaal experiment niet mogelijk is, kan gezocht worden naar natuurlijke experimenten. Bij natuurlijke experimenten zijn door bestaande instituties groepen van vergelijkbare bedrijven ontstaan die wel en niet van het beleidsinstrument gebruik hebben gemaakt. De verschillende methoden om met behulp van een natuurlijk experiment effecten in kaart te krijgen worden in hoofdstuk 2 beschreven. Zonder sociaal of natuurlijk experiment is een betrouwbare effectmeting lastig; de geschatte effecten kunnen dan net zo goed zelfselectie als de effecten van beleid beschrijven.

Behalve econometrische analyses kunnen surveymethoden gebruikt worden om de effecten van beleid te onderzoeken. Vaak wordt er gebruik gemaakt van enquêtes onder bedrijven die wel en niet aan een interventie hebben meegedaan. Bij de surveymethode speelt, net als bij een niet-experimentele kwantitatieve evaluatie, het probleem van zelfselectie. Daarnaast bestaat het risico op strategisch gedrag en een lage en/of selectieve respons. Desalniettemin levert een enquête vaak additionele informatie op over de precieze werking van een instrument en de ervaringen van de gebruikers. Een enquête is daarom aan te bevelen als aanvulling en check op de in dit rapport beschreven kwantitatieve methodes.

Het Directoraat-Generaal Bedrijfsleven en Innovatie (DG B&I) heeft de Expertwerkgroep Effectmeting gevraagd mogelijke evaluatieontwerpen uit te werken voor zes beleidsinstrumenten. De geselecteerde instrumenten bestaan zowel uit traditionele generieke instrumenten als nieuwe, meer vraaggestuurde instrumenten, zoals de Topconsortia voor Kennis en Innovatie. Door de diversiteit aan beleidsinstrumenten ontvouwt zich over de gehele linie van beleidsopties een beeld over wat mogelijk is op het gebied van effectmeting per type instrument. In hoofdstuk 4 van dit rapport worden evaluatieontwerpen beschreven. De werkgroep is tevens gevraagd te reflecteren op de recent uitgevoerde WBSO-evaluatie (hoofdstuk 3). Tabel I geeft enerzijds een overzicht van de evaluatieontwerpen die de Expertwerkgroep het meest kansrijk acht voor een effectiviteitsmeting en geeft anderzijds de voornaamste observaties over de WBSO-evaluatie weer.

Uit deze exercitie blijkt dat het vaak mogelijk is om, op een innovatieve en creatieve manier, effecten op basis van natuurlijke experimenten te meten. Het lukt echter niet altijd om de effecten van een beleidsinstrument voor de gehele doelgroep van bedrijven met natuurlijke experimenten te vatten. Natuurlijke experimenten hebben vaak betrekking op een deel van de doelgroep of regeling met specifieke karakteristieken. In deze gevallen kunnen de effecten van het instrument dan ook alleen worden geschat voor een deel van de doelgroep of een deel van de regeling. Ter illustratie, bij de IPC wordt in één van de evaluatieopties gesuggereerd om voor de tenders van 2011 en 2012 gebruik te maken van de beoordeling van aanvragen door AgentschapNL. Bij de regression discontinuity-methode worden bedrijven die de IPC hebben aangevraagd in twee groepen verdeeld. Bedrijven die net een voldoende scoren en aanspraak kunnen maken op de IPC en bedrijven die net een onvoldoende scoren en dus geen IPC ontvangen. De andere deelnemende bedrijven worden buiten beschouwing gelaten. Vaak is extrapoleren van de resultaten naar de gehele populatie van bedrijven niet zonder risico.

Tabel I Evaluatieontwerp per instrument

Instrument	Advies voorkeursoptie(s) evaluatieontwerpen
<i>Innovatiekredieten (IK) van het Innovatiefonds MKB+</i>	<i>Propensity score matching op basis van AgNL beoordelingsprocedure</i>
<i>Research en Development Aftrek (RDA)</i>	<i>Discontinuïteit rond het maximum WBSO-uren en invoering RDA</i>
<i>Bewijs van Goede Dienst-gemeenten (BvGD)</i>	<i>Regressieanalyse op basis van enquêteresultaten</i>
<i>Centra voor Innovatief Vakmanschap (CIV)</i>	<i>Vergelijking van afgestudeerden en/of instellingen op basis van difference-in-difference</i>
<i>Innovatieprestatiecontracten (IPC)</i>	<i>Toekenning op basis van loting (sociaal experiment) en rangorde (regression discontinuity)</i>
<i>Topconsortia voor Kennis en Innovatie (TKI) en TKI-toeslag</i>	<i>Vergelijking TKI-bedrijven met andere bedrijven of vergelijking TKI-bedrijven met en zonder projectdeelname op basis van difference-in-difference</i>
<i>Reflectie op WBSO-evaluatie</i>	<i>Dynamisch paneldatamodel probeert voor zo veel mogelijk geobserveerde en niet-geobserveerde kenmerken te corrigeren, maar evaluatie van dit soort grote generieke maatregelen is problematisch. Maak in de toekomst meer gebruik van discontinuïteiten.</i>

Om in de toekomst de effecten van beleid nog beter inzichtelijk te maken beveelt de Expertwerkgroep Effectmeting aan om in de ontwikkelingsfase van nieuw beleidsinstrumentarium meer aandacht te besteden aan de logische onderbouwing van beleidinstrumenten (beleidstheorie), alsmede na te denken over hoe de effecten van een instrument in de toekomst kwantitatief kunnen worden geëvalueerd (evaluatieontwerp). Bij voorkeur zou hierbij, waar mogelijk, moeten worden gewerkt met sociale experimenten op basis van (gewogen) loting. Hiernaast verdient het de aanbeveling om, voordat beleid grootschalig wordt uitgerold, te werken met kleinschalige, makkelijk te evalueren pilots. Databeschikbaarheid is cruciaal voor goede evaluaties. Koppeling van bestanden, actieve monitoring van toe- en afgewezen aanvragen, alsmede gebruik van een financiële prikkel om de respons onder geëvalueerden te verhogen, zijn enkele maatregelen die kunnen worden genomen om de databeschikbaarheid te vergroten. Bovendien dienen instrumenten met een vergelijkbaar beleidsdoel zoveel mogelijk gezamenlijk te worden geëvalueerd. Ten slotte beveelt de werkgroep aan om tevens van de overige beleidsinstrumenten van het bedrijvenbeleid (met een grotere financiële omvang) en de andere beleidsterreinen van het ministerie van EZ de effecten in kaart te brengen, op basis van de in dit rapport gehanteerde aanpak.

Inhoudsopgave

Managementsamenvatting	II
Verantwoording en leeswijzer	V
Hoofdstuk 1 Marktfalen en bedrijvenbeleid	1
1.1 Inleiding.....	1
1.2 Marktfalens bij generiek beleid.....	1
1.3 Bedrijfslevenbeleid in de praktijk en de topsectorenaanpak.....	7
1.4 (On)mogelijkheden effectmeting van het bedrijvenbeleid.....	8
Hoofdstuk 2 Statistische methoden voor empirische beleidsevaluatie	11
2.1 Causaliteit en correlatie.....	11
2.2 Potentiële uitkomstenmode.....	11
2.3 Sociale experimenten.....	12
2.4 Natuurlijke experimenten.....	13
2.5 Standaard regressieanalyse.....	17
2.6 Surveymethoden.....	18
2.7 Conclusies.....	19
Hoofdstuk 3 Reflecties op WBSO Evaluatie: Een casestudie	21
3.1 Inleiding.....	21
3.2 De WBSO.....	21
3.3 De evaluatie van de WBSO 2006-2010.....	23
3.4 Conclusie.....	27
Hoofdstuk 4: Evaluatieontwerpen instrumenten	29
4.1 Inleiding.....	29
4.2 Innovatiekrediet (IK).....	29
4.3 Research and Development Aftrek (RDA).....	36
4.4 Bewijs van Goede Dienst (BvGD).....	44
4.5 Centra voor Innovatief Vakmanschap (CIV).....	52
4.6 Innovatieprestatiecontracten (IPC).....	58
4.7 Topconsortia voor Kennis en Innovatie (TKI).....	66
Hoofdstuk 5 Conclusie en aanbevelingen	73
5.1 Conclusie.....	73
5.2 Aanbevelingen.....	73

Verantwoording en leeswijzer

Inleiding

Het ministerie van Economische Zaken (EZ) heeft de behoefte om meer zicht te krijgen op de effectiviteit van zijn beleidsinstrumentarium. Mede naar aanleiding van het regeerakkoord van het Kabinet Rutte/Verhagen² en rapporten van de Algemene Rekenkamer en het CPB is in de Kamerbrief “Naar de Top: Het bedrijvenbeleid in actie(s)”³ een heldere ambitie uitgesproken op het gebied van monitoring en effectmeting van het bedrijvenbeleid. Dit rapport biedt concrete aangrijpingspunten om verdere invulling te geven aan deze ambitie aangaande de effectmeting van de beleidsinstrumenten die onder het bedrijvenbeleid vallen.

Om de effecten van beleid op een verantwoorde en robuuste manier in kaart te brengen, is het uitgangspunt het statistisch kunnen aantonen van een causaal verband van het gebruikte instrument op het gestelde doel. Dit is in de praktijk echter niet altijd eenvoudig. Daarom heeft het ministerie van EZ de “Expertwerkgroep Effectmeting” in het leven geroepen. De werkgroep heeft als doel om verschillende methodes te verkennen om de effectiviteit van beleidsinstrumenten in kaart te brengen en deze op hun relatieve merites te beoordelen.

Onder effectmeting wordt in dit rapport verstaan: *het kwantificeren van de directe effecten van een instrument op het gestelde doel*. Beschrijvende onderzoeksmethodes waarin effecten meer kwalitatief in beeld worden gebracht worden in dit rapport als een terugvaloptie beschouwd in die gevallen waar het niet mogelijk is de effecten van een instrument te kwantificeren. De meer procesmatige onderdelen van beleidsevaluaties (zoals de uitvoering van instrumenten door AgNL) vallen buiten de reikwijdte van dit rapport.

Het Directoraat-Generaal Bedrijfsleven en Innovatie (DG B&I) heeft de Expertwerkgroep Effectmeting gevraagd voor de volgende zes beleidsinstrumenten mogelijke evaluatieontwerpen uit te werken:

- 1) Innovatiekredieten (IK) van het Innovatiefonds MKB+;
- 2) Research en Development Aftrek (RDA);
- 3) Bewijs van Goede Dienst-gemeenten (BvGD);
- 4) Centra voor Innovatief Vakmanschap (CIV);
- 5) Innovatieprestatiecontracten (IPC);
- 6) Topconsortia voor Kennis en Innovatie (TKI) en TKI-toeslag.

De meeste instrumenten betreffen nieuw beleid, zoals dat is aangekondigd in de hierboven genoemde Kamerbrief. De geselecteerde beleidsinstrumenten zijn representatief voor het brede scala aan beleidsinstrumenten dat binnen het DG B&I wordt gehanteerd. De selectie bevat instrumenten met een uiteenlopend budgettair beslag en bestaat zowel uit traditionele generieke instrumenten (IK, RDA, IPC), nieuwe meer vraaggestuurde instrumenten (CIV en TKI) als een meer ‘indirect’ beleidsinstrument (BvGD). Met deze selectie verwacht het DG B&I een goed beeld te krijgen van de mogelijkheden en beperkingen van kwantitatieve effectmeting op een breed spectrum van beleidsopties.

Opdracht

De hoofdpdracht van de werkgroep is tot een advies te komen over de meest geschikte evaluatieontwerpen om de effectiviteit van bovengenoemde instrumenten in kaart te brengen gegeven het doel, de vormgeving, de werking van de instrumenten en de beschikbaarheid van data. Hiernaast is de werkgroep gevraagd een oordeel uit te spreken over de onderlinge kwaliteit van de verschillende evaluatieontwerpen en deze te rangschikken in volgorde van inzichtelijkheid en robuustheid op de effectiviteit van het instrument.

De werkgroep is daarnaast gevraagd de resultaten van lopende evaluaties op kwaliteit te beoordelen en te bezien op welke punten deze evaluaties in de toekomst kunnen worden verbeterd. Het gaat hierbij om de evaluaties van de WBSO en Technopartner SEED. De evaluatie van dit laatste instrument heeft vertraging opgelopen en kon derhalve niet in de werkgroep worden besproken.

² Meer expliciet, de zinsnede: “Subsidies worden alleen verstrekt indien de effectiviteit ervan is bewezen.”

³ Tweede Kamer, vergaderjaar 2010-2011, 32637, nr. 15.

Tot slot heeft het ministerie aan de werkgroep gevraagd om te bezien wat de mogelijkheden zijn om de effecten van beleid op de hogere beleidsdoelstellingen (zoals economische groei, innovativiteit) in kaart te brengen, alsmede de effecten van de sectorale en integrale vraaggestuurde benadering van de topsectorenaanpak.

Reikwijdte rapport

De evaluatieontwerpen die in dit rapport worden behandeld zijn er primair op gericht de directe resultaten die met een instrument worden gerealiseerd, de zogenaamde eerste-orde-effecten, inzichtelijk te maken. Effecten van instrumenten op hogere beleidsdoelen, de zogenaamde tweede- en derde-orde-effecten op bijvoorbeeld meer werkgelegenheid of economische groei, zijn lastig aan een bepaald instrument toe te schrijven. Voor zover de Expertwerkgroep mogelijkheden ziet om de tweede- en derde-orde-effecten van instrumenten binnen bedrijven en sectoren te achterhalen, denk aan innovatieve producten en productiviteit, worden deze vermeld.

De werkgroep spreekt geen oordeel uit over de werking en de economische legitimatie van beleidsinstrumenten. De werkgroep adviseert over evaluatieontwerpen, maar voert evaluaties niet zelf uit. De beslissing over welk bureau de evaluatie uiteindelijk zal uitvoeren wordt genomen door het ministerie.⁴ De werkgroep kan worden gevraagd te ondersteunen/adviseren in de samenstelling van begeleidingscommissies van een evaluatieonderzoek. Als het evaluatieonderzoek is afgerond, geeft de Expertwerkgroep een oordeel over de kwaliteit van de evaluatie.

Werkwijze

De Expertwerkgroep Effectmeting heeft per beleidsinstrument gesprekken gevoerd met de betrokken dossierhouders vanuit het beleid, alsmede de betrokken instrumentuitvoerders vanuit AgentschapNL (AgNL), Kwaliteitsinstituut Nederlandse Gemeenten (KING)⁵ en het Platform Bètatechniek. Ter voorbereiding op deze gesprekken heeft de werkgroep vanuit EZ per instrument factsheets en achtergrondinformatie ontvangen, zoals reeds bestaande evaluaties. Tijdens de gesprekken zijn de instrumenten uitvoerig besproken, zodat een goed beeld over de beoogde werking en uitvoering van de instrumenten ontstond. Op basis hiervan heeft de werkgroep ideeën voor evaluatieontwerpen bediscussieerd en uitgewerkt.

Leeswijzer

Overheidsingrijpen moet in economisch opzicht gerechtvaardigd zijn. Hoofdstuk 1 geeft een overzicht van de marktfalens die de legitimatie vormen voor het bedrijvenbeleid. Om beleid te evalueren kan gebruik worden gemaakt van verschillende methodes, hoofdstuk 2 bevat een overzicht. De WBSO is een goed voorbeeld van een instrument waarbij in de evaluaties gebruik is gemaakt van kwantitatieve methoden. In hoofdstuk 3 reflecteert de werkgroep op de meest recente evaluatie en geeft aan waar nog ruimte is voor verbetering. Hoofdstuk 4 is de kern van dit rapport. Het bevat voor de zes hierboven vermelde beleidsinstrumenten een beschrijving, een overzicht van evaluatieontwerpen en een advies welk evaluatieontwerp het meest kansrijk wordt geacht om de directe effecten van het instrument in kaart te brengen. Dit hoofdstuk is meer technisch van aard. De technische uitwerkingen van de evaluatieontwerpen zijn vooral bedoeld voor de meer ingevoerde lezer. Hoofdstuk 5, tot slot, bevat de algemene conclusies en aanbevelingen.

⁴ In het geval van het BvGD neemt EZ die beslissing in samenspraak met KING. Het betreft hier een gedeeld eigenaarschap. In het geval van CIV is dit het ministerie van OCW in samenspraak met het platform Bètatechniek, met uitzondering van de CIV voor het groene onderwijs. Hiervoor is EZ verantwoordelijk.

⁵ Opgemerkt dient te worden dat KING meer is dan een 'instrumentuitvoerder'. KING is mede-eigenaar en verantwoordelijk voor de doorontwikkeling van het instrument BvGD in de toekomst.

Hoofdstuk 1 Marktfalen en bedrijvenbeleid

1.1 Inleiding

In dit rapport staan zes beleidsinstrumenten centraal die onderdeel uitmaken van het Nederlandse bedrijvenbeleid. Het Nederlandse bedrijvenbeleid kenmerkt zich op de eerste plaats door een inzet op generieke randvoorwaarden. Randvoorwaarden zijn een breed begrip, het gaat van verstandig fiscaal en monetair beleid tot infrastructuur, mededinging, de toegang tot kapitaal, een goed opgeleide beroepsbevolking en een excellente onderzoeksinfrastructuur. Onderdeel van dit randvoorwaardelijke beleid zijn generieke beleidsinstrumenten zoals de WBSO, RDA en het Innovatiekrediet die in dit rapport aan bod komen. Deze generieke instrumenten zorgen, naast een aantrekkelijk vestigingsklimaat, voor het mitigeren van marktfalen waardoor private R&D beslissingen beter in lijn kunnen worden gebracht met het maatschappelijk wenselijke niveau. Marktfalen staat in dit hoofdstuk dan ook centraal als legitimatie voor overheidsingrijpen, zie paragraaf 1.2. Naast deze generieke pijler steunt het bedrijvenbeleid op een tweede pijler; het topsectorenbeleid. De centrale gedachte achter dit beleid is dat sectorspecifieke verschillen er toe doen. Een sectorale aanpak kan dan complementair zijn aan generiek beleid. In paragraaf 1.3 wordt nader ingegaan op het topsectorenbeleid. Tot slot worden in paragraaf 1.4 de (on)mogelijkheden van effectmeting van het bedrijvenbeleid besproken.

1.2 Marktfalens bij generiek beleid

Een belangrijk deel van het bedrijvenbeleid is erop gericht de positieve externe effecten van kennisinvesteringen te benutten (ook wel het internaliseren van externe effecten genoemd). Hier wordt het gros van de publieke middelen dat beschikbaar is voor het bedrijvenbeleid op ingezet. Marktfalen speelt dan ook het meest prominent bij innovatievraagstukken. Ook de legitimatie van andere niet direct op innovatie gerichte instrumenten die in dit rapport aan de orde komen vindt haar oorsprong in het bestaan van markt- of overheidsfalen. In de instrumentparagrafen wordt hier nadrukkelijker bij stilgestaan. Van de zes instrumenten die worden besproken in dit document zijn er vier gericht op het stimuleren van innovatie in de Nederlandse economie. Voor de eenvoud en leesbaarheid wordt daarom onderstaand innovatie als uitgangspunt genomen.

Het belang van innovatie

Innovatie betreft het toepassen van kennis en ideeën in nieuwe productiemethoden en organisatievormen en het op de markt brengen van nieuwe producten. In een concurrerende markteconomie hebben aanbieders er belang bij om nieuwe, kwalitatief betere en goedkopere producten op de markt te brengen. Dit genereert een, meestal tijdelijke, voorsprong op hun concurrenten, een groter marktaandeel en meer dan normale winsten. Deze prikkel tot innovatie zorgt niet alleen voor hogere winst voor de innovatieve ondernemer, maar ook voor een groei van de welvaart voor de consumenten in de vorm van betere en nieuwe producten, zorgt voor werkgelegenheid in de innoverende bedrijven, en als het nieuwe product op de internationale markt wordt verhandeld, ook voor meer export. Innovatie leidt vaak tot hogere productiviteit, die dan kan worden omgezet in lagere prijzen, hogere inkomens of meer vrije tijd.

Succesvolle innovatie is cruciaal voor de toename van de materiële en immateriële welvaart van de bevolking in de loop van de tijd. De structurele groei van het inkomen per hoofd sinds het begin van de industriële revolutie en de toename in vrije tijd en keuzemogelijkheden tussen opeenvolgende generaties is grotendeels te danken aan wat Baumol terecht aanduidt als “The Free Market Innovation Machine”.⁶ Permanente en succesvolle innovatie ontstaat pas als bestaande bedrijven en nieuwe toetreders daartoe worden aangespoord en geprikkeld. Een concurrerende omgeving is daarvoor een eerste voorwaarde, zij het dat de relatie tussen intensiteit van concurrentie en niveau van innovatie niet noodzakelijk rechtlijnig is.⁷ De vraag die tegelijkertijd rijst is of een concurrerende omgeving leidt tot het optimale niveau van innovatie, of in meer algemene termen geformuleerd, kan de samenleving innovatie aan de vrije markt overlaten? Het antwoord is negatief want, zoals hieronder wordt uiteengezet, de vrije markt faalt om een aantal redenen en genereert minder innovatie dan maatschappelijk optimaal is. Marktfalen is mogelijk aanleiding voor de overheid om innovatiebeleid te

⁶ William J. Baumol, 2004. *The Free-Market Innovation Machine: Analyzing the Growth Machine of Capitalism*, Princeton University Press, 2004.

⁷ Philip Aghion, Nick Bloom, Richard Blundell, Rachel Griffith en Peter Howitt, 2005, Competition and Innovation: An inverted U relationship, *Quarterly Journal of Economics*, vol. CXX (2), p. 701-728.

voeren en te streven naar een hoger optimaal niveau van innovatie.

Innovatie komt niet altijd even geleidelijk en gaat soms gepaard met tijdelijke, maar omvangrijke, maatschappelijke kosten. Het op de markt brengen van betere en nieuwe producten heeft bijna altijd tot gevolg dat bedrijven met minder goede producten marktaandeel verliezen (vergelijk de groei van Apple en de neergang van Nokia) of dat nieuwe innovatieve producten andere producten van de markt drukken (de pc met tekstverwerker versus de schrijfmachine). Innovatie is, zoals Schumpeter het formuleerde, een proces van creatieve destructie met winnaars en verliezers. Tot de winnaars horen de investeerders en werknemers in de succesvolle innoverende bedrijven, tot de verliezers de investeerders en werknemers in de bedrijven die worden weggeconcentreerd. Historisch gezien heeft het innovatieproces echter geleid tot een netto toename van de welvaart omdat de baten van de winnaars groter waren dan de verliezen van de verliezers.

Marktfalens

In de *Wealth of Nations* (1776) formuleerde Adam Smith, grondlegger van de economische wetenschap, de stelling dat ook als individuen en bedrijven bij hun beslissingen alleen hun eigenbelang in de gaten houden, er een *onzichtbare hand* is die er voor zorgt dat het algemene belang wordt gediend. Dit intuïtieve vermoeden van Adam Smith is door economen in de loop van de jaren uitgewerkt met als resultaat dat aangetoond kan worden dat onder ideale omstandigheden een concurrerende markt, waarin consumenten en producenten rationeel handelen, resulteert in een efficiënte allocatie van goederen, diensten en productiefactoren. Met efficiënte allocatie wordt bedoeld dat alle transacties worden gerealiseerd die voordelig zijn voor een individu of een bedrijf zonder dat een andere partij er nadeel van ondervindt⁸. Er wordt niet te weinig en ook niet teveel geproduceerd en verhandeld en in die zin leidt de perfect werkende markt tot het efficiënte niveau van de maatschappelijke welvaart. De markt faalt wanneer er ten opzichte van het efficiënte niveau te weinig of teveel wordt geproduceerd. Te weinig betekent dat profijtvolle transacties, met maatschappelijke baten groter dan kosten, niet tot stand komen; te veel betekent dat transacties waarvan de maatschappelijke kosten groter zijn dan de baten (bijvoorbeeld milieuvervuiling) toch worden doorgevoerd.

Marktfalens betekent dat de *onzichtbare hand* niet werkt en dat een vrije markteconomie niet leidt tot het efficiënte niveau van maatschappelijke welvaart. De overheid kan het als haar taak - of anders geformuleerd als een publiek belang - zien om het efficiënte niveau wel te realiseren en zal in dat geval beleid voeren om het marktfalens te corrigeren. Marktfalens als uitgangspunt en motivatie voor het beleid staat centraal in dit rapport⁹. Dat betekent echter niet dat marktfalens het enige mogelijke uitgangspunt is voor het overheidsbeleid. Marktfalens is een efficiëntiecriteria: er zijn te veel of te weinig transacties vergeleken met wat maatschappelijk optimaal is. Een ander belangrijk criterium voor overheidsbeleid is de verdeling van de maatschappelijke welvaart. Vrije marktwerking kan onder ideale omstandigheden leiden tot het ideale welvaartsniveau, maar tegelijk resulteren in een welvaartsverdeling over huishoudens en bedrijven die politiek of ethisch als niet rechtvaardig of niet aanvaardbaar wordt beschouwd. Herverdeling is dan een mogelijk belangrijke motivatie voor het beleid. Herverdelingsbeleid komt niet direct aan bod in dit rapport.

In de economische literatuur worden meestal de volgende vormen van marktfalens onderscheiden: marktmacht, publieke goederen, externe effecten, informatiegebreken en coördinatiegebreken. De meeste daarvan zijn ook relevant voor een mogelijk niet efficiënt niveau van innovatie in de bedrijven en daarmee aangrijpingspunt voor innovatiebeleid.

Marktmacht

Het typische voorbeeld van marktmacht is de monopolist. Een rationeel handelend monopolist zal zijn monopoliewinst maximaliseren door zijn aanbod te beperken, waardoor hij zijn product tegen een hogere prijs kan afzetten. In een monopoliemarkt wordt minder verhandeld dan maatschappelijk efficiënt is.

Monopoliesituaties kunnen ontstaan wanneer bedrijven samenspannen in een kartel, in het geval van een wettelijk

⁸ Deze vorm van efficiënte allocatie wordt een Pareto efficiënte allocatie genoemd.

⁹ De presentatie van de latere Nobelprijswinnaar Economie Kenneth Arrow voor de Joint Economic Committee van het Amerikaanse congres in 1969 is een van de eerste heldere uiteenzettingen van deze beleidsaanpak. Zie K. Arrow. (1969). "The Organization of Economic Activity: Issues Pertinent to the Choice of Market versus Non-market Allocation in," In *The Analysis and Evolution of Public Expenditure: The PPB System*. Vol. 1 U.S. Joint Economic Committee, 91st Congress, 1st Session. Washington D.C.: U.S. Government Printing Office, 59-73. In Nederland is de "De Calculus van het Publieke Belang" geschreven in opdracht van het ministerie van Economische Zaken en het ministerie van Financiën invloedrijk geweest voor deze aanpak. Zie C.N. Teulings, A.L. Bovenberg en H.P. van Dalen, 2003, *De Calculus van het Publieke Belang*, Kenniscentrum voor Orderingsvraagstukken, Ministerie van Economische Zaken, Den Haag.

monopolie wanneer een bedrijf van overheidswege het alleenrecht krijgt om bepaalde producten te verhandelen (de voormalige PTT bijvoorbeeld) of in het geval van een natuurlijk monopolie wanneer de productietechnologie zodanig is dat het altijd goedkoper is om één aanbieder te hebben dan meerdere (het spoorwegennet bijvoorbeeld). Het beleid ter correctie van het marktfalen in het geval van een monopolie kent vele routes. De aanpak van kartelvorming valt onder de mededingingsautoriteit. Het (de)reguleren van wettelijke en natuurlijke monopolies valt onder specifieke toezichthouders (Opta, vervoerskamer etc.¹⁰).

Marktmacht vormt een belemmering voor innovatie. Het samenspannen van een aantal bedrijven om in een sector de concurrentie te beperken, heeft ook tot gevolg dat er minder prikkels zijn om te innoveren. Marktmacht kan ertoe leiden dat de zittende bedrijven toetreding van nieuwe, innoverende bedrijven belemmeren of effectief kunnen dreigen om dat te doen. Dat kan bijvoorbeeld door overcapaciteit te creëren of door het hanteren van roofdierprijzen (predatory pricing). Roofdierprijzen zijn zeer lage (verliesgevende) prijzen die de zittende partijen met 'diepe zakken' net lang genoeg kunnen volhouden om potentiële toetreders te ontmoedigen. Een tweede reden waarom marktmacht kan leiden tot minder innovatie heeft te maken met het probleem van het kannibaliseren van het eigen product. Stel dat de marktmacht zo groot is dat sprake is van een monopoliesituatie (één aanbieder of een kartel). Een monopolist die een nieuw product op de markt brengt concurreert met zijn eigen bestaande product. Hij maakt winst op zijn nieuwe product maar verliest op zijn oude product, waardoor de prikkel om te innoveren minder wordt. Deze afweging is minder gunstig dan bij een nieuwe toetreders die geen eigen producten heeft die hem in de weg staan en die alleen winst maakt met het nieuwe product.

Marktmacht is een vorm van marktfalen die de welvaart van consumenten en eindgebruikers vermindert. Om die reden valt de aanpak van marktmacht onder het mededingingsbeleid en toezichtbeleid. Het mogelijk tekortschieten van innovatie wordt meestal al in de overwegingen van de mededingingsautoriteit en de toezichthouders meegenomen. Marktfalen in de vorm van marktmacht behoeft geen eigenstandig innovatiebeleid.

Publieke goederen

Bij publieke goederen wordt er door de markt te weinig of helemaal niets voorgebracht omdat de baten van het goed non-exclusief zijn, en omdat het goed wordt gekenmerkt door non-rivaliteit in het gebruik. Non-exclusiviteit van baten betekent dat het onmogelijk is om mensen van het gebruik van het goed uit te sluiten. Non-rivaliteit betekent dat het gebruik door de ene consument niet ten koste gaat van gebruik door een ander. Het klassieke voorbeeld van een publiek goed is een dijk. Na de bouw van de dijk wordt iedereen beschermd die achter de dijk woont en de bescherming van de ene bewoner gaat niet ten koste van een ander. Een privaat goed, bijvoorbeeld een appel, kan wel exclusief aan een gebruiker worden verkocht. En als hij de appel opeet, dan heeft niemand anders er nog wat aan. Consumenten kunnen profiteren van de voordelen van een publiek goed zonder er voor te betalen ('free rider' gedrag). Private producenten hebben daarmee onvoldoende prikkels om publieke goederen te produceren. Terwijl de baten van de publieke goederen wel de kosten overstijgen. De hoeveelheid publieke goederen die door de markt wordt geproduceerd is niet efficiënt. De productie van publieke goederen (zoals dijken, defensie, politiebescherming, rechtspraak) wordt meestal door de overheid ter hand genomen.

Succesvolle innovatie leidt tot nieuwe kennis in de vorm van een nieuw product of productieproces. Deze kennis bezit alle kenmerken van een publiek goed. Het is in principe niet mogelijk om anderen van het gebruik van deze kennis uit te sluiten (non-exclusiviteit) en het gebruik van de kennis door het ene bedrijf betekent niet dat de kennis minder wordt voor het volgende bedrijf (non-rivaliteit). Het overheidsbeleid wat betreft de generatie van nieuwe kennis bestaat in grote lijnen uit twee routes: fundamenteel onderzoek en toegepast, commercieel onderzoek. Fundamenteel onderzoek wordt meestal volledig gefinancierd uit overheidsmiddelen. Deze financiering komt ten goede aan hogeronderwijsinstellingen en onderzoeksinstituten. Toegepast en commercieel onderzoek wordt uitgevoerd door bedrijven omdat daarvoor toch nog voldoende prikkel bestaat om het te produceren, zeker in combinatie met de bescherming van de commerciële en toegepaste kennis in de vorm van intellectuele eigendomsrechten. Door gebruik te maken van intellectuele eigendomsrechten (auteursrechten, patenten, merkrechten) wordt het publiek goedkarakter van de nieuwe kennis en innovatie getransformeerd in een privaat goed. Dit gaat echter niet perfect en er blijven externe effecten van kennisproductie over.

¹⁰ Deze toezichthouders zullen per 1 januari 2013 samengaan in de Autoriteit Consument en Markt (ACM). Deze nieuwe toezichthouder ontstaat uit de samenvoeging van de Consumentenautoriteit, de Onafhankelijke Post en Telecommunicatie Autoriteit (OPTA) en de Nederlandse Mededingingsautoriteit (NMa).

Externe effecten

Externe effecten doen zich voor wanneer een transactie tussen twee marktpartijen de consumptie of productiemogelijkheden van een derde partij positief of negatief beïnvloedt zonder dat de derde partij voor de schade wordt gecompenseerd of voor de baat betaalt¹¹. Milieuvervuiling is het klassieke voorbeeld van een negatief extern effect. Bijvoorbeeld de benzinetransactie tussen een autobezitter en een oliemaatschappij leidt tot milieuvervuiling waar de rest van de samenleving last van ondervindt en niet voor wordt gecompenseerd. In het geval van externe effecten weerspiegelt de prijs van een product onvoldoende alle baten en kosten die maatschappelijk relevant zijn. Bij een negatief extern effect is de marktprijs te laag en wordt er meer geproduceerd dan maatschappelijk efficiënt is. Bij een positief effect zijn de private kosten te hoog omdat geen rekening wordt gehouden met de baten van derden en wordt er te weinig geproduceerd.

Innovatie leidt tot *positieve externe effecten*: de innovatieve inspanningen van een bedrijf hebben voordelen voor andere bedrijven en de rest van de samenleving, zonder dat dit bedrijf voor die externe effecten wordt vergoed. Het innoverende bedrijf houdt geen rekening bij zijn beslissingen over de omvang van zijn innovatieve investeringen met deze externe positieve effecten. Vanuit een maatschappelijk oogpunt zal het bedrijf daarom te weinig innoveren.

Deze positieve externe effecten worden ook *spillover effecten* of *uitstralingseffecten* genoemd. Innovatiebeleid door de overheid is er in dit verband op gericht om innoverende bedrijven bijvoorbeeld via subsidies te prikkelen om rekening te houden met deze uitstralingseffecten op derden en om dus meer te innoveren dan ze om private redenen van plan waren.

Meestal worden in dit verband twee soorten spillovers onderscheiden.¹² Ten eerste *kennisspillovers*. De kennisontwikkeling die gepaard gaat met innovatie heeft het karakter van een publiek goed. Zodra een aanbieder een nieuw of beter product op de markt heeft gebracht, kan een concurrent dat onmiddellijk kopiëren en namaken en een soortgelijk product op de markt brengen. Dit maakt het minder aantrekkelijk voor een onderneming om te innoveren. Het op de markt brengen van nieuwe producten of het inzetten van nieuwe productiemethoden is altijd een kwestie van investeren en uitproberen. Indien vervolgens een concurrent bijna gratis kan kopiëren, dan wordt de prikkel om te innoveren fors beknot.

Intellectuele eigendomsrechten beperken de mogelijkheden voor kopiëren en imitatie, maar ze verdwijnen niet helemaal. Omdat een innovator nooit volledig de volle vruchten van zijn innovatie kan plukken, wordt er minder geïnnoveerd dan maatschappelijk optimaal is. Een wat subtieler argument van kennisspillover is dat nieuwe kennis meestal voortbouwt op bestaande kennis ('standing on the shoulders of giants'). Zo kan de eerste generatie van een nieuw product of nieuw productieproces concurrenten inspireren tot een variant en tot een tweede generatie nieuwe producten en tot winstmogelijkheden waar de eerste innovator niet van profiteert.

Een tweede vorm van spillover is de *rentspillover*. Een nieuw product genereert een consumentensurplus. Consumentensurplus ontstaat wanneer eindgebruikers een hogere welvaart ervaren van het nieuwe product dan de marktprijs die ze er voor moeten betalen. Door prijsdiscriminatie (bijvoorbeeld hogere prijzen voor de "early adopters") kan een innovatief bedrijf een deel van het consumentensurplus afromen. Dit betekent dat het innovatieve bedrijf niet alle opbrengsten van zijn innovatie naar zich toe kan trekken en dus minder innoveert dan maatschappelijk optimaal is. Een soortgelijk argument kan worden gemaakt voor bijvoorbeeld een nieuwe machine die de productiekosten verlaagt bij de bedrijven die deze machine kopen. Daardoor stijgt het producentensurplus bij de bedrijven die deze machine gebruiken. Extra surplus dat het innovatieve bedrijf in het beste geval slechts deels kan afromen.

¹¹ Individuele beslissingen hebben bijna altijd een effect op anderen. Niet al deze externe effecten worden als marktfalen beschouwd. Bijvoorbeeld een gepromoveerde academicus die toetreedt tot de arbeidsmarkt verhoogt het arbeidsaanbod van promovendi en zal het loon van andere promovendi (met een onooglijk klein bedrag) verlagen. Een supermarkt die klanten weglukt van een andere supermarkt, is een ander voorbeeld. Dit soort markteffecten worden soms als 'financiële externe effecten' aangeduid. Ze worden niet als marktfalen beschouwd omdat ze geen probleem vormen wat betreft de efficiënte werking van de markt. Het gaat eerder om een herverdeling van inkomen en winsten (de ene partij gaat er op vooruit ten koste van de andere partij die er op achteruit gaat). Externe effecten die wel als marktfalen worden beschouwd worden soms aangeduid als 'technologische externe effecten' om te benadrukken dat ze de consumptie of productiemogelijkheden van een derde partij negatief of positief beïnvloeden (zonder dat daar enige compensatie tegenover staat).

¹² In de CPB publicatie "Innovatiebeleid in Nederland: De (on)mogelijkheden van effectmeting" uit 2011 wordt "business stealing effect" als een derde vorm van spillover genoemd. Dit effect is gerelateerd aan het Schumpeteriaanse creatie- en destructie- effect. Creatie en destructie van bedrijvigheid en de daarmee gaande gepaard gaande faillissementen van falende bedrijven en de groei van succesvolle bedrijven is echter een normaal gevolg van marktwerking.

Informatiegebreken

Als de ene marktpartij meer informatie heeft dan de ander, bestaat de kans op opportunistisch gedrag en suboptimaal gedrag. Het kan gaan om gebrekkige informatie over prijzen, kwaliteit, kosten of risico's waaraan men wordt blootgesteld. Als bijvoorbeeld het bieden van een betere kwaliteit voor de aanbieders meer kosten betekent, terwijl de koper de kwaliteit niet goed kan beoordelen, dan is er geen prikkel voor de aanbieder om een bovengemiddelde kwaliteit te bieden en voor consumenten om hiervoor te betalen. Er ontstaat op de markt een negatieve spiraal waarbij alleen producten van minder goede kwaliteit worden aangeboden met een nog altijd ongunstige prijs/kwaliteitverhouding. Deze vorm van marktfalen staat in de literatuur bekende als *averechtse selectie*.

Op sommige markten hebben niet de vragers, maar de aanbieders last van een informatieachterstand. Denk bijvoorbeeld aan de problemen die zouden ontstaan als een particuliere verzekeraar een verzekering tegen het werkloosheidsrisico zou aanbieden. De kans om een baan te vinden bij werkloosheid hangt niet alleen af van de economische conjunctuur maar ook van het gedrag van de werkzoekende. Een werkloosheidsverzekering kan als gevolg hebben dat minder hard naar een nieuwe baan wordt gezocht waardoor de werkloosheidsduur oploopt. Daardoor loopt de verzekeringspremie op en prijst een particuliere verzekeraar zich mogelijk uit de markt. Dit probleem wordt aangeduid als moral hazard ('moreel risico'). Moreel risico maakt dat sommige risico's niet verzekerd kunnen worden door particuliere verzekeringsmaatschappijen en dat er reden is voor de overheid om in te grijpen.

Het probleem van 'averechtse selectie' wordt in de markt vaak aangepakt doordat de aanbieders die een hogere kwaliteit aanbieden daarbij ook garanties geven (niet goed geld terug) of in de loop van de tijd een reputatie opbouwen, al dan niet gekoppeld aan een merkproduct. Het probleem van 'moral hazard' wordt in de markt aangepakt door nooit een volledige verzekering af te sluiten maar door altijd een deel van het risico op schade bij de verzekerde te laten.

Marktfalen wegens gebrekkige of asymmetrische informatie speelt op het gebied van innovatie vooral op de kapitaalmarkt bij de financiering van innovatieprojecten¹³. Dit falen van de kapitaalmarkten wordt veroorzaakt door gebrek aan informatie over de commerciële slagingskansen van het innovatieve project. De aanvrager van financiering voor een innovatieproject heeft betere informatie over de slagingskansen van het project dan de kredietverschaffer. Deze informatieasymmetrie maakt de kredietverlener terughoudend in het honoreren van innovatieprojecten. Een markt met asymmetrische informatie zoals de kapitaalmarkt voor innovatieve projecten kan ertoe leiden dat kapitaalverschaffers relatief hoge rentepercentages vragen en andere strenge leningsvoorwaarden stellen zodat vooral de slechte risico's (innovatieve projecten met lage slagingskansen) kredieten zullen aanvragen (averechtse selectie). De externe financiering kan ook aanleiding zijn om het innovatieproject minder strak uit te voeren dan het geval zou zijn bij alleen eigen financiering, waardoor het risico groter wordt (moreel risico). Marktfalen wegens informatiegebreken leidt ertoe dat de kapitaalmarkt voor innovatieprojecten minder omvangrijk wordt, wat de mogelijkheden beperkt voor de ondernemer met een innovatief idee om voldoende financieringsmiddelen uit de markt te trekken.

Voor een deel wordt de problematiek van de asymmetrische informatie door de markt zelf opgelost. De reputatie van een kredietaanvrager is een element dat de kredietverschaffer vertrouwen kan geven. Ook het geven van een onderpand bij de lening kan helpen om het probleem van de asymmetrische informatie te overbruggen. Het probleem blijft echter bestaan voor nieuwe beginnende ondernemers die nog geen reputatie hebben kunnen opbouwen of waarvoor het vaak niet mogelijk is om voldoende onderpand of medefinanciering te leveren om de onzekerheid bij de kapitaalverschaffer te verminderen.

Coördinatiegebreken

Een markteconomie werkt alleen perfect en leidt tot een efficiënt welvaartsniveau indien ze beschikt over een complete set aan markten. Dit betekent dat er $N \times T \times S$ markten moeten zijn, zodat alle N goederen en diensten verhandeld kunnen worden, er termijncontracten kunnen worden afgesloten voor alle toekomstige periodes T en verzekeringscontracten kunnen worden afgesloten waarmee men zich kan indekken tegen alle mogelijke risico's S die zich kunnen voordoen. In de praktijk beschikt geen enkele markteconomie over een volledige set aan markten. Er is slechts een beperkt aantal termijnmarkten waarop toekomstige contracten kunnen worden afgesloten en het is ook niet mogelijk om verzekeringscontracten voor alle risico's af te sluiten. Het ontbreken van markten is een vorm van marktfalen omdat

¹³ Noteer dat de soms extreem grote risico's die samenhangen met investeringen in onzekere innovatieprojecten niet als een vorm van marktfalen wordt beschouwd. In de markt worden grote onzekerheid vertaald in hoge risicopremies bij de financiering. Succesvolle innovatieprojecten kunnen ook zeer hoge winsten opleveren.

transacties die voor individuen en bedrijven interessant zouden kunnen zijn, niet kunnen worden uitgevoerd. Er is een aantal redenen waarom markten niet tot stand komen. In de eerste plaats zijn er transactiekosten verbonden aan de werking van de markt. Bijvoorbeeld het opstellen van termijncontracten en verzekeringscontracten vereist dat vooraf in detail alle mogelijke risico's en eventualiteiten moeten worden beschreven en hoe ze al dan niet worden afgedekt. De kosten voor het vergaren van al deze informatie en het beschrijven ervan kunnen zo groot worden dat het niet meer interessant is om dergelijke contracten op te stellen.

Op het gebied van innovatie kunnen transactiekosten tot gevolg hebben dat bijvoorbeeld bepaalde financieringsmarkten niet of niet in voldoende mate tot stand komen. Een mogelijk voorbeeld is dat er wel een markt is voor financiering van zeer risicovolle innovatieprojecten (via gespecialiseerde markten voor durfkapitaal) en voor weinig risicovolle projecten (gemakkelijk te financieren door bankkrediet) maar dat er geen financieringsmarkt tot stand komt voor tussengelegen risico's.

Een tweede reden voor het ontbreken van markten zijn zoek- en informatiekosten voor het vinden van een geschikte tegenpartij. Om een voor beiden voordelige markttransactie af te sluiten moeten een koper en een verkoper elkaar eerst vinden. In de meeste gevallen (de dagelijkse boodschappen) is dat geen probleem, maar in andere gevallen is niet gelijk duidelijk waar de geschikte tegenpartij voor een transactie te vinden is. Partijen moeten elkaar dan zoeken. De arbeidsmarkt en de woningmarkt zijn twee voorbeelden waar dit speelt. Zoeken naar een geschikte tegenpartij vereist de inzet van tijd en middelen. Een werkgever met een vacature moet een advertentie plaatsen en kandidaten selecteren en toetsen. Een huiseigenaar die zijn huis wil verkopen moet dit –eventueel met de hulp van een makelaar- adverteren en zijn huis voor inspectie beschikbaar stellen. Als niet snel een geschikte tegenpartij wordt gevonden, kunnen de zoekkosten oplopen en kan besloten worden om niet langer te blijven zoeken of om de transactie te sluiten met een tegenpartij die weliswaar niet optimaal is maar het beste wat men tot dan toe heeft kunnen krijgen. In beide gevallen faalt de markt om tot een efficiënt resultaat te komen.

In de innovatiemarkt zijn er partijen, zoals onderzoeksinstituten en universitaire onderzoekers die over kennis beschikken die mogelijk interessant is voor bedrijven om commercieel uit te werken en te exploiteren. Om afspraken te kunnen maken moeten beide partijen elkaar eerst vinden. Onderzoekers weten niet welke bedrijven mogelijk in hun kennis zijn geïnteresseerd en bedrijven weten niet bij welke onderzoekers kennis zit die voor hen commercieel interessant zou kunnen zijn. De zoekkosten die ze moeten maken om elkaar te vinden kunnen zo groot zijn dat het contact nooit tot stand komt, waardoor innovaties die voor de samenleving voordelig zijn niet plaatsvinden.

Marktfalen als uitgangspunt

Bij het uitgangspunt dat marktfalen het centrale basisprincipe is voor innovatiebeleid horen een aantal waarschuwingen.

Ten eerste kan beredeneerd worden dat marktfalen in de richting gaat van onderaanbod van innovatie. Zoals we hebben gezien leiden spillovereffecten, kapitaalmarktfalen en coördinatiegebreken allemaal tot een onderaanbod en dus tot een lager niveau van innovatie dan wat maatschappelijk efficiënt is. Tegelijkertijd is het echter niet mogelijk om te meten hoe omvangrijk het marktfalen is en dus hoe groot het onderaanbod. Daarmee is ook de optimale omvang van het overheidsingrijpen onzeker. De omvang van het innovatiebeleid, geredeneerd vanuit spillovereffecten, is optimaal indien het beleid tot gevolg heeft dat alle innoverende bedrijven bij hun investeringen in innovatie niet alleen rekening houden met hun eigen baten maar ook met de baten van de anderen. Het is een haast onmogelijk opgave om te evalueren of de omvang van het innovatiebeleid optimaal is en niet te hoog of te laag is om de spillovereffecten precies te compenseren. Op zijn minst moet effectief innovatiebeleid ter correctie van marktfalen gericht zijn op het verhogen van het innovatiepeil in de economie¹⁴.

Ten tweede geldt dat het opheffen van marktfalen niet het enige mogelijke uitgangspunt is voor innovatiebeleid. Herverdelingsoverwegingen waarbij sommige bedrijven meer innovatiesubsidies krijgen dan andere bedrijven kunnen ook een motief zijn voor innovatiebeleid. Herverdeling is een politieke keuze, die mogelijk tot inefficiënte resultaten leidt.

Ten derde, innovatiebeleid leidt ook tot kosten (uitvoeringskosten, kosten van belastingheffing ter financiering van innovatiesubsidies). Het is niet onmogelijk dat deze beleidskosten groter zijn dan de baten die gegenereerd worden in de

¹⁴ Preadviezen Koninklijke Vereniging voor de Staaathuishoudkunde (2004), Innovatie in Nederland: De markt faalt en de overheid draait, redactie Bas Jacobs en Jules Theeuwes

vorm van een hoger innovatieniveau. In dat geval is het innovatiebeleid niet efficiënt.

Ten vierde, naast marktfalen is er ook overheidsfalen. Het uitgangspunt bij de theorie van marktfalen is dat de overheid het algemene belang nastreeft. Marktfalen betekent dat de private markt niet in staat is om het algemeen belang te realiseren. Dit is dan aanleiding voor de overheid om dit via het beleid te corrigeren en te repareren. Het is echter niet altijd zo dat de overheid het algemeen belang voor ogen heeft en het is zelfs zo dat overheidsbeslissingen er zelf oorzaak van zijn dat het innovatieniveau lager is dan optimaal. Voorbeelden hiervan zijn vormen van regulering op het gebied van veiligheid die innovatieve activiteiten onnodig beperken. Of denk bijvoorbeeld aan het steunen van private belangen in plaats van het algemene belang. Zittende bedrijven in een bepaalde sector kunnen bijvoorbeeld lobbywerk verrichten voor het instellen van startvergunningen, registratieverplichtingen en andere administratieve regelgeving, die voor nieuwe, mogelijk innovatieve toetreders, de toegang onnodig moeilijk maakt.

1.3 Bedrijfslevenbeleid in de praktijk en de topsectorenaanpak

Er wordt door veel bedrijven gebruik gemaakt van de mogelijkheden die het bedrijfslevenbeleid biedt. In de periode 1998-2008 werden circa 20.000 bedrijven bereikt met subsidie- en kredietregelingen. De Top-25 qua aantal deelnames bevat de grote 9 (Philips, ASML, Shell, DSM, NXP, Unilever, Océ, KPN, AkzoNobel), maar ook verschillende kleine bedrijven. De grote participeren vooral in de samenwerkingsregelingen en de kleine in de kennisoverdrachtregelingen¹⁵. De grootste bedragen worden daarbij vertimmerd in de generieke pijler van het bedrijvenbeleid. Zo gaat er in de fiscale innovatie-instrumenten¹⁶ circa 1,8 miljard om. Echter, ondanks deze generieke instrumenten blijft een aantal hardnekkige bottlenecks bestaan, zoals achterblijvende private R&D-investeringen, beperkte kennisvalorisatie door publiek gefinancierde onderzoeksinstituten en universiteiten, een gebrek aan coördinatie tussen de verschillende overheden ten aanzien van economische diplomatie, de aanwezigheid van sectorspecifieke regeldruk en een gebrekkige aansluiting tussen onderwijs en arbeidsmarkt.

Bedrijven leunen, naast hun eigen kennisontwikkeling, steeds meer op andere bedrijven en misschien nog wel meer op (semi-)publieke kennisinstellingen om hun kennisbasis te verbreden en/of te verdiepen. Goede interactie en samenwerking binnen bedrijven, tussen bedrijven en tussen bedrijven en kennisinstellingen zijn daarom van groot belang.

Bovenstaande ontwikkelingen vragen om een sectorale aanpak die naast het generieke beleid kan opereren om de nog steeds aanwezige knelpunten het hoofd te bieden. Het nieuwe op negen topsectoren gerichte bedrijvenbeleid tracht de tekortkomingen van de markt die hierboven uitvoerig besproken zijn op een effectieve manier aan te pakken door het bedrijfsleven, kennisinstellingen en de overheid met elkaar te verbinden.

Bij het adresseren van marktfalen wordt de overheid zelf echter ook geconfronteerd met informatieasymmetrie en coördinatiefalen. Het ontbreekt de overheid vaak aan sectorspecifieke kennis en informatie om marktfalen op een effectieve manier te ondervangen, wat op haar beurt weer kan leiden tot overheidsfalen. Bovendien leidt de manier waarop de overheid is georganiseerd (meerdere autonome lagen op verschillende niveaus) tot coördinatieproblemen, wat tot minder efficiënt beleid leidt. Ook verschillen positieve externe effecten en knelpunten per sector, waardoor het inspelen op kennisspillers om een gedifferentieerde aanpak vraagt.

Om deze problemen het hoofd te kunnen bieden is het nieuwe bedrijvenbeleid sectorgeoriënteerd, vraaggestuurd en geïntegreerd. Een geïntegreerde benadering betekent dat het nieuwe bedrijvenbeleid zich richt op het wegnemen van knelpunten voor het vestigingsklimaat in brede zin (zoals financiering, sectorspecifieke regulering, innovatie, menselijk kapitaal). Bovendien betekent het dat het beleid niet georganiseerd is langs de lijnen van overheidsschotten, maar langs de lijnen van economische sectoren, zodat beleidsthema's integraal worden gezien.

In de praktijk betekent dit dat er voor negen topsectoren teams zijn geformeerd die in overleg met de sectoren tot sectoradviezen zijn gekomen. Deze adviezen rusten op verschillende pijlers, te weten innovatie- en kennisroadmaps, human capital agenda's en internationaliseringsagenda's. Hiernaast bevatten de adviezen ook voorstellen op het gebied van regulering en duurzaamheid. Deze dialoog met de sectoren stelt ondernemers, wetenschappers en de overheid in

¹⁵ Zie AgNL, Deelname bedrijven aan innovatieregelingen 1998-2008, november 2010

¹⁶ WBSO, RDA en Innovatiebox.

staat op maat gemaakte oplossingen te vinden die meer succes moeten oogsten bij het tegengaan van sectorspecifieke knelpunten dan voorgekookt overheidsbeleid. De TKI's en de CIV's zijn voorbeelden van nieuwe instrumenten die binnen deze aanpak passen en coördinatiegebreken proberen op te lossen. De sectoragenda's zijn echter breder en bevatten specifieke acties voor bedrijfsleven, kennisinstellingen en overheid. Deze acties zijn zeer divers. Ter illustratie, de sector logistiek wil een synchromodaal transportsysteem ontwikkelen om de beladingsgraad te verhogen. Dit betekent dat het bedrijfsleven zelf gaat investeren in ICT en het vaker en slimmer bundelen van ladingen, de overheid zal wet- en regelgeving aanpassen, terwijl kennisinstellingen onderzoek doen naar nieuwe contractvormen, businessmodellen en complexe besluitvormingsmodellen.

1.4 (On)mogelijkheden effectmeting van het bedrijvenbeleid

De eerder in dit hoofdstuk beschreven vormen van marktfalen worden door het ministerie van EZ gebruikt als rechtvaardiging van beleid. Een goede beleidstheoretische onderbouwing betekent nog niet automatisch dat het beleid ook daadwerkelijk effectief zal zijn. Zoals zo vaak is "the proof of the pudding in the eating". Gezien het belastinggeld dat gemoeid is met beleid is het van het grootste belang om de effectiviteit van beleid in kaart te brengen. Daartoe heeft deze werkgroep de opdracht gekregen om, als belangrijke eerste stap, de directe effecten van beleidsinstrumenten in kaart te brengen gegeven de doelen van de instrumenten en de beschikbaarheid van data. Het is hier belangrijk de mogelijkheden, maar vooral ook de onmogelijkheden van effectmeting te identificeren.

De werkzaamheden van de expertwerkgroep effectmeting richten zich vooral op het meten van de additionaliteit van de *output* van de verschillende beleidsmaatregelen (zie figuur 1). Het meten van hogere orde effecten van beleid op bijvoorbeeld economische groei of de concurrentiekracht is daarentegen een stuk lastiger, zo niet onmogelijk gezien de stand van de wetenschap en methodologie. Bij elk stapje hogerop in de diverse niveaus (zie figuur 1) verwatert de invloed van het beleidsinstrument. Ook de factor tijd speelt een beperkende rol. Doorwerking van effecten neemt vaak langere tijd in beslag; extra R&D vertaalt zich niet direct in meer innovatie en deze innovatie vertaalt zich niet direct in meer economische groei of een betere concurrentiepositie. In de ideale situatie waarin een beleidsinstrument slechts één beleidsdoel heeft, kan de stelling worden aangegaan dat indien geen direct effect in een evaluatie wordt gevonden, het onwaarschijnlijk is dat dit beleidsinstrument een bijdrage heeft geleverd aan het halen van hogere-orde-effecten.

De interpretatie van effecten op outputniveau worden vaak doorgetrokken naar effecten op outcomeniveau op basis van theoretische redeneringen. Het theoretisch overbruggen van de kloof tussen output en outcome kent echter gevaren. Een theoretische correlatie is immers nog geen causaliteit. Bovendien worden correlaties vaak overschat en worden afnemende meeropbrengsten nog weleens uit het oog verloren.

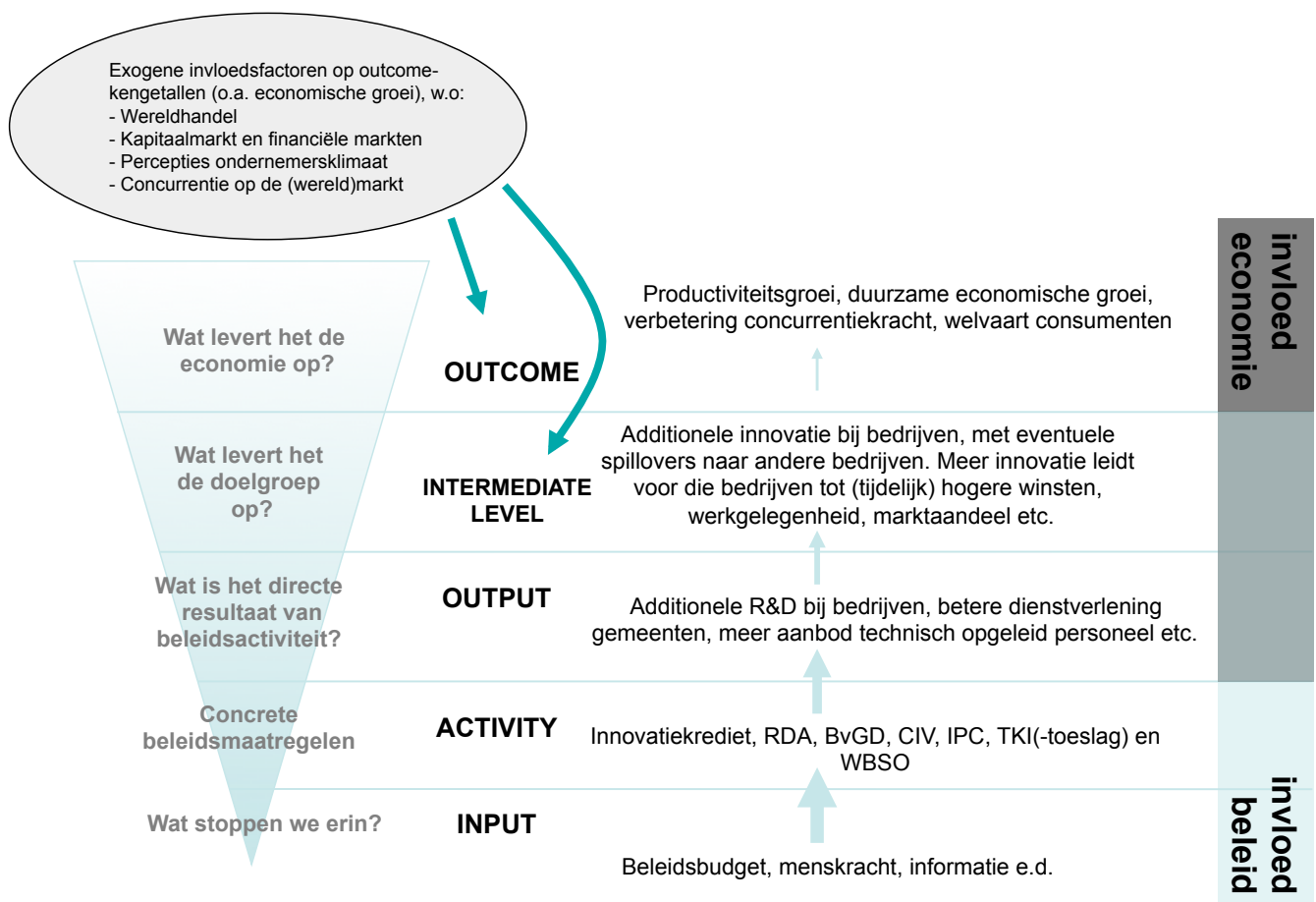
Als alternatief kan op basis van wetenschappelijke studies in een bredere context worden gekeken naar outcome effecten. Zo zijn in de econometrische literatuur wel tal van studies gedaan naar het *outcome* effect van bijvoorbeeld extra R&D. Deze worden onder andere samengevat in de recente CPB studie van Lanser en Van der Wiel (2011) die in belangrijke mate leunt op het overzichtsartikel van Hall et al.¹⁷ Dit overzichtsartikel concludeert dat het maatschappelijk rendement van R&D, gemeten als de (marginale) bijdrage aan het BBP van de R&D-inspanning, hoog is maar onzeker. De uitkomsten hangen in sterke mate af van de gekozen tijdsperiode, sectoren, landen en de gebruikte econometrische specificatie. Het private rendement van R&D kan bovendien niet als een constante worden beschouwd, maar is het resultaat van een complexe interactie tussen bedrijfsstrategie, de strategie van concurrenten en een stochastische macro-economische omgeving. Ook hier is dus voorzichtigheid geboden.

Daar waar mogelijk biedt de werkgroep wel een doorkijk naar het meten van effecten op hogere niveaus. Het gaat hierbij vooral om doorwerking van extra R&D-uitgaven in meer innovatieve producten (een tweede-orde-effect). Deze effecten kunnen ook in een evaluatie worden onderzocht wanneer voldoende data over een lange periode beschikbaar is. Het meten van deze effecten blijft, zoals eerder aangegeven, echter lastig. Indien een goede controlegroep voorhanden is, zoals dat bij een sociaal experiment het geval is (zie hoofdstuk 2), wordt het makkelijker om tweede-orde-effecten in kaart te brengen.

¹⁷ B. H. Hall, J. Mairesse en P. Mohnen, 2009, Measuring the returns to R&D, Working Paper 15622, National Bureau of Economic Research, Cambridge, Massachusetts. Een meer recente NBER working paper levert ook een bijdrage in deze discussie en heeft een lijst met referenties: Atkeson & Burstein, 2011, Aggregate implications of innovation policy, NBER WP 17493.

Evaluatie van de integrale topsectorenaanpak is lastig. Het door het DG B&I ontwikkelde monitoringsysteem is een nuttig instrument om de voortgang van de aanpak te meten en deze onderling tussen de verschillende topsectoren te vergelijken. Het dwingt partijen hun ambitieniveau te onderbouwen en het maakt vergelijkingen tussen topsectoren mogelijk. Voor effectmeting zijn de doelenbomen en tabellen minder relevant, aangezien de doelen endoogeen zijn. De topteams hebben ze zelf gekozen, inclusief streefwaarde, waardoor het probleem van selectiviteit zich voordoet. Het meten van de effectiviteit van beleidsprogramma's als de topsectorenaanpak staat nog in de kinderschoenen. De Organisatie voor Economische Samenwerking en Ontwikkeling (OESO) heeft een expertgroep op het gebied van *Industrial Policy* in het leven geroepen waar landen hun ervaringen met evaluaties van beleidsinstrumenten, maar ook beleidsprogramma's zullen gaan uitwisselen. Dit biedt voor de toekomst wellicht verdere aanknopingspunten.

Figuur 1 - Effecten van (bedrijfsleven)beleid op verschillende niveaus



Hoofdstuk 2 Statistische methoden voor empirische beleidsevaluatie

2.1 Causaliteit en correlatie

In dit hoofdstuk geven we een beknopte beschrijving van de bestaande econometrische methoden die worden gebruikt om de effecten van beleidsinterventies te bepalen. Bij empirische beleidsevaluatie willen we weten hoe uitkomsten veranderen door de inzet van beleidsinstrumenten. Hierbij proberen we te achterhalen of de doelvariabele¹⁸ zónder de beleidsinterventie een ander niveau heeft dan mét de beleidsinterventie. Met andere woorden, is er sprake van een causaal effect van de beleidsinterventie op de doelvariabele? Dit is niet eenvoudig te meten omdat de doelvariabele niet alleen door de beleidsinterventie kan veranderen, maar ook door andere factoren. Stel bijvoorbeeld dat het doel van een overheidssubsidie aan een bedrijf is om bedrijven meer te laten innoveren en dat moet worden vastgesteld of deze interventie effect heeft gehad. De relevante doelvariabele is dan bijvoorbeeld uitgaven aan R&D, aantal innovatieve producten, omzet, winst, of de arbeidsproductiviteit in een bedrijf. Omdat deze doelvariabelen van veel zaken afhangen, zoals conjunctuur, innovaties in het buitenland en eventuele andere beleidswijzigingen, is het lastig om het causale effect tussen de beleidsinterventie en de doelvariabele vast te stellen.

Het grote probleem bij empirische beleidsevaluatie is dat instrumenten zoals subsidies of kredieten vaak selectief ingezet worden. Dat wil zeggen dat de groep van bedrijven die heeft meegedaan aan een bepaalde interventie niet vergelijkbaar is met de groep van bedrijven die hieraan niet heeft meegedaan. Een verschil in uitkomsten tussen beide groepen kan daarom net zo goed het gevolg zijn van de verschillen in bedrijfskarakteristieken als van de beleidsinterventie. De correlatie tussen deelname en uitkomst is daarom bij selectieve inzet een onbetrouwbare schatting van het causale effect van de interventie.

Hieronder bespreken we een aantal methoden om met dit probleem om te gaan. Daarbij kijken we naar sociale en natuurlijke experimenten. Bij een sociaal experiment is door willekeurige toewijzing bepaald welke bedrijven meedoen aan de interventie en welke niet. Bij natuurlijke experimenten is deze willekeurige toewijzing er niet geweest met als doel het beleid te evalueren, maar zijn door bestaande instituties groepen vergelijkbare bedrijven ontstaan die al dan niet meegedaan hebben aan de interventie. Dit hoofdstuk geeft slechts een beknopte samenvatting van bestaande methoden. Uitgebreidere overzichten kunnen bijvoorbeeld worden gevonden in Angrist en Pischke (2009), Blundell en Costa Dias (2007), DiNardo en Lee (2011) en Imbens en Wooldridge (2009).¹⁹

2.2 Potentiële uitkomstenmodel

We beginnen met de bespreking van een stilistisch model voor beleidsevaluatie. Het potentiële uitkomstenmodel veronderstelt dat ieder bedrijf²⁰ twee mogelijke uitkomsten heeft. De eerste uitkomst Y_1^* geeft de uitkomst weer die zou worden waargenomen als het bedrijf heeft meegedaan aan de interventie. En als het bedrijf niet heeft meegedaan wordt de uitkomst Y_0^* waargenomen. Het verschil tussen beide uitkomsten $\Delta = Y_1^* - Y_0^*$ is het causale effect van de beleidsinterventie. In dit geval is de beleidsinterventie bijvoorbeeld het ontvangen van een innovatiekrediet en de relevante uitkomst de R&D-uitgaven van een bedrijf.

¹⁸ Doelvariabele is een ander woord voor afhankelijke variabele.

¹⁹ Angrist, J.D. & S-S Pischke (2009), *Mostly Harmless Econometrics*, Princeton University Press. Blundell, R. & M. Costa Dias (2007), *Alternative Approaches to Evaluation in Empirical Microeconomics*, *Journal of Human Resources* 44, 565-640. DiNardo, J. and D.S. Lee (2011), *Program evaluation and research designs*, in O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics, Volume 4A*, Amsterdam: Elsevier Science. Imbens, G.W. & J.M. Wooldridge (2009), *Recent Developments in the Econometrics of Program Evaluation*, *Journal of Economic Literature* 47, 5-86.

²⁰ In sommige gevallen is het beleid niet gericht op bedrijven, maar bijvoorbeeld op scholen (CIV), gemeenten (BvGD) of kennisinstellingen (TKI). We spreken in dit hoofdstuk alleen over bedrijven vanwege de leesbaarheid.

Het grote probleem is dat het causale effect Δ voor een bedrijf niet kan worden geobserveerd. Een bedrijf heeft óf meegedaan aan de interventie óf niet meegedaan aan de interventie, dus maar één van de mogelijke uitkomsten Y_1^* en Y_0^* wordt geobserveerd. De uitkomst die ongeobserveerd blijft, wordt de *counterfactual* genoemd.

Vervolgens definiëren we een variabele D die aangeeft of het bedrijf heeft meegedaan aan de interventie; zo ja dan is D gelijk aan 1 en anders gelijk aan 0. Het standaardprobleem is dat data vaak afkomstig zijn uit een situatie waarin bedrijven zelf kiezen of ze meedoen aan een interventie: dit noemen we het “selectiviteitsprobleem”. Een voorbeeld hiervan is een situatie waarin alle bedrijven aanspraak kunnen maken op een subsidie, maar waarin niet elk bedrijf ervoor kiest om dat te doen. Een soortgelijk probleem doet zich voor als een subsidieverstrekker uit een aantal aanvragen alleen de beste aanvragen honoreert. In dat geval geldt dat het meedoen van de interventie niet onafhankelijk is van de mogelijke uitkomsten. Het vergelijken van bedrijven die meedoen aan de interventie met bedrijven die niet meedoen, geeft dan geen zuivere schatting van het effect van de interventie.

2.3 Sociale experimenten

De meest zuivere manier om het effect van een interventie te schatten is om een (sociaal) experiment op te zetten waarin deelname willekeurig wordt toegewezen. Of een bedrijf aan de interventie meedoet, is dan bepaald door loting. In praktijk zal de loting van toekenning van de subsidie pas plaatsvinden nadat bedrijven een aanvraag gedaan hebben voor de subsidie. De doelgroep bestaat dan uit bedrijven die interesse hebben in de subsidie. Er ontstaan zo een *experimentele groep*²¹ en een *controlegroep* die onderling vergelijkbaar zijn, behalve als het gaat om het wel of niet meedoen aan de interventie. In dit geval is er geen selectiviteitsprobleem en geeft het verschil tussen de gemiddelde uitkomsten in de experimentele groep en controlegroep een goede schatting van het effect van de interventie.

Er liggen hier twee veronderstellingen aan ten grondslag. Ten eerste, de deelname aan de interventie D is onafhankelijk van de mogelijke uitkomsten Y_1^* en Y_0^* . Dit is een cruciale veronderstelling, waarvan de geldigheid wordt gegarandeerd door de willekeurige toekenning van de interventie. De tweede veronderstelling is dat de uitkomst van een bedrijf niet afhankelijk is van hoeveel en welke bedrijven meedoen aan de interventie. Ook bij een experiment is niet op voorhand duidelijk of deze veronderstelling geldig is. Als bijvoorbeeld een directe concurrent een innovatiesubsidie krijgt, kan dit ook gevolgen hebben voor de uitkomsten van het bedrijf. Als de bedrijven in de controlegroep meeprofiteren van de extra innovatie die wordt gedaan door de bedrijven in de experimentele groep, dan wordt het effect van de interventie onderschat. Het omgekeerde kan ook het geval zijn. Per situatie zal moeten worden bekeken of er interacties tussen bedrijven in de experimentele en controlegroep zijn en wat de mogelijke gevolgen daarvan zijn.

Sociale experimenten komen bij bedrijvenbeleid niet zo vaak voor als bijvoorbeeld bij onderwijs-, arbeidsmarkt- of ontwikkelingsbeleid. Sociale experimenten zijn dan idealiter een pilot om vooraf te toetsen of een interventie voldoet aan de verwachtingen. Dat betekent ook dat sociale experimenten vaak redelijk kleinschalig zijn. De schaal moet vooraf bepaald worden door een poweranalyse²², waarbij rekening gehouden wordt met het verwachte effect van de interventie en de normale onverklaarde variatie in uitkomsten.

In veel gevallen zijn data niet afkomstig van een sociaal experiment en kan ook niet zomaar aangenomen worden dat deelname aan de interventie onafhankelijk is van de mogelijke uitkomsten. In dat geval kan ervoor gekozen worden om een volledig economisch model uit te werken dat alle relaties tussen de interventie en de uitkomsten specificeert, of om een regressie uit te voeren waarin gecontroleerd wordt voor alle relevante verschillen tussen bedrijven die wel en niet deelgenomen hebben aan de interventie. De veronderstellingen van deze beide aanpakken zijn echter sterk: de modelspecificatie moet correct zijn, inclusief de veronderstelde oorzaken voor deelname aan de interventie. Verder moet de data heel veel informatie over de bedrijven bevatten.

²¹ Experimentele groep wordt ook wel treatment groep genoemd

²² Een poweranalyse geeft een inschatting van de grootte van de steekproef die minimaal nodig is om bij de evaluatie tot statistisch significante uitkomsten te kunnen komen.

Verdiepingsbox 2.1 sociaal experiment

Het vergelijken van gemiddelde uitkomsten in de experimentele en controlegroep met data van een sociaal experiment geeft een zuivere (niet-parametrische) schatting voor het effect van de interventie. De *difference-in-means* schatter is:

$$\Delta = \frac{\sum_{i=1}^n D_i Y_i}{\sum_{i=1}^n D_i} - \frac{\sum_{i=1}^n (1-D_i) Y_i}{\sum_{i=1}^n (1-D_i)}$$

waarbij Y_i de geobserveerde uitkomst is en D_i de onafhankelijke variabele die via loting is toebedeeld aan de experimentele of een controlegroep. Dezelfde schatting van het effect van de interventie kan worden verkregen door de regressie:

$$Y_i = \alpha + \Delta D_i + U_i$$

waarbij U_i storingstermen zijn en α de constante. Omdat D_i onafhankelijk is van de storingsterm U_i , meet de parameter Δ het causale effect van D_i op Y_i . Deze regressie wordt geschat met de kleinste kwadraten methode. Vaak worden ook nog andere geobserveerde kenmerken X_i als controlevariabelen opgenomen in de regressie

$$Y_i = \alpha + \Delta D_i + \beta X_i + U_i$$

De belangrijkste reden voor het opnemen van de andere geobserveerde kenmerken is dat het mogelijk is dat standaardfouten daardoor kleiner worden, omdat de variantie van U_i kleiner wordt. Het toevoegen van andere geobserveerde kenmerken mag niet van invloed zijn op de grootte van het geschatte effect van de interventie. Dit zal in de meeste gevallen nauwelijks invloed hebben op de omvang van de geschatte coëfficiënt.

2.4 Natuurlijke experimenten

Wanneer een sociaal experiment niet mogelijk is, kan met behulp van niet-experimentele data geprobeerd worden het causale effect van een interventie te schatten. We bespreken achtereenvolgens regression discontinuity, difference-in-difference, instrumentele variabelen en propensity score matching.

De methoden die we bespreken zijn gebaseerd op het zoeken naar exogene variatie bij het meedoen aan de interventie. Deze exogene variatie zorgt ervoor dat van twee vergelijkbare bedrijven er slechts één meedoet aan de interventie. Door de uitkomsten van deze bedrijven te vergelijken, kan het causale effect van de interventie geschat worden. De exogene variatie in het meedoen aan een interventie wordt vaak veroorzaakt door bestaande instituties. Hieronder bespreken we een aantal methoden in meer detail.

Regression discontinuity²³

De eerste methode die we bespreken is *regression discontinuity*. Het idee hierachter is dat er een grens is die bepaalt in welke mate bedrijven aanspraak kunnen maken op een interventie. Een voorbeeld is dat bedrijven die een aanvraag doen voor een subsidie een score krijgen voor hun aanvraag. De aanvragen met een score boven bijvoorbeeld een 6 worden gehonoreerd, terwijl de aanvragen met een lagere score worden afgewezen. Het idee van regression discontinuity is om aanvragen die net wel en net niet gehonoreerd zijn met elkaar te vergelijken. Dus neem bijvoorbeeld alle aanvragen met een score binnen een nauw gekozen interval tussen de 5.9 en 6.1. De aanvragen zullen van bedrijven komen die goed vergelijkbaar zijn, met als uitzondering dat degenen die net boven de 6 zaten wel subsidie gehad hebben en degenen die daaronder zaten net niet. Naarmate het interval ruimer wordt genomen, worden de bedrijven minder goed vergelijkbaar en kan het causale effect van de subsidie niet worden vastgesteld.

Regression discontinuity maakt als veronderstelling dat rond het afkappunt de score dusdanig wordt vastgesteld dat bedrijven daarop niet kunnen anticiperen. Bedrijven kunnen dus niet manipuleren of ze net boven of net onder het

²³Zie voor een theoretische discussie over regression discontinuity Hahn, J., P. Todd & W. van der Klaauw (2001), Identification and estimation of treatment effects with a regression-discontinuity design, *Econometrica* 69, 201-209.

afkappunt terecht komen. Er bestaan statistische toetsen die dit onderzoeken. Een eenvoudige check is bijvoorbeeld om te toetsen of bedrijven boven en onder het afkappunt gemiddeld dezelfde geobserveerde kenmerken hebben. Bij een exogeen bepaald afkappunt zouden de gemiddelden van de geobserveerde kenmerken tussen beide groepen vergelijkbaar moeten zijn. Als aan deze veronderstelling voldaan is, dan is er sprake van een lokaal sociaal experiment. En daarom kunnen dezelfde empirische methoden gebruikt worden als bij een sociaal experiment. Het grote verschil is echter de interpretatie van het gevonden causale effect. Omdat er bij een regression discontinuity sprake is van een lokaal sociaal experiment, moet het gevonden causale effect ook lokaal worden geïnterpreteerd. Het geschatte effect is dus het causale effect voor bedrijven die op de marge zitten van wel of geen interventie krijgen. Het grote nadeel van regression discontinuity is dat het dus geen totaal effect geeft van een interventie, in het bijzonder wordt niet geschat wat het effect is voor bedrijven die ver boven het afkappunt zitten en dus sowieso meedoen aan de interventie.

Een meer praktisch probleem is de keuze van de bandbreedte die bepaalt of bedrijven wel of niet meegenomen worden in de analyse. In het bovenstaande voorbeeld werd gesproken over bedrijven die voor hun aanvraag een score tussen de 5.9 en de 6.1 hadden gekregen. Het zou echter zo kunnen zijn dat er maar weinig bedrijven zijn die hieronder vallen. In dat geval wordt het (lokale) effect van de interventie geschat met een kleine steekproef, wat een grotere onzekerheid geeft. Het vergroten van het interval zorgt ervoor dat er meer bedrijven meegenomen worden in de empirische analyse, waardoor het effect van de interventie preciezer geschat wordt. Het risico hiervan is echter dat er ook andere (niet-geobserveerde) verschillen ontstaan tussen de groep bedrijven die heeft meegedaan aan de interventie en de groep die daaraan niet heeft meegedaan. In dat geval wordt het effect van de interventie onzuiver geschat. Vaak wordt er gebruik gemaakt van regressietechnieken om hiervoor enigszins te corrigeren (zie bijvoorbeeld Lalive, 2008).²⁴

Difference-in-difference

De tweede methode die we bespreken is *difference-in-difference*. Dit is een paneldatamethode, wat betekent dat bedrijven over de tijd worden gevolgd. Difference-in-difference is een nuttige methode als gedurende de periode dat uitkomsten van bedrijven worden geobserveerd een aantal bedrijven een interventie krijgt. Het idee van difference-in-difference is dat de trendmatige ontwikkeling – de tijdstrend - in uitkomsten van de bedrijven die de interventie wel en niet krijgen ongeveer gelijk is; dit heet de *common trend* aanname. De verandering in uitkomsten van de bedrijven met de interventie over tijd wordt vergeleken met de verandering in uitkomsten van bedrijven zonder de interventie. Als de common trend aanname geldig is, dan meet dit het causale effect van de interventie.

Neem als voorbeeld een subsidie aan een aantal bedrijven die gedurende een tijdsinterval rond de interventie gevolgd worden: een aantal perioden voorafgaand aan de subsidie en een aantal perioden nadat de subsidie is verstrekt. Omdat het een paneldatamodel betreft, kan er voor elk bedrijf in de steekproef een dummyvariabele worden opgenomen. Daarnaast wordt de doelvariabele geresseerd op een tijdstrend, die dus de trendmatige groei over de gehele periode registreert. Verder bevat het model een indicatorvariabele voor de gesubsidieerde bedrijven vanaf het moment dat ze de subsidie hebben ontvangen. Deze variabele registreert het causale effect van de subsidie op de doelvariabele.

Verdiepingsbox 2.2 difference-in-difference

Het difference-in-difference model is een standaard lineair panel data model

$$Y_{it} = \alpha_i + \delta D_{it} + \beta X_{it} + \mu_t + U_{it}$$

waarbij α_i de bedrijfsspecifieke effecten zijn en de dummyvariabelen μ_t de tijdstrend modelleren. De variabelen U_{it} beschrijven de storingstermen. In de modelspecificatie kunnen ook andere verklarende variabelen X_{it} van bedrijf i op tijdstip t meegenomen worden, wat de precisie van het geschatte causale effect van de interventie δ kan verbeteren. Daarnaast is het mogelijk om in plaats van een dummyvariabele voor wel of niet meedoen aan de interventie ook een continue variabele voor de experimentele groep op te nemen. In dat geval wordt in de specificatie niet meegenomen of een bedrijf heeft meegedaan aan de interventie, maar bijvoorbeeld welk bedrag het aan subsidie ontvangen heeft. De interpretatie van δ wordt dan hoe de uitkomst verandert bij een extra euro subsidie, wat soms ook wel *bang for the buck* genoemd wordt.²⁵

²⁴Lalive, R. (2008), How do Extended Benefits Affect Unemployment Duration? A Regression Discontinuity Approach, *Journal of Econometrics* 142, 785-806.

²⁵Zie voor toepassingen van difference-in-difference bijvoorbeeld Card, D. & A.B. Krueger (1994), Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania, *American Economic Review* 84, 772-793. Duflou, E. (2001), Schooling and Labor Market

Het voordeel van difference-in-difference is dat het causale effect van de interventie wordt gemeten voor de gehele populatie van bedrijven die meegedaan hebben aan de interventie. De common trend aanname is echter in sommige gevallen een te sterke veronderstelling. Deze veronderstelling sluit bijvoorbeeld uit dat bedrijven die het in de afgelopen periode erg goed gedaan hebben, in aanmerking komen voor een interventie. Het zoeken is dus naar een controlegroep van bedrijven die een vergelijkbare trend hebben als de bedrijven die een interventie gehad hebben. De geldigheid van de common trend aanname wordt vaak beargumenteerd door te kijken naar de trend in uitkomsten in de perioden voorafgaande aan de interventie.

Instrumentele variabelen

De volgende methode die we bespreken is *instrumentele variabelen*. Bij instrumentele variabelen worden één of meerdere variabelen gezocht die wel van invloed zijn op welke bedrijven de interventie krijgen, maar die geen direct effect hebben op de uitkomsten van dit bedrijf. Het vinden van dit soort variabelen blijkt in de praktijk erg lastig te zijn. In sommige gevallen kunnen er instituties zijn die deze keuze beïnvloeden. Dit is bijvoorbeeld het geval als er een afkappunt is zoals bij regression discontinuity, maar dat dit afkappunt niet met zekerheid bepaalt welke bedrijven een interventie krijgen. Bedrijven net boven het afkappunt hebben een hoge kans op de interventie, maar krijgen die niet allemaal, terwijl een veel lager percentage van de bedrijven net onder het afkappunt de interventie krijgt. Deze situatie wordt ook wel een *fuzzy regression discontinuity* genoemd.

Het is afhankelijk van hoe de instrumentele variabele van invloed is op de kans dat een bedrijf een interventie krijgt, hoe het geschatte causale effect geïnterpreteerd moet worden. Als de instrumentele variabele slechts rond een afkappunt de kans op een interventie beïnvloedt, zoals bij fuzzy regression discontinuity, dan moet het geschatte effect lokaal geïnterpreteerd worden. Als de waarde van instrumentele variabele voor alle bedrijven relevant is, dan kan een meer algemene interpretatie aan het gevonden effect gegeven worden. Bij instrumentele variabelen is het belangrijk dat de instrumentele variabele een voldoende groot effect heeft op het krijgen van de interventie. Als dit effect klein is of erg onnauwkeurig gemeten, dan kan het geschatte causale effect erg onzuiver zijn.

Verdiepingsbox 2.3 instrumentele variabelen

Een instrument is een variabele Z_i die gecorreleerd is met het wel of niet meedoen aan de interventie, en die geen direct effect heeft op de uitkomst. Het instrumentele variabelen model heeft daarom twee vergelijkingen

$$Y_i = \alpha + \delta D_i + \beta X_i + U_i$$

$$D_i = \gamma + \pi Z_i + \rho X_i + V_i$$

Instrumentele variabelen methoden kunnen bij kleine steekproeven een onzuivere schatting van het effect van de interventie geven. Het is hiervoor van belang naar de π te kijken; deze geeft het effect van Z_i op de interventie D_i . De vuistregel is dat de test-statistic van een F -test voor significantie van π groter dan 10 moet zijn. Dit zorgt ervoor dat de instrumenten niet zwak zijn.

Instrumentele variabelen methoden hebben toepassing onder andere bij het schatten van vraag/aanbod-modellen en dynamische paneldatamodelen. Bij het schatten van bijvoorbeeld een vraagvergelijking wordt de prijs geïnstrumenteerd met factoren die wel het aanbod, maar niet de vraag beïnvloeden. Bij dynamische paneldatamodelen

$$Y_{it} = \alpha_i + \gamma Y_{it-1} + \delta D_{it} + \beta X_{it} + \mu_t + U_{it}$$

wordt vaak GMM gebruikt om de parameters te schatten. Het idee is dat vertraagde endogene variabelen Y_{it-2}, Y_{it-3} , etc. gebruikt worden om de parameter γ te kunnen schatten. Een veronderstelling die hierbij gemaakt wordt is dat er geen autocorrelatie in de storingstermen U_{it} zit. Het risico van GMM is dat er te veel vertraagde variabelen als instrument opgenomen worden, waardoor de instrumenten zwak worden en de onzuiverheid in de schatters toeneemt. GMM is er in allerlei variaties, o.a. de system estimator.²⁶

Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment, *American Economic Review* 91, 795-813.

²⁶Verbeek, M. (2012), *A Guide to Modern Econometrics*, fourth edition, Wiley.

Propensity score matching

De volgende methode die we bespreken is *propensity score matching*. Dit is eigenlijk een verzamelnaam voor technieken waarbij verondersteld wordt dat alle achtergrondvariabelen die relevant zijn voor het wel of niet krijgen van de interventie en die ook van invloed zijn op de uitkomsten van het bedrijf, geobserveerd worden.²⁷ Bij propensity score matching wordt gezocht naar paren van twee bedrijven die vergelijkbaar zijn in achtergrondvariabelen, maar waarvan er slechts één meegedaan heeft aan de interventie. Door de uitkomsten van deze twee identieke bedrijven met elkaar te vergelijken wordt het causale effect van de interventie geschat. Aan de hand van de propensity score wordt bepaald hoe identiek bedrijven zijn. De propensity score geeft de kans dat een bedrijf een interventie krijgt, gegeven de achtergrondkenmerken van dat bedrijf. Een voorbeeld van toepassing van propensity score is een subsidie aan een bedrijf die gebaseerd is op een aantal criteria. Elk van de bedrijven krijgt voor deze criteria een score. Allereerst wordt de kans geschat dat een bedrijf de subsidie krijgt, gegeven de scores op de verschillende criteria. Deze kans wordt aangeduid als de propensity score. Vervolgens kunnen verschillende methoden worden toegepast om het effect van de propensity score op de doelvariabele te bepalen. Door bedrijven met een vergelijkbare propensity score, waarvan de één wel de subsidie heeft gekregen en de andere niet, aan elkaar te koppelen, ontstaan vanzelf twee vergelijkbare groepen.

Verdiepingsbox 2.4 Propensity score matching

De propensity score p_i is de kans dat gegeven X_i bedrijf i meedoet aan de interventie, dus

$$p_i = Pr(D_i = 1 | X_i)$$

Veronderstel dat er N_1 bedrijven meegedaan hebben aan de interventie en N_0 bedrijven niet. Een algemene specificatie voor de schatter voor het effect van de interventie is:

$$\Delta = \frac{1}{N_1} \sum_{i=1}^{N_1} \left(Y_i - \sum_{j=1}^{N_0} w(i, j) Y_j \right)$$

Voor elke N_1 observaties in de experimentele groep wordt een counterfactual gekozen die tot stand komt door de N_0 observaties in de controlegroep te wege. De gewichten $w(i, j)$ geven aan hoe vergelijkbaar observatie i in de experimentele groep is met observatie j in de controlegroep. Deze gewichten worden bepaald met behulp van de geschatte propensity scores en hiervoor zijn verschillende algoritmes. Als een observatie in de experimentele groep wordt gekoppeld aan de observatie in de controlegroep met de dichtstbijzijnde propensity score is sprake van nearest-neighbour matching. Als deze gekoppeld wordt aan de gemiddelde uitkomst van alle observaties in de controlegroep met een ongeveer vergelijkbare propensity score is sprake van neighborhood matching. Als observaties op intervallen van de propensity score met elkaar vergeleken worden heet dit blocking. En het is ook mogelijk om elke observatie in de experimentele en controlegroep te wege met de inverse van de propensity score (weighting).

Er zijn allerlei manieren om gebruik te maken van propensity score. Met behulp van de propensity score kunnen twee identieke bedrijven bepaald worden die aan elkaar gekoppeld worden, zoals hierboven beschreven. Echter aan bedrijven die een interventie krijgen kunnen ook meerdere bedrijven zonder interventie gekoppeld worden, of omgekeerd. Er bestaan ook methoden waarbij elke observatie gewogen wordt met behulp van de propensity score, zodat geen observaties verloren gaan.

Het voordeel van propensity score matching is dat het effect van de interventie geschat wordt voor alle bedrijven waarvoor counterfactuals bestaan. Bedrijven met achtergrondkenmerken die ervoor zorgen dat ze sowieso een interventie krijgen, worden buiten beschouwing gelaten.

²⁷Zie voor een algemene beschrijving: Becker, S.O. & A. Ichino (2002), Estimation of Average Treatment Effects Based on Propensity Scores, *Stata Journal* 4, 358-353. Caliendo, M. and S. Kopeinig (2008), Some Practical Guidance for the Implementation of Propensity Score Matching, *Journal of Economic Surveys* 22, 31-72.

Nadeel van propensity score matching is dat er impliciet verondersteld wordt dat toekenning van een interventie willekeurig is voor bedrijven met dezelfde propensity score. Dit nadeel kan minder belangrijk zijn als er veel achtergrondinformatie beschikbaar is, maar de veronderstelling dat alle relevante achtergrondkenmerken geobserveerd worden kan niet worden getoetst. De onderzoeker zal dus op basis van institutionele kennis moeten beargumenteren dat dit het geval is.

2.5 Standaard regressieanalyse

De methoden die hierboven besproken zijn proberen allemaal het probleem van selectiviteit bij de inzet van een interventie te ondervangen door te zoeken naar vergelijkbare bedrijven met uitzondering van het wel of niet meedoen aan de interventie. In praktijk is het niet altijd mogelijk om één van de bovenstaande methoden daadwerkelijk toe te passen. Vaak wordt in dat geval gebruikt gemaakt van een standaard lineair regressiemodel.

Bij een regressieanalyse onderscheiden we bedrijven die wel of niet hebben meegedaan aan de interventie. De uitkomsten van beide groepen kunnen worden vergeleken aan de hand van een regressievergelijking, daarbij controlerend voor zoveel als mogelijk observeerbare factoren inclusief een indicatorvariabele voor het bedrijf.

Verdiepingsbox 2.5 Regressieanalyse

Het standaard lineair regressiemodel heeft de vorm:

$$Y_i = \alpha + \delta D_i + \beta X_i + U_i$$

De coëfficiënt δ geeft de correlatie tussen de interventie D_i en de uitkomst Y_i (conditioneel op de andere geobserveerde kenmerken X_i). Vaak is een (conditionele) correlatie geen causaal effect. Om de correlatie wel causaal te kunnen interpreteren moet gelden dat het meedoen aan de interventie D_i niet gecorreleerd is met niet-geobserveerde kenmerken U_i . In een niet-experimentele context betekent dat vaak dat veel geobserveerde kenmerken X_i moeten worden opgenomen in de regressieanalyse. Echter het opnemen van veel geobserveerde kenmerken heeft ook het risico dat zogenaamde confounding (“verwarrende”) variabelen opgenomen worden. Dit zijn variabelen die beïnvloed worden door de interventie en ook een effect hebben op de uitkomst. Stel bijvoorbeeld dat een interventie tot doel heeft de arbeidsproductiviteit in een bedrijf te verhogen. Als we het effect van de interventie op winst willen schatten en arbeidsproductiviteit als verklarende variabele meenemen, dan wordt in de regressie het causale effect van de interventie op winst onderschat.

Hierbij moet gewaakt worden voor de volgende problemen. Ten eerste kan er sprake zijn van zelfselectie van bedrijven. Vaak zijn de gegevens afkomstig uit een situatie waarin bedrijven zelf kiezen of ze meedoen aan een interventie. Bijvoorbeeld alle bedrijven kunnen aanspraak maken op een subsidie, maar niet elk bedrijf kiest ervoor om dat te doen. Ten tweede doet zich een probleem voor wanneer een subsidieverstrekker uit een aantal aanvragen de beste selecteert om te honoreren. In dat geval geldt dat het meedoen van de interventie niet onafhankelijk is van de mogelijke uitkomsten. Het vergelijken van bedrijven die meedoen aan de interventie en niet meedoen geeft in veel gevallen een overschatting van het effect van de interventie. Bij een overschatting is de gemeten waarde van het effect groter dan het feitelijke causale effect.

Een imperfecte oplossing van selectiviteit is om een volledig economisch model uit te werken dat alle relaties tussen de interventie en de uitkomsten specificiert, of om een regressie uit te voeren waarin gecontroleerd wordt voor alle relevante verschillen tussen bedrijven die wel en niet deelgenomen hebben aan de interventie. De veronderstellingen van deze beide aanpakken zijn echter sterk. De modelspecificatie moet correct zijn en de data heel veel informatie bevatten.

De bovenstaande bespreking van regressieanalyse gaat er impliciet van uit dat gebruik wordt gemaakt van een crosssectie van bedrijven. Vaak zijn er echter gegevens beschikbaar over meerdere tijdsperioden. In dat geval zijn dus paneldata beschikbaar. Bij panel data kan worden gecorrigeerd voor bepaalde vormen van niet-geobserveerde bedrijfskenmerken. Het gaat dan om bedrijfskenmerken die niet variëren over tijd (zie verdiepingsbox 2.3).

Naast de standaard fixed-effect paneldatamethoden, die ook gebruikt worden om het difference-in-difference model te schatten, is het ook mogelijk om meer dynamische modellen te analyseren. Modellen zijn dynamisch als uitkomsten in eerdere perioden de huidige uitkomst van het bedrijf beïnvloeden (zogenaamde vertraagde endogene variabelen). Zo'n dynamisch paneldatamodel kan ook gebruikt worden om geaggregeerde uitkomsten voor een bepaalde sector over tijd te modelleren.

Door de aanwezigheid van vertraagde endogene variabelen is het schatten van dynamische paneldatamodelen lastiger dan het schatten van een standaard paneldatamodel. Het probleem is dat de vertraagde endogene variabelen die aan de rechtekant van de vergelijking staan gecorreleerd kunnen zijn met de storingstermen. Om dit probleem op te lossen wordt vaak gebruik gemaakt van instrumentele variabelen. De onderliggende veronderstelling is dat storingstermen over tijd niet gecorreleerd zijn. In dat geval kunnen verder vertraagde endogene regressoren gebruikt worden als instrumentele variabelen. Het aantal mogelijke instrumentele variabelen is in dat geval erg groot en daarom wordt vaak gebruik gemaakt van GMM als schattingsmethode. GMM moet beschouwd worden als een generalisatie van instrumentele variabele methoden, maar waarbij veel instrumenten gebruikt kunnen worden. Zoals ook genoemd in de diepingsbox 2.3 is het grote risico dat er te veel (zwakke) instrumenten opgenomen worden, waardoor schatters onzuiver worden. Binnen GMM zijn er allerlei variaties zoals first-difference GMM²⁸ en system GMM²⁹.

2.6 Surveymethoden

In het verleden is bij de evaluatie van bedrijvenbeleid vaak gebruik gemaakt van enquêtes onder bedrijven. Er wordt dan zowel een groep bedrijven ondervraagd die heeft deelgenomen aan een interventie als een groep bedrijven die niet heeft meegedaan. Bedrijven wordt dan direct gevraagd naar het effect van de interventie, maar ook naar een aantal kenmerken en uitkomsten van het bedrijf. Deze enquêtes zijn vaak erg uitgebreid en kunnen potentieel een volledig beeld geven van effecten van de interventie.

Bij surveymethoden worden echter ook een aantal veronderstellingen gemaakt. De eerste veronderstelling is dat antwoorden van bedrijven bij de enquêtes waarheidsgetrouw zijn en dat er dus geen systematische foutrapportage is. Denk bijvoorbeeld aan het risico op sociaal wenselijk of strategisch antwoorden. Bijkomend probleem is dat bij enquêtes onder bedrijven het responspercentage vaak niet heel hoog is. De tweede veronderstelling is dan ook dat respons niet selectief verschilt tussen bedrijven die wel en niet meegedaan hebben aan de interventie. Dit speelt wanneer, ten gevolge van verschillen in de responspercentages, bepaalde groepen onder- of oververtegenwoordigd zijn in het onderzoek. Gedraagt een ondervertegenwoordigde groep zich duidelijk anders met betrekking tot de te onderzoeken variabelen dan de oververtegenwoordigde groep, dan leidt dit tot een vertekening in de uitkomsten. Een derde veronderstelling is dat in de survey bedrijven die wel en niet meegedaan hebben aan de interventie vergelijkbaar zijn. Bij het analyseren van gegevens van een survey worden vaak verschillen in gemiddelden bepaald. Er wordt dan meestal niet gecorrigeerd voor verschillen in achtergrondkenmerken tussen bedrijven. Dat maakt het des te belangrijker dat de groepen van bedrijven die wel en niet meegedaan hebben aan de interventie vergelijkbaar zijn. Dit kan gedaan worden door bij de keuze welke bedrijven te enquêteren ervoor te zorgen dat in beide groepen vergelijkbare bedrijven bekeken worden. Ten slotte wordt ook bij een surveymethode verondersteld dat er geen directe spillovers van de interventie tussen bedrijven zijn. Dit is dus dezelfde veronderstelling die expliciet besproken is bij de sociale experimenten en ook bij alle andere bovenstaand besproken methoden noodzakelijk is. Ondanks deze veronderstellingen kan de surveymethode vaak wel belangrijke additionele informatie opleveren over de precieze werking van het instrument en de ervaringen van de gebruikers. Het is daarom vaak een belangrijke aanvullende methode naast de hierboven geschetste statistische methodes.

²⁸ Arellano, M. and S. Bond (1991), Some tests of specification for panel data: monte carlo evidence and an application to employment equations, *Review of Economic Studies* 58, 277-297.

²⁹ Blundell, R. and S. Bond (1998), Initial conditions and moment restrictions in dynamic panel data models, *Journal of Econometrics* 87, 115-143.

2.7 Conclusies

Hierboven hebben we een aantal methoden besproken voor empirische beleidsevaluatie. Het grote probleem bij de evaluatie is dat interventies vaak selectief ingezet worden, wat de empirische analyse lastig maakt. In zulke gevallen geeft een lineaire regressie vaak een onzuivere schatting van het causale effect van de interventie.

Een sociaal experiment met willekeurige toewijzing geeft de mogelijkheid het causale effect van een interventie het meest zuiver te schatten. Het geschatte effect heeft in dit geval betrekking op de gehele interventie. In de regel is de opzet van een sociaal experiment beperkt in omvang, wat gevolgen kan hebben voor de precisie waarmee het effect van de interventie geschat wordt. Loting bij toekenning van interventies wordt echter nauwelijks gebruikt voor de evaluatie van innovatiebeleid. In gevallen waar geen sociaal experiment voorhanden is, zijn er een aantal andere methoden die gebruikt kunnen worden. Deze methoden verschillen in het effect dat ze schatten, en de consistentie en precisie van de geschatte effecten.

Regression discontinuity lijkt op een sociaal experiment, maar schat slechts lokaal een causaal effect. Difference-in-difference geeft een algemener beeld en maakt gebruik van paneldata waardoor ook rekening gehouden kan worden met sommige niet-geobserveerde bedrijfskenmerken, maar er moet wel aan de common trend aanname zijn voldaan. Het vinden van instrumentele variabelen is vaak lastig. Propensity score matching is als methode vrijwel altijd mogelijk en geeft vaak ook een volledig beeld van een interventie. Het grote probleem van propensity score matching is dat de onderliggende veronderstelling erg sterk is en nauwkeurige data vereist. Aan surveys liggen ook veronderstellingen ten grondslag, bijvoorbeeld over selectiviteit. Een aanvullend risico bij surveys is dat de gegevens vaak gebaseerd zijn op de perceptie van ondervraagden in plaats van objectieve, administratieve gegevens. Door middel van surveys kan wel veel informatie op tafel komen over de werking en uitvoering van het instrument die complementair is aan de resultaten van de hierboven genoemde kwantitatieve methoden.

Hoofdstuk 3 Reflecties op WBSO Evaluatie: Een casestudie

3.1 Inleiding

Het DG B&I van het ministerie van EZ heeft een ruime ervaring met het laten uitvoeren van beleidsevaluaties. Bij deze evaluaties wordt veel gebruikt gemaakt van kwalitatieve onderzoeksmethodes, zoals de surveymethode. Hiernaast heeft het DG B&I ook reeds de nodige ervaring opgedaan met evaluaties waarbij gebruik wordt gemaakt van meer kwantitatieve methodes. Een goed voorbeeld hiervan zijn de evaluaties van de in 1994 aangenomen Wet Bevordering Speur- en Ontwikkelingswerk (WBSO). In 2001-2002 is WBSO door PWC, Dialogic en de TU Delft geëvalueerd, waarbij de effecten voor de eerste keer zijn gekwantificeerd met behulp van lineaire regressie. Bij een eerdere evaluatie uit 1998 lukte dit nog niet vanwege databeperkingen. In 2005 heeft het CPB naar aanpassingen in de opzet van de WBSO-regeling gekeken en gekozen voor regression discontinuity. In 2006-2007 is door EIM in samenwerking met UNU-MERIT een evaluatie uitgevoerd van de WBSO over de periode 2001-2005, waarbij een fixed effects paneldatamodel is gebruikt. De meest recente evaluatie van de WBSO over de periode 2006-2010, uitgevoerd door EIM maakt gebruik van een dynamisch paneldatamodel dat geschat is met behulp van GMM.³⁰ De Expertwerkgroep Effectmeting is gevraagd de meeste recente WBSO-evaluatie te beoordelen en te kijken waar verbetering mogelijk is om de effectiviteit in kaart te kunnen brengen.

3.2 De WBSO

Doelstelling

De Wet Bevordering Speur- en Ontwikkelingswerk (WBSO) heeft als doelstelling om onderzoek en ontwikkeling door Nederlandse bedrijven en ondernemers te stimuleren door de loonkosten van onderzoekend personeel te verlagen door middel van een fiscale tegemoetkoming.

Marktfalen

Bij de motivatie voor de invoering van de WBSO in de jaren negentig wordt niet verwezen naar marktfalen; in deze jaren was de focus op marktfalen nog relatief onbekend. Er werd daarentegen verwezen naar de grote onzekerheid die samengaat met investeringen in onderzoek en ontwikkeling en de omvangrijke investeringen in apparatuur en laboratoria die nodig zijn. Hierdoor zouden de private R&D-uitgaven achterblijven. Verder werden de loonkosten in Nederland als relatief hoog gezien, met een mogelijk negatieve invloed op het vestigingsklimaat voor (R&D-intensieve) bedrijven in ons land. Een onderliggend doel van de WBSO is dan ook om een aantrekkelijk fiscaal klimaat te hebben voor buitenlandse bedrijven om hier hun R&D-activiteiten te vestigen.

Bij de ontwikkeling van kennis en technologie kunnen positieve externe effecten optreden; de maatschappelijke baten zijn groter dan de private baten. Het bestaan van positieve externe effecten leidt tot een onderinvestering in R&D vanuit maatschappelijk oogpunt bekeken. Een fiscale tegemoetkoming voor de loonkosten van R&D-personeel (een loonkostensubsidie) is dan een mogelijk instrument om innovatieve investeringen goedkoper te maken en te stimuleren (zie ook in hoofdstuk 4 de bespreking van de Research en Development Aftrek (paragraaf 4.3)).

Werking

De WBSO is de verzamelnaam voor een tweetal fiscale faciliteiten ter bevordering van speur- en ontwikkelingswerk (S&O):

- S&O-afdrachtvermindering loonbelasting:
 - Voor inhoudingsplichtigen die een onderneming drijven met werknemers die R&D verrichten (bedrijven met personeel);
 - Voor inhoudingsplichtigen die geen onderneming drijven, indien zij voor rekening van een (samenwerkingsverband van) onderneming(en) of een product of bedrijfschap R&D verrichten (kennisinstellingen met personeel).
- S&O-aftrek inkomstenbelasting: Voor zelfstandigen die een onderneming drijven, indien zij meer dan een vastgesteld aantal uren aan R&D besteden

³⁰ EIM, 2012, Evaluatie WBSO 2006-2010, Zoetermeer.

De afdrachtvermindering loonbelasting voor inhoudingsplichtigen bestaat uit een tweedelig schijvensysteem. In 2010 gold een afdrachtvermindering van 50% voor de eerste € 220.000 aan gerealiseerde loonkosten van onderzoekend personeel en van 18% over het restant. Voor belastingplichtigen (zelfstandigen) gold een vaste aftrek van € 12.031 op het belastbaar inkomen in de inkomstenbelasting wanneer zij minimaal 500 R&D-uren maken op jaarbasis. In 2012 zijn het percentage in de eerste en tweede schijf verlaagd naar respectievelijk 42% en 14%. De grens van het plafond van de eerste schijf is verlaagd naar € 110.000.

Het voornemen voor 2013 is om de bovengrens van de eerste schijf weer te verhogen van 110.000 naar 200.000 euro. Om dat te financieren gaat het percentage van de eerste schijf van 42% naar 38% en het percentage van de eerste schijf voor starters van 60% naar 50%

Startende ondernemers worden extra gestimuleerd R&D te ontplooiën. Voor inhoudingsplichtige starters gold een percentage van 64% over het bedrag in de eerste schijf. Dit percentage is inmiddels verlaagd naar 60%. Voor belastingplichtige starters gold een aanvullende aftrek van € 6.017 op hun belastbaar inkomen.

Om te voorkomen dat grote bedrijven zich een onevenredig groot deel van het budget toe-eigenen, geldt een maximum van R&D-loonkosten waarover het fiscale voordeel van toepassing is. Het plafond binnen de WBSO is gelijk aan €14 miljoen per kalenderjaar.

Uitvoering

Ondernemers en kennisinstellingen kunnen maximaal driemaal per jaar een digitale aanvraag indienen bij NL Innovatie, minimaal een maand voorafgaand aan de periode waarin de werkzaamheden worden uitgevoerd. De aanvraag dient minimaal drie en maximaal zes maanden te betreffen. In bepaalde gevallen is een jaaraanvraag mogelijk. Voor zelfstandigen geldt geen maximum aan het aantal aanvragen op jaarbasis, maar zij kunnen hun aanvraag uiterlijk op de laatste dag van september indienen.

Het proces dat door een aanvrager doorlopen wordt, kan in een achttal stappen onderverdeeld worden:

1. de aanvrager wil iets ontwikkelen of onderzoeken
2. indienen aanvraag en indien nodig aanlevering BSN's
3. inrichten S&O-administratie
4. controle volledigheid aanvraag door NL Innovatie
5. inhoudelijke beoordeling aanvraag door NL Innovatie
6. verrekening financieel voordeel in aangifte door aanvrager
7. indien nodig doorgeven gerealiseerde S&O-uren aan NL Innovatie
8. mogelijk bezoek of controle aanvrager door NL Innovatie

Budget

Het budget voor de WBSO bedraagt in 2012 872 miljoen euro. Sinds 1994 wordt het budget van de WBSO jaarlijks vastgesteld in het Belastingplan, de realisatie kan afwijken van het budget.

Evaluatie

In 2011-2012 is de WBSO over de periode 2006-2010 geëvalueerd. De eerstvolgende evaluatie zal in 2016 plaatsvinden.

3.3 De evaluatie van de WBSO 2006-2010

Hypotheses

Kernhypothese³¹

- 1) De WBSO heeft een positief effect op de totale R&D-loonuitgaven van een onderneming.

Hypothese ter ondersteuning kernhypothese

- 2) Bedrijven die relatief veel subsidie krijgen geven relatief meer uit aan R&D-loon dan bedrijven die een lager subsidiepercentage hebben.

Hypotheses voor toetsing uitstralingseffecten

- 3) De WBSO zorgt voor meer innovatie op bedrijfsniveau.
- 4) De WBSO leidt tot spillovers in de eigen sector.
- 5) De WBSO leidt tot spillovers buiten de eigen sector.
- 6) De WBSO zorgt voor een hogere toegevoegde waarde per werkende.

In de onderstaande bespreking kijken we primair naar de eerste hypothese en de ondersteunende kernhypothese. Dit betekent dat alleen wordt gekeken naar het basismodel dat gebruikt is in de WBSO evaluatie 2006-2010.

Econometrische specificatie en identificatie

Om het effect van de WBSO te meten, heeft EIM gekeken naar de relatie tussen (de logaritme van) de loonsom van het R&D-personeel dat bedrijf i in dienst heeft (hierna R&D-loonsom) en de WBSO-afdrachtvermindering die het bedrijf ontvangt gedeeld door de loonsom (het WBSO-kortingspercentage). Deze methode is de standaard regressieanalyse die beschreven staat in Sectie 2.5 van dit rapport. Omdat de R&D-loonsom (sterk) gecorreleerd is over de tijd, is er een vertraagde endogene variabele opgenomen in de regressievergelijking. Het model meet dus de relatie tussen veranderingen in de loonsom en het subsidiepercentage. Het EIM verwacht een positieve relatie tussen (de groei van) de loonsom en het subsidiepercentage; als de WBSO effectief is, geven bedrijven die relatief veel subsidie krijgen meer uit aan onderzoek dan bedrijven die een lager subsidiepercentage hebben (zie EIM 2012, blz. 12 van de bijlage). Dit model test dus de bovenstaande hypothese ter ondersteuning van de kernhypothese.

Bij het schatten van het model heeft het EIM verder rekening gehouden met een aantal controlevariabelen en met (niet-geobserveerde) verschillen tussen bedrijven voor zover deze constant zijn over de tijd. Dit zou ervoor moeten zorgen dat het geschatte effect causaal geïnterpreteerd kan worden. Het EIM heeft de volgende specificatie gebruikt:

$$\ln Y_{it} = \alpha_i + \gamma \ln Y_{it-1} + \delta \frac{WBSO_{it-1}}{Y_{it-1}} + \beta X_i + \mu_t + U_{it}$$

Hierin is Y de totale R&D-loonsom in een bedrijf in een jaar, α_i een fixed effect per bedrijf, $WBSO$ het bedrag dat vanwege de WBSO minder aan belasting hoeft te worden betaald, X een reeks aan controlevariabelen en μ_t een fixed effect voor de jaren. U is de storingsterm.

Deze specificatie verschilt van de specificatie die in de voorgaande evaluatie van de WBSO is gebruikt bij het schatten van het effect van de WBSO op de R&D-loonsom. Toen werd de WBSO als verklarende factor niet in verhouding tot de R&D-loonsom uitgedrukt, maar als (logaritme van een) absoluut bedrag (EIM en UNU-MERIT 2007).³² Daarnaast werd toen geen vertraagde endogene opgenomen, omdat dat in combinatie met fixed effects vertekende resultaten zou geven.

³¹ In de evaluatie van EIM wordt niet gewerkt met hypotheses, maar met onderzoeksvragen. Om de consistentie met de rest van dit rapport te bewaken, zijn de onderzoeksvragen vertaald in hypotheses.

³² In de voorgaande evaluatie van de WBSO is ook een andere aanpak gevolgd, waarbij het effect van de WBSO is afgeleid van de correlatie tussen de loonsom van onderzoekers en de gebruikerskosten van speur- en ontwikkelingswerk.

De reden hiervoor is dat de vertraagde endogene variabele gecorreleerd is met storingstermen, wat bij het schatten complicaties geeft (zie de discussie in de volgende paragraaf).

Het model suggereert een relatief eenvoudig verband tussen het subsidiepercentage en de loonsom. De hypothese dat de WBSO een positief effect heeft op R&D-loonuitgaven zou moeten worden geaccepteerd als δ significant groter is dan nul. Er zijn drie mogelijke valkuilen. Ten eerste bepaalt de omvang van de loonsom het subsidiepercentage: hoe hoger de loonsom, hoe lager het marginale WBSO-tarief. De vertraagde endogene variabele $\ln Y_{it-1}$ corrigeert voor de hoogte van de loonsom, maar er wordt hier wel een lineaire functionele vorm opgelegd die misschien te restrictief is.

De tweede valkuil is dat bedrijven waaraan de WBSO een effectieve stimulans biedt, vanwege de groei van hun onderzoeksactiviteiten vaker in een hogere WBSO-schijf terecht zullen komen. Voor dit selectie-effect wordt in principe gecorrigeerd door het de fixed-effects, maar dit hangt op de veronderstelling dat deze constant zijn over de tijd.

De derde valkuil is het dynamische karakter van de specificatie. De hypothese is door EIM geformuleerd met betrekking tot niveaus. De geschatte coëfficiënt is echter beter vergelijkbaar met de relatie tussen de *verandering* van de loonsom en het subsidiepercentage. Dit kan problematisch zijn als het groeipotentieel van een bedrijf afneemt naarmate een bedrijf groter wordt. Bij welke bedrijfsomvang de gemiddelde groei het grootst is, is een empirische vraag. Het rapport van het EIM bevat aanwijzingen dat bedrijven met een snelgroeiende loonsom meer voorkomen in de hogere schijven van de WBSO (zie bijlage III van EIM 2012, p. 57). Welk deel van de loonsomgroei door de WBSO komt, is op deze manier niet vast te stellen.

Voor een generieke maatregel zoals de WBSO is het lastig om deze problemen met niet-experimentele data op te lossen. Zoals hierboven beschreven is het zonder goede controle groep noodzakelijk om extra veronderstellingen te maken over de modelspecificatie.

De basisspecificatie van EIM is niet eenvoudig om te schatten omdat het fixed effect ook van invloed is op de vertraagde endogene variabele. EIM heeft verschillende dynamische paneldata methoden gebruikt om voor dit probleem te corrigeren. Deze methoden maken gebruik van instrumentele variabelen, veelal vertraagde afhankelijke variabelen en de eerste verschillen daarvan. In principe kan met behulp van instrumentele variabelen ook getracht worden om selectie-effecten en andere vormen van endogeniteit weg te nemen. Of dit ook lukt, hangt sterk af van welke instrumentele variabelen voorhanden zijn. In het EIM rapport wordt niet ingegaan op de bovengenoemde problemen. Daarnaast geldt dat bij het schatten van dynamische paneldatamodellen nog extra veronderstellingen moeten worden gemaakt. Hieronder bespreken we hoe EIM deze econometrische technieken heeft gebruikt bij het schatten van de basisspecificatie.

Schattingmethoden

Het gebruikte dynamische paneldatamodel bevat aan de rechterkant een vertraagde endogene, wat impliceert dat de standaard fixed effects schattingsmethoden niet gebruikt kunnen worden. Het is in dit soort gevallen gebruikelijk om eerste verschillen te nemen, dus:

$$\ln Y_{it} - \ln Y_{it-1} = \gamma (\ln Y_{it-1} - \ln Y_{it-2}) + \delta \left(\frac{WBSO_{it-1}}{Y_{it-1}} - \frac{WBSO_{it-2}}{Y_{it-2}} \right) + \beta (X_{it} - X_{it-1}) + \mu_t - \mu_{t-1} + (U_{it} - U_{it-1})$$

Het grote probleem bij het schatten van dit soort modellen is de correlatie tussen de regressor ($\ln Y_{it-1} - \ln Y_{it-2}$) en de storingsterm ($U_{it} - U_{it-1}$). Hetzelfde zou kunnen gelden voor de regressor ($\frac{WBSO_{it-1}}{Y_{it-1}} - \frac{WBSO_{it-2}}{Y_{it-2}}$) en dat zou op eenzelfde manier behandeld moeten worden. De gebruikelijke aanpak is om instrumenten te zoeken dat zijn variabelen die gecorreleerd zijn met de endogene regressoren, maar ongecorrleerd met de storingsterm. Bij dynamische paneldatamodellen worden de instrumenten vaak gezocht in verder vertraagde endogene variabelen. Dus mogelijke instrumenten zijn $\ln Y_{it-2}$, $\ln Y_{it-3}$, etc. Als er slechts gebruik gemaakt wordt van één vertraagde variabele als instrument, in dit geval $\ln Y_{it-2}$, dan is sprake van de Anderson-Hsiao schatter. Als er meer instrumentele variabelen dan regressoren zijn, dan is GMM een efficiënte methode om de parameters te schatten. Deze methode wordt in de WBSO evaluatie GMM-dif genoemd.

Deze GMM-aanpak maakt twee onderliggende veronderstellingen. De eerste is dat de instrumenten $\ln Y_{it-2}$, $\ln Y_{it-3}$, etc. exogeen zijn, wat betekent dat ze niet gecorreleerd zijn met de storingsterm ($U_{it} - U_{it-1}$). Dit is het geval als er geen

autocorrelatie zit in de storingstermen U_{it} . Deze veronderstelling kan getoetst worden nadat het model geschat is. De tweede veronderstelling is dat de instrumenten relevant zijn, wat wil zeggen dat ze gecorreleerd zijn met de endogene regressor ($\ln Y_{it-1} - \ln Y_{it-2}$). Deze correlatie moet voldoende groot zijn om ervoor te zorgen dat er geen al te grote onzuiverheid in de schatters komt door het zogenaamde *weak instruments* probleem. Er zijn een aantal redenen waarom instrumenten *weak* zouden kunnen zijn. De eerste is als γ dicht bij 1 ligt en de tweede reden is dat de fixed effects α_i een relatief grote invloed op de uitkomsten hebben ten opzichte van de storingstermen U_{it} . Het probleem van *weak instruments* wordt verergerd door het opnemen van veel instrumentele variabelen. Er zijn allerlei testen zoals F -testen voor de instrumenten in de *first-stage* regressie en de Gragg-Donald test die kunnen aangeven of instrumenten voldoende sterk zijn.

Als γ gelijk aan 1 is, dan is het dynamische paneldatamodel niet stationair en wordt het schatten van het model met de bovenstaande procedure lastig omdat de instrumentele variabelen irrelevant worden. In dit geval wordt vaak gebruikt gemaakt van de system GMM schattingsmethode. Binnen deze methode wordt geen gebruik gemaakt van het model in eerste verschillen, maar wordt het model in niveaus gebruikt en zijn juist de instrumenten in eerste verschillen gespecificeerd. Naar deze methode wordt in de WBSO evaluatie gerefereerd als GMM-sys.

In de WBSO-evaluatie wordt zowel bij GMM-dif als GMM-sys het maximaal aantal instrumenten gebruikt dat beschikbaar is. Het opnemen van veel instrumenten verhoogt het risico dat er onzuiverheid in de schattingen optreedt door het *weak instruments* probleem. Er kan echter niet worden beoordeeld of dit hier ook een probleem is, omdat de gebruikelijke toetsen voor *weak instruments* niet gerapporteerd worden in het rapport. Daarnaast worden er ook geen toetsen gerapporteerd over de autocorrelatie in de storingstermen.

De schattingsresultaten met de Anderson-Hsiao methode liggen dicht bij die van GMM-dif. Dat is bemoedigend en impliceert een bepaalde mate van robuustheid van de resultaten. De GMM-sys schattingsresultaten liggen echter relatief ver van de resultaten van de andere twee schattingsmethoden. Dat zou kunnen duiden op mogelijke problemen met de GMM-sys schatter, maar hiervoor worden geen toetsen gerapporteerd. In het rapport wordt waarschijnlijk de juiste keuze gemaakt om bij de interpretatie van de schattingsresultaten uit te gaan van de schattingsresultaten die zijn verkregen met GMM-dif en Anderson-Hsiao. Maar zonder de bovengenoemde specificatietoetsen blijft de validiteit van de resultaten moeilijk te beoordelen.

In de volgende paragraaf geven we een economische interpretatie van de gevonden resultaten. We veronderstellen daarbij dat het effect van de WBSO op R&D-loonuitgaven correct geschat is, hoewel daar onzekerheid over blijft bestaan.

Economische interpretatie van de gemeten effecten

De belangrijkste variabele in het EIM rapport om de bijdrage van de WBSO aan onderzoek en ontwikkeling in kaart te brengen, is de 'bang for the buck' (BFTB) maatstaf. Deze meet de extra R&D-loonkostenuitgaven³³ die worden veroorzaakt door een euro belastingvermindering.³⁴ In het EIM-rapport wordt hiermee het eerste-orde-effect van de WBSO in kaart gebracht. Omdat de BFTB ook bij andere evaluaties is berekend, kunnen verschillende schattingsmethoden met de BFTB eenvoudig vergeleken worden.

De gemiddelde BFTB wordt voor de periode 2006-2010 geschat tussen de 1,55 en 1,99 euro, met als meest waarschijnlijke schatting 1,77. Voor de periode 2001-2005 werd de gemiddelde BFTB geschat op 1,99. Het verschil kan te wijten zijn aan zowel de verbeterde econometrische aanpak als een sterk verbeterd databestand, maar ook aan een gevolg van de 'wet van de afnemende meeropbrengsten', waarbij het marginale effect afneemt naarmate de regeling ruimer wordt.

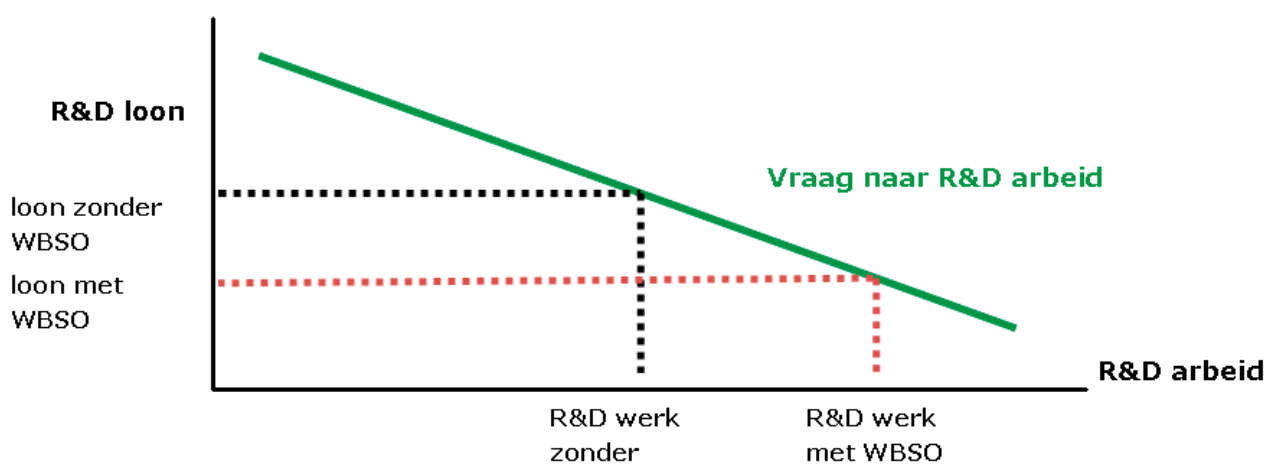
Hoe moet de gemiddelde BFTB worden geïnterpreteerd? Het rapport (p. 62) geeft een voorbeeldberekening voor het jaar 2007. In dat jaar waren de totale bruto R&D-loonkosten afgerond € 3,4 mld. Daarvan wordt 60% of circa € 2 mld. ondersteund door WBSO. In 2007 ontvingen bedrijven vanwege WBSO in totaal € 0,46 mld. aan belastingvoordeel. Het gaat hier om de door AgNL toegekende S&O-afrachtvermindering, oftewel een korting op de loonbelasting. Gemiddeld is dit belastingvoordeel dus 23 procent (0,46 gedeeld door 2,0 mld.). Met een gemiddelde BFTB van 1,77 betekent dit volgens het rapport een totaal extra R&D-loon van € 0,46 x 1,77 = € 0,8 mld.

³³ In haar rapport heeft EIM het in dit verband onterecht over het effect op extra R&D-uitgaven. Dat is een ruimer begrip dan alleen de extra R&D-loonkostenuitgaven. In schattingsmodel (Bijlagenrapport, pag. 11, vergelijking 1) is de parameter die ten grondslag licht aan de BFTB variabelen echter gerelateerd aan de R&D loonkosten (variabele *rdw*) en niet aan de totale R&D-kosten.

³⁴ De formule hiervoor is te vinden in Bijlage 2 van de econometrische bijlage bij het hoofd rapport van de WBSO-evaluatie 2006-2010.

Vervolgens wordt in het rapport gesteld dat het verschil van € 1,2 mld. tussen het totaal WBSO loon (€ 2 mld.) en het extra R&D loon als gevolg van de WBSO (€ 0,8 mld.) gelijk is aan de deadweight loss, gedefinieerd als het 'door WBSO ondersteund loon dat ook zonder WBSO zou zijn uitgevoerd'. Dit is vanuit de economische theorie gezien een bijzondere interpretatie van het begrip 'deadweight loss'. In de economische literatuur wordt dit begrip geassocieerd met het verlies aan productie ten gevolge van een beperking van het aanbod door een monopolist of een bedrijf met monopoliekracht. Wat soms ook –bij extensie – wordt omschreven als 'deadweight' loss is het subsidiebedrag dat wordt toegekend voor activiteiten die ook zonder de subsidie zouden hebben plaatsgevonden. In deze meer extensieve betekenis is 'deadweight loss' de tegenpool van additionaliteit. De in het rapport gebruikte betekenis van 'deadweight loss' is verwarrend. Er zou misschien beter gesproken kunnen worden over een 'windfall gain'.

Figuur 2: Relatie vraag naar R&D-aanbod en loon



In wezen is de WBSO een loonkostensubsidie, die weliswaar niet wordt berekend als een percentage van het uurloon, maar via een vermindering van de loonsom uiteindelijk wel leidt tot een lager uurloon voor R&D-arbeid. In Figuur 2 wordt de vraag naar R&D-arbeid getekend als een dalende functie van het uurloon. Door de WBSO-afdrachtvermindering daalt het uurloon van 'loon zonder WBSO' naar 'loon met WBSO' en dat heeft tot gevolg dat de vraag naar WBSO arbeid stijgt van 'werkgelegenheid zonder WBSO' naar 'werkgelegenheid met WBSO'.

Achter de schatting van de BFTB in het rapport zit een schatting van de vraagelasticiteit naar R&D arbeid. In het rapport (p. 66) wordt gesproken over het verband tussen elasticiteit en de BFTB. De elasticiteit die daarin wordt beschreven is echter niet de vraagelasticiteit van arbeid. De elasticiteit die in het rapport wordt beschreven, meet het relatieve effect van een toename van 1 procent in de WBSO-afdrachtvermindering op de R&D-loonsom. Deze toename is 0,38 procent. Stel dat we de toename van 1% van de WBSO-afdrachtvermindering interpreteren als een loonvoetkorting van 0,23%.³⁵ Dan zou dit tot gevolg hebben dat de R&D-loonsom met 0,38 procent toeneemt. De procentuele toename van de R&D-loonsom (+0,38% in dit geval) is bij een gegeven R&D-loon (vóór aftrek van WBSO-afdrachtvermindering) gelijk aan de procentuele toename van de R&D werkgelegenheid. Indien deze interpretatie klopt, dan zou de impliciete schatting van de vraagelasticiteit die achter de BFTB schuil gaat gelijk zijn aan 1,65 (= 0,38/0,23) en dat is in de empirische literatuur een relatief hoge elasticiteit, zeker omdat het hier om gespecialiseerde werknemers gaat.

Tweede- en derde-orde effecten

In de evaluatie tracht men ook hypothesen drie tot en met zes te toetsen door de effecten van de WBSO op innovatie (2^e orde effect) en de effecten op de bedrijfsprestaties te meten (3^e orde effect), zowel binnen als buiten de eigen sector. Deze effecten probeert men met behulp van elasticiteiten (2^e-orde effect) en een productiefunctiemodel (3^e-orde effect) te identificeren. Het is echter lastig hier een conclusie aan te verbinden aangezien de problemen bij het meten van de hierboven besproken eerste-orde-effecten nog prominenter zijn bij hogere orde effecten. Voor wat betreft de interpretatie van deze effecten dient men dus nog voorzichtiger te zijn.

³⁵ We volgen hiermee de aanname van EIM dat het aanbod van onderzoekers volledig elastisch is. Deze veronderstelling is op de korte termijn niet erg realistisch voor R&D-medewerkers.

3.4 Conclusie

Bij de evaluatie van de WBSO is ervoor gekozen om de kwantitatieve analyse een prominente rol te geven. Echter het evalueren van een generieke maatregel waarvan op grote schaal gebruik wordt gemaakt, is niet eenvoudig. Het gekozen dynamische paneldatamodel wordt vaker gebruikt bij de evaluatie van dit soort maatregelen, maar heeft zeker beperkingen. Omdat onduidelijk is waar de variatie in gebruik van WBSO vandaan komt, moeten in het model een aantal veronderstellingen gemaakt worden, zowel bij de specificatie als bij het schatten. De empirische analyse houdt rekening met een aantal lastige econometrische problemen. Voor een echt goede beoordeling van de analyse ontbreken een aantal specificatietoetsen. Het causaal interpreteren van het gevonden effect blijft problematisch omdat het niet voor de hand ligt dat de instrumentele variabelen correleren met het WBSO-kortingspercentage en tegelijkertijd onafhankelijk zijn van de R&D-loonsom.

Om in de toekomst meer over de effectiviteit van de WBSO te kunnen zeggen beveelt de werkgroep het volgende aan:

- Besteed in een toekomstige evaluatie meer aandacht aan discontinuïteiten. In de laatste evaluatie is er ook gebruik gemaakt van discontinuïteiten, maar omdat er toen veel wijzigingen tegelijkertijd doorgevoerd zijn, was het moeilijk om de verschillende effecten uit elkaar te trekken. In de toekomst zou met behulp van een difference-in-difference analyse het effect van de schijfverlenging naar 200.000 euro, die in 2013 ingaat, kunnen worden geanalyseerd. Gegeven het feit dat in het regeerakkoord “Bruggen slaan” een schuif wordt voorgesteld van fiscale innovatieregelingen naar de TKI-toeslag, zal dit in de toekomst tot verdere wijzigingen kunnen leiden. Indien hier handig mee wordt omgegaan, kan dit interessante evaluatiemogelijkheden bieden. Een plafondverhoging wordt voor een difference-in-difference analyse bijvoorbeeld niet opportuun geacht, aangezien het aantal observaties te beperkt zal zijn. In de onderhavige evaluatie is ook naar het effect van veranderingen in het schijvenstelsel gekeken. Hier speelde echter het probleem dat de schijfverruiming vooral als crisismaatregel genomen zijn, waardoor de difference-in-difference geen robuuste resultaten oplevert.
- Toekomstige evaluaties zouden verder aan kracht kunnen winnen wanneer duidelijker de relatie wordt gelegd tussen de resultaten van de surveystudie en de econometrische resultaten.
- Om te voorkomen dat het effect van de WBSO wordt overschat, is het belangrijk om goed te weten of de bedrijven die een S&O-verklaring hebben ook daadwerkelijk aan S&O doen. Voor aanvragen vanaf januari 2013 zijn WBSO-aanvragers verplicht (in lijn met de RDA) om na afloop van het jaar te melden hoeveel S&O-uren daadwerkelijk zijn uitgevoerd. Nu is dat alleen noodzakelijk wanneer een WBSO-aanvrager minder dan 90% van de oorspronkelijke S&O-toekenning heeft gerealiseerd. Bij een toekomstige evaluatie zal hierop dus beter zicht zijn.
- Voor het tweede-orde-effect zou naar de verandering van de schijven op innovatieve activiteiten kunnen worden gekeken.

Hoofdstuk 4: Evaluatieontwerpen instrumenten

4.1 Inleiding

Dit hoofdstuk vormt de kern van dit rapport. In dit hoofdstuk adviseert de expertwerkgroep over verschillende evaluatieontwerpen voor elk van de zes aangewezen instrumenten. Elk evaluatieontwerp wordt daarbij op zijn merites beoordeeld. Voordat de evaluatieontwerpen die bij een instrument horen besproken worden, gaan we eerst in op het doel, de legitimatie op basis van markt- of overheidsfalen, de werking, uitvoering en het budget van het instrument. De selectie bevat instrumenten met uiteenlopende budgetten en bestaat zowel uit traditionele generieke instrumenten (Innovatiekrediet, Research & Development Aftrek, Innovatieprestatiecontracten), nieuwe meer vraaggestuurde instrumenten (Centra voor Innovatief Vakmanschap en Topconsortia voor Kennis en Innovatie) als een meer 'indirect' beleidsinstrument (Bewijs van Goede Diensten-gemeenten). In dit hoofdstuk wordt dus een breed spectrum aan beleidsopties behandeld, zodat over de gehele linie een beeld ontstaat over wat men per type instrument van effectmeting kan verwachten, en vooral ook wat men niet kan verwachten. De beschrijving van evaluatiemethodes gehanteerd in dit hoofdstuk grijpt terug naar hoofdstuk 2, waar de evaluatietechnieken uitvoerig zijn beschreven. Waar in de vorige hoofdstukken zoveel mogelijk is geprobeerd de lezer bij de hand te nemen, zal dit hoofdstuk echter wat meer technische kennis vereisen en dieper ingaan op de econometrische materie. Desalniettemin is getracht de intuïtie van de evaluatieontwerpen zo helder mogelijk neer te zetten. De econometrische vergelijkingen betreffen een uitwerking van deze intuïtie en kunnen de welingevoerde lezer wellicht intellectueel wat meer uitdagen. Voor toekomstige evaluaties bieden deze uitwerkingen bovendien handvatten om op voort te borduren.

4.2 Innovatiekrediet (IK)

Doelstelling

Het innovatiekrediet (IK) beoogt private R&D-uitgaven gericht op technische en/of klinisch risicovolle ontwikkelingsprojecten voor nieuwe producten, diensten en processen te verhogen.³⁶

Marktfalen

Vanuit de economische welvaartstheorie beredeneerd zijn er twee redenen waarom de overheid innovatiekrediet wil verstrekken. Ten eerste kan het maatschappelijk rendement van een innovatieproject hoger liggen dan het private rendement. Een belangrijke reden waarom het maatschappelijk rendement hoger kan liggen dan het private rendement zijn kennisspillovers. Ten tweede zou een innovatiekrediet kapitaalmarktfalen kunnen tegengaan als de overheid beter geïnformeerd is over een innovatieproject dan private investeerders. Dat kan als redenen hebben dat ondernemers bereid zijn om meer informatie te delen met AgentschapNL (AgNL) dan met private investeerders³⁷ en dat AgNL meer specifieke expertise op het terrein van technologie en innovatie heeft dan private investeerders. AgNL kan hierdoor mogelijk een betere inschatting van de risico's van een specifiek innovatieproject maken.

Werking

De markt voor de financiering van innovatieprojecten kan op basis van risico en rendement³⁸ worden onderverdeeld in deelsegmenten. Aan de ene kant zijn er projecten met een hoog risico en een hoog rendement. Deze projecten vinden hun financiering op de risicokapitaalmarkt (venture capital). Projecten met een laag risico en met een laag rendement worden vooral bancair gefinancierd. Het innovatiekrediet richt zich op het middensegment. Dit zijn projecten waarvoor het risico voor banken te hoog is, maar het rendement voor venture capitalists (VC) te laag. Daardoor is het vinden van private financiering voor deze projecten lastig.

In tegenstelling tot andere beleidsinstrumenten past het innovatiekrediet bij technisch succes het profijtbeginsel toe. Gemiddeld wordt 70-80% van de kredieten terugbetaald. Het rendement op de successen is te laag om het verlies van de mislukte projecten te dekken. Dit komt omdat 1) het innovatiekrediet een vergoeding vraagt voor een lening die dicht bij

³⁶ Het gaat om projecten die zich bevinden in de experimentele fase en nieuw zijn voor Nederland.

³⁷ In de economische literatuur wordt dit het hold-up probleem genoemd.

³⁸ Met rendement wordt hier altijd verwacht rendement bedoeld.

marktrentes uit het lagere risicosegment (bancaire kredieten) ligt, dit terwijl het risico hoger is, en 2) de overheid niet, zoals bij eigen vermogen, meedeelt in de winst van succesvolle projecten, maar wel voor de verliezen opdraait voor het kredietdeel bij technisch mislukken van het project. Het impliciete subsidie-element van het innovatiekrediet komt ten voordele van de VC of de ondernemer.

Het innovatiekrediet wordt verstrekt met een rente die gekoppeld is aan Euribor. Als basis geldt 4% voor technische projecten van bedrijven met voldoende cashflow. Dit wordt verhoogd naar 7% wanneer de cashflow onzeker is. Voor klinische projecten wordt 7% rente in rekening gebracht, met een verhoging naar 10% wanneer de cashflow onzeker is. Via het innovatiekrediet is medefinanciering mogelijk tot maximaal 35% van de subsidiabele kosten van bovengenoemde ontwikkelingsprojecten. Een aanvrager van innovatiekrediet moet zicht hebben op financiering en zal dus ook bij andere, private financiers langsmoeten.

Uitvoering

De adviseurs/experts binnen het team Kredieten van AgNL beoordelen kredietaanvragen aan de hand van 6 criteria: ondernemerschap/onderneming, business case/verdienpotentieel, nieuwigheid/innovativiteit, technisch risico, projectplan en financiering. Per criterium wordt een rating gegeven, variërend van – tot ++. Hierna wordt de kredietcommissie (= interne commissie met een externe ondernemer) om advies gevraagd. Dit is een onafhankelijk oordeel en geldt als second opinion. Deze commissie geeft een genuanceerd advies over de aanvraag; met redenen onderbouwd. In ongeveer 20% van de gevallen zijn de 'bureaucommissie' AgNL en de kredietcommissie het oneens met elkaar. Het advies van de kredietcommissie wordt niet in 100% van de gevallen opgevolgd. Uiteindelijk neemt de manager Innovatiekrediet van AgNL, namens de Minister, een ja/nee-beslissing. De aanvraagtijd van een innovatiekrediet duurt gemiddeld 3 à 4 maanden (maximaal 4 maanden); dit is niet langer dan bij andere, private financiers. Bij meer dan 50% van de innovatiekredietaanvragen staat het project gelijk aan het gehele bedrijf.

De beoordeling wordt niet beïnvloed door de hoogte van het gevraagde krediet. Het gemiddelde bedrag van een aanvraag is € 1,5-2 miljoen. Er komen ongeveer 80 aanvragen per jaar binnen. Hiervan worden er 20 ingetrokken door de aanvrager, na feedback op de plannen door de adviseurs bij AgNL. Van de resterende 60 krijgen er uiteindelijk 30 een innovatiekrediet toegekend. Vanaf 2012 is het budget bijna verdubbeld, het aantal aanvragen zal waarschijnlijk toenemen. Enerzijds door de beleidswijziging waarbij subsidies zijn geschrapt en deels zijn omgezet in fiscale maatregelen en kredieten. Daarnaast is de doelgroep van het innovatiekrediet per 1-1-2012 verbreed met het grootbedrijf.

Budget

Voor het innovatiekrediet is in 2012 € 95 miljoen beschikbaar. Dit is inclusief de uitvoeringskosten die in 2012 ongeveer € 3,7 miljoen bedragen. Budget dat niet wordt benut vloeit terug naar het fonds en gaat zo door naar volgende begrotingsjaren. Hierdoor is er een lagere kans op voortijdige sluiting van het instrument wegens budgetgebrek zonder het nadeel dat er teveel wordt uitgegeven zoals bij een open eind-regeling. Het fonds kan zo flexibel meebewegen in omvang waar noodzakelijk en binnen de totale budgettaire ruimte van het Innovatiefonds MKB+. Inkomsten, zoals aflossingen van verleende kredieten, vloeien ook terug naar dit fonds en kunnen zo worden gebruikt voor het uitzetten van nieuwe financieringen. Deze onzekere inkomsten worden niet geraamd in de Rijksbegroting, maar dragen wel bij aan de revolverendheid van het fonds en de onderliggende pijlers zoals het Innovatiekrediet.

Looptijd evaluatieperiode

Het innovatiekrediet wordt in 2012-2013 geëvalueerd. Voor deze evaluatie heeft de werkgroep een voorlopig advies uitgebracht. Recentelijk is een onderzoeksbureau in de arm genomen om de evaluatie uit te voeren.

Hypotheses

Kernhypothese

- 1) Door het ontvangen van een innovatiekrediet besteden bedrijven meer aan R&D.

Hypothese ter ondersteuning kernhypothese

- 2) Projecten die gesteund worden met het innovatiekrediet komen sneller tot stand dan afgewezen projecten.

Hypothese voor toetsing uitstralingseffecten aanvraagprocedure

- 3) Het AgNL-aanvraagproces verhoogt de kans op kredieten van private financiers voor bedrijven die zijn afgewezen voor het innovatiekrediet.

Hypothese voor toetsing maatschappelijke baten

- 4) De maatschappelijke baten van toegewezen innovatiekredietaanvragen zijn groter dan de maatschappelijke kosten van het impliciete subsidie-element van de regeling.

Indicatoren

Om de bovenstaande hypothesen te toetsen moet een aantal indicatoren gebruikt worden:

- Wel of niet doorgaan van het project, inclusief de scores op de onderliggende criteria en het advies van de kredietcommissie. Het innovatiekrediet zou zich moeten richten op projecten die anders niet uitgevoerd worden. Daarnaast zijn invloeden op het tempo en de grootte van het project belangrijk (hypothese 1 en 2).
- Cofinanciering; ter beantwoording van de vraag: welke invloed heeft het al dan niet verstrekken van het innovatiekrediet gehad op cofinanciering door andere investeerders? (hypothese 3)
- Succes van het project, zowel technisch als commercieel. Het wel of niet technisch slagen van het project wordt gedurende de periode van verstrekking van het innovatiekrediet bekend. De terugbetaling van het krediet is daarvan afhankelijk. De mate van technisch succes kan in sommige gevallen op kortere termijn gemeten worden aan de uitkomst van het nieuwheidsrapport (search report)³⁹ van een octrooibureau en op lange termijn de geëffectueerde patenten. Verkoop en doorverkoop van patenten zegt iets over commercieel succes. Technisch succes kan ook worden afgeleid uit de beoordeling van AgNL of het krediet moet worden terugbetaald. Dat is een veel directere meting. (hypothese 4)
- Commercieel succes kan gemeten worden met groei/krimp van de bedrijven in termen van omzet, investeringen, winst, vervolgprijzen en personeelsgroei. Eventueel kan ook nog worden gekeken naar internationalisering van het bedrijf of een maatstaf voor marktaandeel. (hypothese 4)
- Er kan apart naar spillovers gekeken worden, in de eigen sector en in andere sectoren. Het is de vraag in hoeverre dat goed meetbaar is voor de projecten die met innovatiekrediet zijn ondersteund. Resultaten voor spillovereffecten zijn gemakkelijker te verkrijgen indien de effecten van de totale R&D-uitgaven in de eigen sector en andere sectoren op individuele bedrijfsprestaties worden onderzocht (conform benadering die gevolgd is bij de WBSO-evaluatie). (hypothese 4)

Een groot deel van de benodigde gegevens voor een evaluatie is beschikbaar bij AgNL. Zij hebben gegevens over alle bedrijven die ooit een innovatiekrediet aangevraagd hebben. Omdat het innovatiekrediet betrekking heeft op specifieke projecten, wordt in eerste instantie gekeken naar de projecten waarvoor innovatiekrediet aangevraagd is. De groep bedrijven die succesvol een innovatiekrediet aangevraagd heeft, kan relatief makkelijk gevolgd worden. Uit projectinformatie van AgNL is beschikbaar of het project technisch geslaagd is en in welke mate het krediet wordt terugbetaald. Daarnaast kan aan deze bedrijven eenvoudig informatie gevraagd worden over de omvang van de R&D-uitgaven, het economisch succes van de ondersteunde projecten, de financiering door andere investeerders en de economische prestaties van het bedrijf in het algemeen.

Ook bedrijven bij wie een aanvraag voor innovatiekrediet is afgewezen, kunnen bevroegd worden. Bijvoorbeeld om een antwoord te krijgen op vraag of eventueel alternatieve financiering is verkregen. Het is de vraag in hoeverre het haalbaar is om over deze bedrijven voldoende gegevens beschikbaar te krijgen. In het geval van afgewezen aanvragen waar het project gelijkstaat aan de onderneming, kan WBSO-data wellicht worden gebruikt om te bezien of het toch gelukt is om financiering te krijgen⁴⁰. Als een bedrijf in de WBSO-data voorkomt dan kan verondersteld worden dat het bedrijf erin geslaagd is om de financiering rond te krijgen. Langs dezelfde lijn kunnen de WBSO-data wellicht ook gebruikt worden om niet-startende bedrijven met een afgewezen innovatiekredietaanvraag te volgen, zolang de aanvraag "groot" is ten opzichte van de totale omvang van het bedrijf. Als een innovatiekrediet het huidige balanstotaal van de aanvragende onderneming met ten minste 100% zou verhogen, dan is het een redelijke veronderstelling dat de prestaties van de gehele onderneming nauw samenhangen met het slagen van het innovatieproject. Voor grote bedrijven kan dit niet

³⁹Verlening van een Europees patent kan jaren duren. Het 'search report' geeft al in een eerder stadium aan of een aanvraag kans van slagen heeft.

⁴⁰Mits men WBSO krijgt voor hetzelfde project, van dezelfde omvang. Daarbij dient ook rekening te worden gehouden met het feit dat circa 70% van de aanvragers van het Innovatiekrediet terug te vinden is in de WBSO-database.

verondersteld worden, het slagen van een innovatieproject binnen een groot bedrijf kan niet worden afgelezen aan de jaarcijfers. Het idee is dus dat WBSO-data een indicatie geven van de groei van de R&D uitgaven van een bedrijf. Zonder groei na een afgewezen aanvraag is de veronderstelling dat er geen alternatieve geldschieter is gevonden.

Voor diverse gegevens over bedrijven (waaronder omzet en aantal werknemers) kan gebruik worden gemaakt van statistische informatie van het CBS. Hiervoor is een koppeling nodig van AgNL-data aan CBS-data (ABR) en aan gegevens van de Belastingdienst. De koppeling verloopt via KvK-nummers. Een beperking daarbij is dat gegevens over kleinere bedrijven uit CBS enquêtes zeer onvolledig zijn. Voor inzicht in de R&D-uitgaven van vooral kleinere bedrijven vormen WBSO-data van AgNL een belangrijke bron in aanvulling op de R&D-data van het CBS. WBSO-data zijn in principe dekkend voor alle bedrijven die substantieel aan R&D doen (inclusief bedrijven met minder dan 10 werkzame personen, die niet worden waargenomen met de R&D-enquête van het CBS). Deze gegevens kunnen ook nodig zijn bij het toetsen van Hypothese 4. Voor een antwoord op de vraag of dankzij het innovatiekrediet projecten *sneller* tot stand komen, bieden CBS- en WBSO-data geen oplossing. Voor de beantwoording van die vraag blijft een aanvullende survey waarschijnlijk noodzakelijk. Octrooigegevens zijn beschikbaar via een KvK-koppeling met gegevens van het NL Octrooiencentrum.

Tabel 2 Overzicht indicatoren bij evaluatieontwerp IK

Hypotheses	Indicatoren		Databronnen
	Afhankelijk	Onafhankelijk	
1) Door het ontvangen van een innovatiekrediet besteden bedrijven meer aan R&D.	<ul style="list-style-type: none"> • Wel/niet doorgaan van project • Grootte van het project 	<ul style="list-style-type: none"> • Dummy voor toekennen innovatiekrediet 	<ul style="list-style-type: none"> • AgNL • CBS
2) Projecten die gesteund worden met het innovatiekrediet komen sneller tot stand dan afgewezen projecten.	<ul style="list-style-type: none"> • Tempo van het project 	<ul style="list-style-type: none"> • Dummy voor toekennen innovatiekrediet 	<ul style="list-style-type: none"> • AgNL • Enquête
3) Het AgNL-aanvraagproces verhoogt de kans waarmee bedrijven die zijn afgewezen voor het innovatiekrediet, een krediet kunnen aanvragen bij private financiers.	<ul style="list-style-type: none"> • Beschikbare cofinanciering 	<ul style="list-style-type: none"> • Dummy voor toekennen innovatiekrediet 	<ul style="list-style-type: none"> • AgNL • CBS
4) De maatschappelijke baten van toegewezen innovatiekredietaanvragen zijn groter dan de maatschappelijke kosten van het impliciete subsidie-element van de regeling.	<ul style="list-style-type: none"> • AgNL beoordeling over technisch succes • Terugbetaling krediet • Octrooien en patenten • Groei van bedrijven in termen van omzet, investingen, vervolgprojecten, personeel, internationalisering, etc. • R&D binnen de sector om spillovers te meten. 	<ul style="list-style-type: none"> • Dummy voor toekennen innovatiekrediet • Hoogte toegekend innovatiekrediet 	<ul style="list-style-type: none"> • AgNL • WBSO-data • NL octrooiencentrum • CBS
Controlevariabelen [niet uitputtend]	<ul style="list-style-type: none"> • Scores op beoordelingscriteria van AgNL en advies kredietcommissie • Projectinformatie • Bedrijfskenmerken 		<ul style="list-style-type: none"> • AgNL • CBS

Evaluatieontwerp

Om goed het primaire effect van het innovatiekrediet aan te tonen (hypothese 1) is het belangrijk om vergelijkbare groepen te onderscheiden waarbij de ene groep het innovatiekrediet niet heeft ontvangen en de andere groep wel. Idealiter zouden de vergelijkbare groepen het gevolg zijn van gerandomiseerde toewijzing van het innovatiekrediet aan bedrijven die hiervoor in aanmerking komen (zie de discussie in hoofdstuk 2). Omdat bij het innovatiekrediet geen gerandomiseerd experiment heeft plaatsgevonden, moet gebruik worden gemaakt van andere evaluatiemethoden (zie hoofdstuk 2).

Optie 1: Propensity score matching

Experimentele en controlegroep

Bedrijven die een innovatiekrediet aanvragen kunnen in drie groepen worden onderverdeeld: zij die geen enkele kans maken op een innovatiekrediet, zij die zonder twijfel een innovatiekrediet zouden moeten krijgen en het zogenaamde grijze gebied waar op basis van de beoordelingsprocedure de uitkomst zowel positief als negatief kan uitpakken. Omdat innovatiekredieten niet willekeurig worden toegewezen, kan een eenvoudige vergelijking van gehonoreerde en afgewezen aanvragen niet gebruikt worden om de hypothesen te toetsen.

De hypothesen (met uitzondering van de derde hypothese) hebben betrekking op alle toegewezen innovatiekredietaanvragen en de evaluatie zou zich moeten richten op bedrijven die zonder twijfel een krediet krijgen en het grijze gebied. Het evaluatieprobleem is dat voor aanvragers die met zekerheid gehonoreerd worden geen counterfactuals bestaan, waardoor voor die groep een empirische evaluatie op basis van een vergelijking met een controlegroep theoretisch onmogelijk is. De vergelijkbare groepen moeten uit het grijze gebied komen, maar bij de interpretatie moet rekening gehouden worden met het feit dat niet alle verstrekte innovatiekredieten meegenomen worden. Als de zekerheid waarmee een aanvraag gehonoreerd wordt gerelateerd is aan de (positieve) effecten van het innovatiekrediet, dan levert een vergelijking in het grijze gebied een onderschatting van de daadwerkelijke effectiviteit. Dat betekent dat als in het grijze gebied bewijs wordt gevonden voor een bepaalde hypothese, deze hypothese waarschijnlijk ook geldig is met betrekking tot alle verstrekte innovatiekredieten.

Identificatiestrategie

Er is veel informatie bekend over het proces van honorering van aanvragen. AgNL beoordeelt aanvragen op zes criteria en ook de kredietcommissie geeft advies. Daarnaast zijn ook andere kenmerken van de aanvraag en de aanvrager bekend. Identificatie van het effect van een innovatiekrediet vindt plaats in twee stappen. Op basis van de geobserveerde kenmerken wordt eerst de kans op honorering van elke aanvraag bepaald (de zogenaamde propensity score).⁴¹ In stap twee wordt het effect van het innovatiekrediet bepaald door toegekende projecten te vergelijken met afgewezen projecten die (ongeveer) dezelfde kans op toekenning hebben.

Econometrische specificatie en methoden

Nadat de 'propensity score' is bepaald voor elke observatie, zijn er verschillende methoden beschikbaar om het effect van het innovatiekrediet op de doelvariabelen vast te stellen (zie hoofdstuk 2). Het werken met gewichten heeft wellicht de voorkeur boven het één-op-één matchen van observaties met en zonder innovatiekrediet, omdat het tot het minste verlies van data leidt. Het gewicht van een observatie is groter naarmate het bedrijf uit de controlegroep beter vergelijkbaar is met het bedrijf wat een innovatiekrediet ontvangen heeft. De beperkte steekproefomvang is waarschijnlijk het grootste struikelblok in de empirische evaluatie. Omdat een groot deel van de innovatiekredieten naar start-ups gaan, is de leeftijd van het bedrijf een belangrijke variabele om op te matchen. De methoden kunnen met behulp van het softwarepakket Stata worden toegepast. Een korte uitleg van de techniek staat in Box 2.4. De gewichten worden bepaald door de doelvariabele te wegen met de inverse van de geschatte propensity score (kans op toewijzing van gehonoreerde aanvragen en kans op afwijzing voor de niet-gehonoreerde aanvragen).

⁴¹In principe zou de score-sheet van AgNL voldoende moeten zijn, omdat deze alle informatie weegt die relevant is voor de succeskans van de aanvraag. Daarnaast is het waarschijnlijk lastig om op basis van maximaal 80 aanvragen nog meer informatie mee te nemen. Toch is het wel handig om andere kenmerken ook te beschouwen. Je zou je namelijk zorgen moeten gaan maken over de 'propensity score' als andere kenmerken naast de informatie van de scoresheet van AgNL ook nog van invloed zijn op de succeskans van een aanvraag.

Het bovenstaande is alleen mogelijk als het effect van het wel of niet honoreren van een innovatiekrediet de relevante keuze is. Het is lastiger om direct de opbrengsten van een verstrekte euro krediet te schatten, de zogenaamde bang for the buck. In dat geval is niet alleen de selectiviteit bij de toekenning van belang, maar ook de selectiviteit bij de grootte van de aanvraag. Echter er kan wel een geaggregeerd idee verkregen worden omtrent de bang for the buck door het causale effect van een toekenning (op commercieel succes, winst, omzet, etc.) te delen door de gemiddelde grootte van het krediet. Hierbij moet meegenomen worden dat gemiddeld het grootste deel van het krediet terugbetaald wordt, want dit zijn ook opbrengsten van een project.

Gevoeligheden

Het aantal observaties in het grijze gebied dat nodig is voor het uitvoeren van een zinvolle effectmeting is moeilijk op voorhand vast te stellen. Dit hangt om twee redenen af van de variatie in de gegevens. Ten eerste moet er voldoende overlap zijn in de verdeling van propensity scores van gehonoreerde en afgewezen aanvragen. Ofwel, hoe ruimer het grijze gebied, des te preciezer en completer kan de schattingsmethode zijn. Ten tweede maakt het uit hoe risicovol de projecten zijn. Bij risicovolle projecten is een grotere steekproef nodig om een betrouwbare schatting van een gemiddelde uitkomst (bijvoorbeeld voltooiing project, R&D-uitgaven) te krijgen dan bij minder risicovolle projecten. Omdat het aantal aanvragen voor een innovatiekrediet niet groot is en op voorhand niet duidelijk is hoeveel aanvragen in het grijze gebied zullen vallen, is het succes van de boven beschreven evaluatiemethode vooraf onzeker.

Bovenstaand werd al opgemerkt dat de 'propensity score'-methode alleen antwoord geeft op de effecten van het innovatiekrediet in het grijze gebied. Voor de volledigheid moet ook worden gekeken naar projecten die zonder twijfel gehonoreerd worden. Ten eerste moet worden vastgesteld hoe groot deze groep daadwerkelijk is. Mogelijk kan iets worden gezegd over de effectiviteit in dit gebied door de interactie van het effect van het innovatiekrediet met de 'propensity score' te bekijken. Als projecten met een hogere 'propensity score' effectiever zijn, dan ligt het voor de hand dat de 'propensity score'-methode een onderschatting geeft van de totale effectiviteit van het innovatiekrediet. Maar het succes van deze aanpak hangt af van de steekproefomvang en de interpretatie blijft een extrapolatie buiten het observatiegebied. Het dient daarom opgemerkt te worden dat volledige representativiteit alleen behaald kan worden met het opzetten van een experiment met (gedeeltelijk of conditioneel) gerandomiseerde toewijzing.

Optie 2: Panelregressie

Experimentele en controlegroep

Bij deze optie wordt de correlatie tussen het toekennen van het innovatiekrediet en latere uitkomsten geschat. Er worden drie groepen bedrijven onderscheiden: bedrijven waaraan een innovatiekrediet is toegekend, bedrijven waarvan een aanvraag voor een innovatiekrediet is afgewezen, en bedrijven die geen aanvraag hebben gedaan.

Identificatiestrategie

Omdat het CBS beschikking heeft over paneldata kan er een panelregressie gedaan worden met als onafhankelijke variabele zowel een dummy voor bedrijven met een aanvraag als een dummy voor bedrijven waaraan een innovatiekrediet is toegekend. Hierbij moet wel worden opgemerkt dat aanvraag of toekenning van het innovatiekrediet waarschijnlijk niet voldoet aan de voorwaarde van strikte exogeniteit. Bedrijven die een innovatiekrediet aanvragen doen dat omdat ze in een fase van ontwikkeling en/of groei zitten.

Econometrische specificatie en methoden

Eenvoudige 'fixed effects'-paneldatatechnieken kunnen een vertekend beeld geven omdat uitval uit het panel waarschijnlijk selectief is. Daarom is het raadzaam om te kijken naar het moment van invoering van het innovatiekrediet en na te denken over een soort 'difference-in-difference'-model. Onduidelijk is wat een goede controlegroep is die dezelfde tijdstrend heeft. De volgende difference-in-difference vergelijking zou kunnen worden geschat:

$$Y_{it} = \beta_0 + \beta_1 IK_{it} + \beta_2 A_{it} + \beta_3 X_{it} + \eta_i + \theta_t + U_{it}$$

In deze vergelijking geeft Y aan hoeveel bedrijf i aan R&D uitgeeft, zijn de β 's de te schatten coëfficiënten, is IK een indicator die aangeeft of het bedrijf een innovatiekrediet heeft ontvangen, is A een indicator die aangeeft of een bedrijf een aanvraag gedaan heeft en is X een vector met controlevariabelen. Verder dient er gecorrigeerd te worden voor

ongeobserveerde eigenschappen van bedrijven en tijdseffecten (“fixed effects” voor bedrijven, η , en tijd, θ), U is de storingsterm.

Gevoeligheden

Het grootste probleem van deze aanpak is dat aan correlaties niet altijd een causale interpretatie kan worden gegeven. Dit kan slechts als gecorrigeerd wordt voor alle selectiviteit in het proces van toekenning van het innovatiekrediet (zie hoofdstuk 2). Deze selectiviteit kan ondervangen worden door in het regressiemodel te corrigeren voor heterogeniteit, al moet worden uitgekeken met het toevoegen van zogenaamde confounders. Confounders zijn variabelen die een deel van het effect waarin de onderzoeker geïnteresseerd is, absorberen. Verder is de steekproefomvang beperkt waardoor het grijze gebied te weinig observaties kan bevatten voor een degelijke evaluatie.

Optie 3: Surveymethode

Het bevragen van innovatiekredietontvangers en -afgewezenen over het belang van het instrument (conditionele vragen). Dit is vergelijkbaar met de eerder uitgevoerde TOK-evaluatie.⁴² Bijvoorbeeld: welke invloed heeft het krediet gehad op doorgang, omvang en snelheid van uitvoeren van het project en wat zijn de economische resultaten ervan. Door middel van controlevragen kan geprobeerd worden om opportunistische beantwoording uit te sluiten.

Conclusies en advies

Het kwaliteitsverschil tussen de eerste evaluatieoptie en de tweede en derde optie is groot. Bij de eerste optie wordt meer informatie over aanvragen en aanvragers gebruikt om tot een vergelijkbare controlegroep te komen, wat in mindere mate gedaan wordt bij beide andere opties. Dit geeft een betrouwbaarder inzicht in de effectiviteit van het innovatiekrediet dan beide andere ontwerpen. Het is op voorhand onduidelijk of de eerste evaluatieoptie praktisch uitvoerbaar is. Het totaal aantal aanvragers is gering en onduidelijk is hoe groot het grijze gebied is waar de evaluatie zich op richt (en op voorhand kan ook niet worden gezegd hoe groot dat moet zijn). Een tweede nadeel is dat de hypothesen alleen getoetst worden voor aanvragen die in het grijze gebied vallen. Ook dit kan problematisch zijn als het grijze gebied klein is. Als dit het geval is dan is de interpretatie van het gevonden effect beperkt.

Met de propensity score methode (optie 1) kan de eerste hypothese getoetst worden. Hypothese 2 kan worden getoetst met behulp van panelregressie (optie 2). Ten aanzien van hypothese 3 adviseert de werkgroep om gebruik te maken van een survey met conditionele vragen (optie 3), aangezien de kosten van het in kaart brengen van een controlegroep van niet-gebruikers zeer hoog zullen liggen. Hypothese 4 is niet eenvoudig te toetsen. Hiervoor moet worden gekeken naar alle uitstralingseffecten van het innovatiekrediet. Omdat het innovatiekrediet pas in 2008 van start is gegaan, zullen langetermijneffecten en spillovereffecten op dit moment echter beperkt zijn.

Omdat op voorhand niet valt vast te stellen hoe groot het grijze gebied is en omdat de propensity score methode niet alle hypothesen kan toetsen, ligt het voor de hand naast de propensity score methode ook beide andere evaluatieopties te gebruiken. Dit geeft een zo volledig mogelijk beeld en kan ook inzicht opleveren in hoe de verschillende evaluatieontwerpen zich tot elkaar verhouden. Dat laatste is nuttige informatie als moet worden bepaald hoe toekomstige evaluaties uitgevoerd worden. Het innovatiekrediet is in 2008 gestart en pas bij een volgende evaluatie zullen serieus verregaande analyses mogelijk zijn. De eerste evaluatie moet dus ook gebruikt worden om te leren wat voor het innovatiekrediet het beste evaluatieontwerp is.

⁴²Bureau Bartels (2000). Evaluatie van de TOK-regeling, Amersfoort.

4.3 Research and Development Aftrek (RDA)

Doelstelling

De Research and Development-aftrek (RDA) heeft als doel private R&D-investeringen te verhogen. Naast het doel om private R&D-investeringen uit te lokken is een onderliggend doel van de RDA (net als bij WBSO en Innovatiebox) een aantrekkelijk fiscaal klimaat te hebben voor buitenlandse bedrijven om hier hun R&D-activiteiten te vestigen. Aan dit doel wordt in dit rapport geen aandacht besteed.

Marktfalen

De legitimatie van de RDA komt voort uit het feit dat bedrijven minder aan R&D besteden dan maatschappelijk gewenst is, omdat zij de kennispillovers waar andere bedrijven mogelijk baat van hebben niet meewegen in hun beslissing over hun eigen R&D-uitgaven. Dit leidt tot een suboptimaal niveau van investeringen in R&D (zie de discussie in hoofdstuk 1). De financiële lasten voor de loonkosten voor R&D werden al verlaagd door de WBSO. Met de RDA worden ook de overige R&D-kosten (investeringen in apparatuur, verbruikte materialen e.d.) verlaagd, waardoor de fiscale behandeling van R&D-arbeid en -kapitaal meer in evenwicht wordt gebracht en de keuze voor de juiste kapitaal/arbeidsverhouding minder wordt verstoord.

Werking

De WBSO en de RDA zijn nauw met elkaar verweven. Om voor de RDA in aanmerking te komen dient ook een WBSO-aanvraag te worden ingediend. Er is één aanvraagformulier waarop dit tegelijkertijd wordt aangevraagd. Er is voor gekozen deze koppeling te maken om de administratieve lasten en uitvoeringskosten van de regeling zo veel mogelijk te beperken.

De RDA kent twee regimes:

- Aftrek op basis van *werkelijke* kosten en uitgaven⁴³: R&D-kosten en uitgaven mogen (uitgaande van 2012) voor 40% extra worden afgetrokken in de vennootschapsbelasting (VpB) of de inkomstenbelasting (IB). Bij een VpB-tarief van 25% is het nettovoordeel dus 10% van de R&D-kosten en -uitgaven die onder de RDA vallen.
- Aftrek op basis van *forfaitaire* kosten: aanvragers mogen 40% van €15 per WBSO-uur van de winst aftrekken, voor een maximum van 150 WBSO-uren per maand, ofwel 1800 WBSO-uren per jaar (dit is grofweg 1 fte). De €15 is gebaseerd op de verhouding arbeidskosten versus overige kosten bij R&D van kleine bedrijven en bedoeld als vergoeding voor die overige kosten. Gegeven dat de hoogte van het totale bedrag dat mag worden afgetrokken wordt bepaald door het aantal WBSO-uren, is de RDA in het forfaitaire regime feitelijk een fiscale kostenverlaging van R&D-arbeid. Tot die 1800 uur geldt: hoe meer R&D-arbeid een bedrijf inzet, hoe meer aftrek dat bedrijf krijgt. (€15 RDA per WBSO-uur resulteert daarbij, uitgaande van 2012, in $€15 \times 40\% = €6$ aftrek van de fiscale winst per WBSO-uur). Het forfaitairekostenregime kent een escape. Bedrijven die wat betreft het aantal uren onder het forfaitaire regime vallen, maar meer dan € 50.000 aan kosten en uitgaven hebben, komen in aanmerking voor aftrek op basis van werkelijke kosten. Eenmaal voor de escape gekozen, kan een bedrijf in dat jaar niet meer van het forfaitairekostenregime gebruikmaken.

De voornaamste reden om het forfaitairekostenregime te introduceren is het laag houden van de administratieve lasten van (kleine) ondernemingen met relatief lage R&D-inspanningen en het beperken van de uitvoeringslasten. De inschatting is dat het budgettaire beslag van de groep die in het forfaitaire regime valt in 2012 ongeveer €12 miljoen bedraagt (5% van het budget) en ongeveer 60% van de aanvragen beslaat.

Uitvoering

Binnen 3 maanden na afloop van het jaar moet een aanvrager melding maken van de gerealiseerde uren en (indien van toepassing, voor de niet-forfaitaire aanvrager) kosten en uitgaven doen. Waar nodig kan er een correctie op basis van realisatie ten opzichte van toekenning plaatsvinden. Alleen correctie naar beneden is mogelijk (voor beide regimes), bijplussen bij hogere realisatie is niet toegestaan.

⁴³Niet alle kosten zijn aftrekbaar. Zo zijn de kosten voor uitbesteed onderzoek bijvoorbeeld uitgesloten. Voor een compleet overzicht van aftrekbare en niet- aftrekbare kosten en uitgaven wordt verwezen naar de handleiding Research and Development Aftrek (RDA), AgentschapNL (2012).

De RDA loopt mee met de reguliere verliesrekening. Dit betekent dat maximaal 13 jaar carry-forward en/of carry backward mogelijk is.

Budget

Het budget voor de RDA bedraagt voor het begrotingsjaar 2012 €250 miljoen. De intentie is het budget voor de RDA de komende jaren op te laten lopen naar €375 miljoen in 2013 om vervolgens in 2014 op het maximum van € 500 miljoen te belanden dat is beoogd voor 2014 en verder. Indien de budgetverhogingen in 2013 en 2014 doorgaan en afhankelijk van het gebruik in 2012, zal het huidige aftrekpercentage van 40% worden verhoogd. Dit zal dan voor beide regimes gelden.

Voor de RDA en WBSO geldt de zogenaamde t-1/t+1 budgettaire spelregel. Dit betekent dat over- of onderschrijdingen in t-1 moeten worden gecompenseerd met het budget in jaar t+1 (welke wordt vastgesteld in jaar t). Dit kan door aanpassingen van de parameters bij gelijkblijvend budget of door extra budget.

Looptijd evaluatie periode

De RDA is op 1 januari 2012 van start gegaan; vanaf 1 mei 2012 kunnen aanvragen worden ingediend, zowel nieuwe als met terugwerkende kracht vanaf 1 januari 2012. De eerste evaluatie van de RDA staat gepland voor 2016.

Hypotheses

Kernhypothese

- 1) De RDA heeft een positief effect op de totale R&D-uitgaven van een onderneming voor aftrek van RDA: de RDA verlaagt de totaalprijs van R&D-activiteiten, waardoor ondernemingen meer aan R&D uit gaan geven (er is zowel een “inkomenseffect” als een “substitutie-effect”).

Hypotheses ter ondersteuning kernhypothese

- 2) De RDA gebaseerd op werkelijke kosten verandert de verhouding van R&D-loonkosten tot overige R&D-kosten ten gunste van overige R&D-kosten *via een verandering van relatieve prijzen* (er is een “substitutie-effect” binnen de R&D-uitgaven). Bij de bepaling van het substitutie-effect dient gecorrigeerd te worden voor een eventuele verandering in de verhouding van R&D-loonkosten tot overige R&D-kosten als gevolg van een grotere omvang van de R&D-activiteiten (grotere ondernemingen zijn vaak kapitaalintensiever).
- 3) De RDA gebaseerd op forfaitaire kosten verandert de verhouding van R&D-loonkosten tot overige R&D-kosten ten gunste van R&D-loonkosten *via een verandering van relatieve prijzen* (er is een “substitutie-effect” binnen de R&D-uitgaven).

Hypotheses tweede- en derde-orde effecten

- 4) De extra R&D-uitgaven van bedrijven als gevolg van de RDA (eerste-orde-effect) heeft positieve effecten op de innovativiteit (tweede-orde-effect) en de bedrijfsprestaties (derde-orde-effecten).
- 5) Via spillovers van kennis zijn er naast private rendementen bij de bedrijven die de extra R&D uitvoeren ook externe effecten bij andere bedrijven. Het maatschappelijk (economisch) rendement van de RDA (en WBSO) is daardoor groter dan het private rendement.

Tabel 3 Overzicht indicatoren bij evaluatieontwerp RDA

Hypotheses	Indicatoren		Databronnen⁴⁵
	Afhankelijk	Onafhankelijk	
1) De RDA heeft een positief effect op de totale R&D-uitgaven.	<ul style="list-style-type: none"> • Uitgaven aan R&D 	<ul style="list-style-type: none"> • Dummy voor gebruik RDA 	<ul style="list-style-type: none"> • AgNL • CBS
2) Door RDA gebaseerd op werkelijke kosten is er een “substitutie-effect” binnen de R&D-uitgaven.	<ul style="list-style-type: none"> • Aantal WBSO-uren 	<ul style="list-style-type: none"> • Dummy voor gebruik RDA • Dummy voor niet-forfaitaire deel 	<ul style="list-style-type: none"> • AgNL
3) Door RDA gebaseerd op forfaitaire kosten is er een “substitutie-effect” binnen de R&D-uitgaven.	<ul style="list-style-type: none"> • Aantal WBSO-uren 	<ul style="list-style-type: none"> • Dummy voor gebruik RDA • Dummy voor forfaitaire deel 	<ul style="list-style-type: none"> • AgNL
4) RDA heeft positieve effecten op de innovativiteit en de bedrijfsprestaties.	<ul style="list-style-type: none"> • (Arbeids-) productiviteit • Nieuwe producten & processen • Octrooien en patenten • Groei van bedrijven 	<ul style="list-style-type: none"> • Dummy voor gebruik RDA 	<ul style="list-style-type: none"> • AgNL/WBSO • AgNL/OCNL • CBS
5) Het maatschappelijk (economisch) rendement van de RDA (en WBSO) is groter dan het private rendement.	<ul style="list-style-type: none"> • Innovatie per sector 	<ul style="list-style-type: none"> • Gebruik RDA per sector 	<ul style="list-style-type: none"> • AgNL • CBS
Controlevariabelen [niet uitputtend]	<ul style="list-style-type: none"> • Projectinformatie • Bedrijfskenmerken • Bedrijfsspecifieke variabelen die samenhangen met het aanbod van en de vraag naar R&D-diensten⁴⁶ 		<ul style="list-style-type: none"> • AgNL • CBS

Evaluatieontwerp

Om goed het primaire effect van RDA aan te tonen (hypothese 1) is het belangrijk om vergelijkbare groepen te onderscheiden waarbij de ene groep geen RDA heeft ontvangen en de andere groep wel. Idealiter zouden de vergelijkbare groepen het gevolg zijn van gerandomiseerde toewijzing van RDA aan bedrijven die hiervoor in aanmerking komen (zie de discussie in hoofdstuk 2). Omdat bij RDA geen gerandomiseerd experiment heeft plaatsgevonden, moet gebruik worden gemaakt van andere evaluatiemethoden (zie hoofdstuk 2).

⁴⁴AgNL beschikt over alle aanvraaggegevens van de WBSO- en RDA-aanvragers. Deze gegevens zouden voor het evaluatieonderzoek kunnen worden gekoppeld aan CBS-data en belastinggegevens. Uit de WBSO-evaluatie is gebleken dat bij circa 95% van de bedrijven uit het WBSO-bestand een koppeling kan worden gemaakt met CBS-data. Hiernaast kan gebruik worden gemaakt van informatie uit de bedrijfsbezoeken van AgNL. Het is mogelijk data in te delen naar topsector. Het CBS heeft hier recent via maatwerk en een nulmeting een standaard voor beschikbaar.

⁴⁵Alleen nodig bij evaluatieopzet 3. Zie bijvoorbeeld: Lokshin & Mohnen, 2010. *How effective are level-based R&D tax credits? Evidence from the Netherlands*, UNU-MERIT Working Paper Series 040, United Nations University, Maastricht Economic and social Research and training centre on Innovation and Technology.

Optie 1: Gefaseerde asymmetrische uitbreiding RDA

Experimentele en controlegroep

De RDA zit nog in de uitrolfase. Het budget van de RDA wordt tussen 2012 en 2014 wellicht uitgebreid van 250 miljoen euro naar 500 miljoen euro. In het algemeen is deze fase ideaal om exogene variatie in de deelname aan RDA aan te brengen, zodanig dat later een betrouwbare effectmeting kan plaatsvinden. Vanwege de budgettaire omvang van de RDA en de verhouding tussen niet-forfaitaire en forfaitaire aanvragen wat betreft budgettaire beslag van ongeveer 1 staat tot 20, ligt het voor de hand om het niet-forfaitaire deel in eerste instantie te verhogen. De groep bedrijven die gebruik maakt van de forfaitaire RDA is dan de controlegroep. Bij de volgende verhoging, zal zowel het forfaitaire als het niet-forfaitaire deel verhoogd moeten worden, waarbij het forfaitaire deel wat betreft budgettaire beslag dan ruim verdubbelt ten opzichte van de situatie in 2012.⁴⁶ Beide groepen worden dan experimentele groep.

Identificatiestrategie

Bij een asymmetrische uitrol kan het effect van RDA bepaald worden door trends in indicatoren voor de experimentele groep te vergelijken met de trends voor de controlegroep. Als er een verschil in trends waarneembaar is tijdens de asymmetrische uitrolfase, dan kan geconcludeerd worden dat deze verschillen veroorzaakt worden door de uitbereiding van de RDA. Het onderzoek volgt dan een zogenaamd difference-in-difference design. Omdat het onderscheid tussen de experimentele en controlegroep gemaakt wordt met het idee van evaluatie, kan deze keuze als exogeen verondersteld worden. Deze manier van uitbreiding van het instrument maakt het mogelijk om het effect van het instrument relatief zuiver te schatten. Overigens moet hierbij worden aangetekend dat deze aanpak het effect meet van de *verhoging* tijdens de uitrolfase. Onder de veronderstelling dat meeropbrengsten van het fiscale instrument dalend zijn, geeft deze aanpak een ondergrens van het totale effect.

Bij de eerste asymmetrische uitbreiding kunnen hypothesen 1 en 2 getoetst worden. De verhoging van de niet-forfaitaire RDA zou volgens hypothese 1 moeten leiden tot hogere R&D-uitgaven bij de experimentele groep ten opzichte van de controlegroep. Volgens hypothese 2 zou het niet-loonaandeel van de R&D uitgaven bij de experimentele groep hoger moeten zijn dan bij de controlegroep. Bij de tweede asymmetrische uitbreiding kan naast hypothesen 1 en 2 ook hypothese 3 getoetst worden, omdat de forfaitaire RDA dan ook verhoogd wordt. Bij de tweede uitbreiding is de verhoging voor de forfaitaire groep relatief hoger dan voor de niet-forfaitaire groep.

Als blijkt dat de R&D-inspanningen van een bedrijf omhoog gaan, dan kan ook hypothese 4 getoetst worden door als uitkomstmaten te kijken naar de arbeidsproductiviteit, het aantal nieuwe producten (en/of processen), octrooien en patenten en de groei van het bedrijf. Dat laatste kan zowel in personeel als in omzet of winst van het bedrijf. Om hypothese 5 te toetsen moet worden gekeken naar hoe innovatief een bepaalde sector is. Dan zullen sectoren met elkaar vergeleken moeten worden die veel en weinig gebruik maken van het forfaitaire deel.

Econometrische specificatie en methoden

De voorgestelde aanpak kan met een standaard lineair model geschat worden volgens de difference-in-difference methode (zie de discussie in hoofdstuk 2). De afhankelijke variabele is daarbij (de jaarlijkse verandering in) een output- of outcome-indicator en de onafhankelijke variabelen zijn een dummy voor treatment en eventuele controlevariabelen. Difference-in-difference wordt soms gekoppeld aan (propensity score) matching, om zogenaamde double-robust schatters van een treatment effect te krijgen. Deze aanpak kan als robuustheidsanalyse meegenomen worden.

Gevoeligheden

Bedrijven die in aanmerking komen voor de forfaitaire RDA hebben de mogelijkheid om te kiezen voor de niet-forfaitaire RDA indien het bedrijf €50.000 of meer aan kosten maakt per jaar. De samenstelling van de controlegroep is daarom afhankelijk van de omvang van de verhoging van de niet-forfaitaire RDA. Eventueel kunnen ondernemingen die recht hebben op forfaitaire RDA, maar gebruik maken van niet-forfaitaire RDA buiten de sample gehouden worden.

Als bedrijven weten dat de forfaitaire RDA een jaar later wordt verhoogd dan de niet-forfaitaire RDA, dan is het mogelijk dat er bedrijven zijn die anticiperen op de verhoging. Dit kan leiden tot een onderschatting van de effectiviteit van de RDA. Als niet van tevoren bekend is dat de forfaitaire RDA in de toekomst verhoogd gaat worden, dan zullen bedrijven mogelijk anders reageren.

⁴⁶Beleidsmatig is het beter te rechtvaardigen om de forfaitaire groep een jaar verhoging te onthouden, aangezien de RDA bij de forfaitaire groep de prijs van R&D-arbeid verlaagt, terwijl de regeling juist bedoeld is om de overige R&D-input goedkoper te maken.

Optie 2: Discontinuïteit rond het maximum WBSO-uren en invoering RDA

Experimentele en controlegroep

Bedrijven boven de 1800 WBSO-uren krijgen de beleidsinterventie waar de RDA zich op richt: het bevorderen van R&D door het stimuleren van kosten/uitgaven van R&D anders dan loon. Bedrijven beneden deze grenswaarde krijgen deze beleidsinterventie niet, maar krijgen een RDA-beschikking op basis van forfaitaire kosten. De groep bedrijven die gebruik maakt van de niet-forfaitaire RDA kan dan vergeleken worden met de groep die gebruik maakt van de forfaitaire RDA. Beide groepen zijn experimentele groepen. De vergelijkbaarheid van de groepen wordt groter als bedrijven met veel minder of met veel meer dan 1800 WBSO-uren buiten beschouwing worden gelaten.

Identificatiestrategie

De discontinuïteit rond het maximum WBSO-uren kan gebruikt worden om het effect van de RDA op het aandeel loonkosten in de totale R&D-kosten te identificeren. Als forfaitaire bedrijven een hoger aandeel loonkosten hebben dan niet-forfaitaire bedrijven, dan kan hypothese 2 en/of hypothese 3 waar zijn. Hypotheses 2 en 3 worden dan gezamenlijk getoetst: er kan geen onderscheid gemaakt worden tussen beide hypothesen.

Daarnaast kan onderscheid gemaakt worden tussen bedrijven op basis van het aandeel loonkosten in hun R&D-uitgaven. Dit loonkostenaandeel kan extra als continue variabele opgenomen worden in het regression-discontinuity model. Volgens hypothesen 2 en 3 zullen forfaitaire bedrijven hun loonkostenaandeel vergroten als ze meer aan R&D uitgeven vanwege de RDA, terwijl niet-forfaitaire bedrijven hun loonkostenaandeel juist zullen verkleinen als ze hun R&D uitgaven verhogen. De hypothesen worden daarom ondersteund als de correlatie tussen het loonkostenaandeel en R&D-uitgaven groter is bij forfaitaire bedrijven dan bij niet-forfaitaire bedrijven. Deze aanpak toetst dus gelijktijdig of hypothesen 2 en 3 waar zijn, maar kan bij een bevestigend antwoord niet zeggen welke afzonderlijke hypothese waar is.

Ten slotte kan gebruikt worden dat de RDA pas recentelijk ingevoerd is. Bij invoering van de RDA kan een onderscheid gemaakt worden tussen forfaitaire en niet-forfaitaire bedrijven. Beide groepen zijn experimentele groepen. Omdat bedrijven in verschillende mate gebruik hebben gemaakt van innovatiesubsidies die zijn afgeschaft kort voor invoering van de RDA, zal de invoering van de RDA ook verschillende effecten hebben binnen beide regimegroepen. Een optie is om bedrijven daarom niet alleen op basis van RDA-regime te groeperen, maar ook op basis van uitgangspositie. Net als bij Optie 1 kan de difference-in-difference methode gebruikt worden, maar dan rond het moment van invoeren van RDA. Het is van belang goed in kaart te brengen welke financiële prikkels er waren voorafgaande aan de invoering van RDA en hoe deze de verschillende bedrijven beïnvloedden. Welke hypothesen getoetst kunnen worden, hangt af van de oude situatie.

Deze aanvullingen kunnen los, maar ook in combinatie met elkaar geanalyseerd worden. In dat laatste geval moet worden gekozen voor een difference-in-difference model waarbij de regression-discontinuity meegemodelleerd wordt.

Econometrische specificatie en methoden

Alle bovenstaande analyses kunnen worden uitgevoerd in een standaard lineair difference-in-difference model zoals besproken is in hoofdstuk 2. De afhankelijke variabele is daarbij een output- of outcome-indicator en de onafhankelijke variabelen zijn een dummy voor treatment en eventuele controlevariabelen. Hierin moeten dummyvariabelen opgenomen worden die aangeven in welk regime de bedrijven vallen en moet ook gekeken worden naar welke financiële prikkels er waren voorafgaande aan RDA.

De robuustheid van de geschatte effecten kan gecontroleerd worden door te variëren met de bandbreedte rond 1800 WBSO-uren die gebruikt wordt om de experimentele en controlegroep samen te stellen. Bij een grotere bandbreedte vermindert de vergelijkbaarheid van de controlegroep en de experimentele groep, maar neemt het aantal waarnemingen toe.

Gevoeligheden

De voorgestelde methode toetst alleen hoe het aandeel R&D-loonkosten reageert op RDA. Zonder additionele veronderstellingen is het niet mogelijk om te zeggen hoe RDA de totale RDA-uitgaven beïnvloedt. Dus bovenstaande methode geeft slechts een gedeeltelijke evaluatie.

Bedrijven die in aanmerking komen voor de forfaitaire RDA hebben de mogelijkheid om te kiezen voor de niet-forfaitaire RDA als de kosten/uitgaven boven de € 50.000 liggen. De samenstelling van de controlegroep is daarom afhankelijk van de omvang van de verhoging van de niet-forfaitaire RDA. Voor een zuiver beeld kunnen ondernemingen die recht hebben op forfaitaire RDA, maar gebruik maken van niet-forfaitaire RDA, buiten de sample gehouden worden.

In tegenstelling tot de situatie bij optie 1 is er geen controlegroep, maar zijn er alleen groepen die verschillen in treatment. Een zuivere meting van de effecten van de RDA is daarom niet waarschijnlijk. De subsidies waar bedrijven gebruik van maakten voor aanvang van de RDA waren vrij talrijk en divers. Het zal daarom moeilijk zijn om op bedrijfsniveau vast te stellen wat de treatment is geweest.

Optie 3: Vraag- en aanbodsysteem (traditionele regressieanalyse met tijdreeksen)

Experimentele en controlegroep

Niet van toepassing.

Identificatiestrategie

Bij deze optie wordt gekeken of er een correlatie is tussen de prijsverlaging van R&D als gevolg van WBSO/RDA en de R&D-uitgaven van bedrijven. Omdat de kosten van R&D marktuitskomsten zijn, is er sprake van een endogeniteitsprobleem: een negatieve correlatie tussen R&D-prijzen en R&D-uitgaven kan het gevolg zijn van meer aanbod van R&D-productiefactoren of van teruglopende vraag naar R&D-productiefactoren. Er zal daarom een instrument moeten worden gevonden dat wel met de vraag samenhangt, maar niet met het aanbod (en/of vice versa). Dit is een bekend probleem bij het schatten van vraag- en aanbodsysteem. Mohnen en Lokshin (2010) hebben een vraag- en aanbod-systeem geschat voor de WBSO.

Econometrische specificatie en methoden

Er worden twee vergelijkingen gebruikt, één voor de vraag naar R&D-diensten en één voor het aanbod van R&D-diensten in jaar t .⁴⁷

$$vraag_t = \alpha_v + \delta_v (1 - \tau_t) prijs_t + \beta_v X_t + \gamma_v Z_t^v + U_t^v$$

$$aanbod_t = \alpha_a + \delta_a prijs_t + \beta_a X_t + \gamma_a Z_t^a + U_t^a$$

Zowel de vraag als het aanbod hangen af van de prijs en van geobserveerde factoren X . RDA verlaagt voor de vragers van R&D-diensten de prijs met een fractie τ .⁴⁸ Omdat RDA gefaseerd geïntroduceerd is verandert deze fractie over jaren. Het grote probleem bij het schatten van dit soort vraag/aanbod-modellen is dat de prijs niet exogeen is. De prijs komt tot stand op de markt waar vraag en aanbod aan elkaar gelijk worden. Dus de geobserveerde evenwichtsprijs in een bepaald jaar is een functie van de vraag- en aanbodfunctie in dat jaar. Om de bovenstaande modelvergelijkingen te schatten moet daarom gebruikt gemaakt worden van instrumentele variabelen Z . Dit zijn variabelen die of alleen de vraag of alleen het aanbod beïnvloeden.

De parameters δ zijn de belangrijkste beleidsparameters. Deze laten zien hoe vraag en aanbod reageren op de prijs. Echter, door RDA verandert in eerste instantie de vraag, deze neemt toe. Als de vraag boven het aanbod ligt, dan zal de prijs toenemen totdat vraag en aanbod weer aan elkaar gelijk zijn. Hierdoor is het effect van RDA op R&D-diensten (in absolute waarde) kleiner dan δ_v . Nadat het model geschat is zal het dus gesimuleerd moeten worden om het nieuwe evenwicht te bepalen na invoering van RDA.

Bij het schatten van dit soort vraag/aanbod-modellen worden vaak ook tijdreeks-elementen meegenomen, en wordt bijvoorbeeld *seemingly unrelated regression* gebruikt. Dit is een standaard tijdreeksmethode om met geaggregeerde data vraag/aanbod-modellen te schatten.

⁴⁷Onder R&D-diensten kunnen zowel R&D-arbeid als andere R&D-uitgaven verstaan worden.

⁴⁸Deze opzet is een vereenvoudiging. De RDA is geen directe inputsubsidie zoals de WBSO, maar grijpt aan op de winstbelasting. Als een bedrijf geen winst maakt, is het voordeel van de RDA dus nul.

Gevoeligheden

Het is vaak moeilijk om variabelen Z te vinden die alleen van invloed zijn op het aanbod en/of alleen van invloed zijn op de vraag. Ook in dit geval zijn er geen overduidelijke kandidaten voor deze variabelen. Zonder zulke instrumentele variabelen is het niet mogelijk het endogeniteitsprobleem van de prijs van R&D-diensten op te lossen.

Als lineaire vergelijkingen worden verondersteld, zal deze aanpak geen goede indicatie geven van de marginale effecten van veranderingen in de prijzen van R&D-productiefactoren. Een niet-lineair systeem schatten is niet altijd goed mogelijk.

Optie 4: Vergelijk Nederland met andere landen

Experimentele en controlegroep

Er zijn twee experimentele groepen forfaitaire en niet-forfaitaire Nederlandse bedrijven. De controlegroep zijn vergelijkbare bedrijven in het buitenland.

Identificatiestrategie

Het ligt hierbij voor de hand om Duitsland te nemen waar zoiets als RDA niet bestaat. Ook dit zou moeten gebeuren in een difference-in-difference model. Echter deze vergelijking is lastig omdat zowel inzicht in de Nederlandse instituties als de Duitse instituties moet worden verkregen.⁴⁹ Daarnaast is gekoppelde informatie noodzakelijk (al is de koppeling al wel gemaakt bij CBS). Het voordeel van deze methode is wel dat het inzicht geeft in de effectiviteit van het gehele instrument en niet alleen van de uitbereiding of het verschil tussen beide regimes. Het nadeel is dat moet worden verondersteld dat de (conditionele) trend hetzelfde is in Nederland als in Duitsland.

Econometrische specificatie en methoden

De voorgestelde aanpak kan met een standaard lineair model geschat worden volgens de difference-in-difference methode. De afhankelijke variabele is daarbij (de jaarlijkse verandering in) een outputindicator en de onafhankelijke variabelen zijn een dummy voor treatment en eventuele controlevariabelen. Difference-in-difference wordt soms gekoppeld aan (propensity score) matching, om zogenaamde double-robust schatters van een treatment effect te krijgen. Deze aanpak kan als robuustheidsanalyse meegenomen worden.

Gevoeligheden

Identificatie: Het nadeel is dat moet worden verondersteld dat de (conditionele) trend hetzelfde is in Nederland als in Duitsland.

Data: Bij deze aanpak worden gegevens uit nogal verschillende bronnen gebruikt. Zo kunnen de definities en enquêtemethoden die het Duitse bureau voor de statistiek gebruikt verschillen van die van het CBS. Dit probleem is beperkt als Eurostatstandaarden gebruikt kunnen worden.

Optie 5: Surveymethode

Experimentele en controlegroep

Niet van toepassing.

Identificatiestrategie

RDA-aanvragers kunnen rechtstreeks gevraagd worden naar hun mening over de RDA en naar veranderingen in hun gedrag door de RDA. "Bent u een nieuw onderzoeksproject gestart door de RDA?" "Heeft u nieuwe investeringen gedaan door de RDA?" Een voorbeeld van zo'n enquête kan gevonden worden in Sectie 4.7 van de Evaluatie WBSO 2006-2010 door EIM (2012).

⁴⁹Duitsland heeft bijvoorbeeld geen federale WBSO maar via de Länder wel een intensief innovatiebeleid.

Er kunnen zowel vragen worden gesteld over het verleden, het heden en de toekomstverwachting. In de survey zal een vraag worden opgenomen over de winstverwachting van bedrijven in hun sector, waarbij moet worden aangegeven voor welke periode dit is.

Econometrische specificatie en methoden

De uitkomsten van een enquête hangen deels af van de manier waarop vragen geformuleerd zijn. Een pilotstudie met open interviews is nodig om de vragenlijst te valideren. De consistentie van een antwoord dient getoetst te worden door belangrijke vragen in een iets andere vorm te herhalen. “Markers” dienen te worden gebruikt om de uitkomsten te normaliseren. (Een werknemer zal bijvoorbeeld mogelijk anders antwoorden dan een directeur.)

Gevoeligheden

Uitkomsten: Antwoorden geven niet altijd een realistisch beeld van de werkelijkheid. Met name het geven van sociaal wenselijke antwoorden kan de uitkomst verstoren.

Conclusies en advies

Recentelijk is al beslist over het Belastingplan 2013 en daarin is niet gekozen voor asymmetrisch uitrollen van RDA. Omdat waarschijnlijk in 2014 RDA haar definitieve vorm heeft, is de mogelijkheid van asymmetrische uitrollen om exogene variatie te krijgen beperkt. Toch dient het aanbeveling om deze optie in gedachten te houden als besloten wordt RDA verder uit te breiden.

Optie 2 heeft de voorkeur boven optie 3, optie 4 en optie 5. Bij optie 2 worden verschillende mogelijkheden genoemd om onder bepaalde voorwaarden te corrigeren voor selectiviteit. Dit is lastiger om te doen bij een survey zoals besproken bij optie 5. Een vergelijking tussen landen zoals besproken in optie 4 is ook niet altijd even betrouwbaar, omdat landen vaak in meer dingen verschillen dan alleen de introductie van nieuw beleid. Het is daarom niet zeker dat landen dezelfde trend hebben. Het schatten van een vraag/aanbod-model zoals genoemd in optie 3 maakt gebruik van instrumentele variabelen. In dit geval is niet duidelijk welke variabelen goede instrumenten zouden kunnen zijn en het gebrek aan valide instrumenten zorgt voor onzuivere schatters voor de causale effecten van RDA. Bij optie 2 is wel een beperking dat het onderzoek zich alleen richt op relatief kleine RDA-gebruikers rond de grens van 1800 WBSO-uren. Tevens is bij deze optie een beperking dat slechts het effect van de RDA op het aandeel van loonkosten binnen de R&D-uitgaven onderzocht wordt, niet in hoeverre de RDA de totale R&D-uitgaven bij bedrijven heeft gestimuleerd.

4.4 Bewijs van Goede Dienst (BvGD)

Doelstelling

Het Bewijs van Goede Dienst (BvGD) heeft als doel om een grotere tevredenheid van ondernemers over gemeentelijke transactiele dienstverlening te bereiken. Hieronder wordt alle (simpele en complexe) dienstverlening die voortkomt uit individueel contact met de gemeente verstaan. Om dit doel te bereiken beoogt het instrument zowel de *objectieve* kwaliteit als de *gepercipieerde* kwaliteit van de dienstverlening te verbeteren. De objectieve kwaliteit kan verhoogd worden door bijvoorbeeld te voldoen aan gestelde aanvraag- en hersteltermijnen, responstijd te verbeteren en/of het aanbieden van actuele gemeentelijke informatie. De gepercipieerde kwaliteit hangt nauw samen met de informatieverstrekking over en het betrekken van ondernemers bij de (voorgenomen) verbetering in de objectieve kwaliteit. Deze kan worden verhoogd door ondernemers actief te betrekken bij verbeterplannen en actief informatie te verstrekken aan ondernemers.

Overheidsfalen

In tegenstelling tot de andere instrumenten die in dit rapport onder de loep zijn genomen adresseert het BvGD geen marktfaalen, maar overheidsfaalen. Te hoge gemeentelijke kosten, bijvoorbeeld vergunningaanvragen, administratieve lasten, onjuiste en/of ontoegankelijke informatie, om een activiteit die de maatschappij ten goede komt te starten kunnen de beschikbaarheid van bepaalde goederen en diensten die door ondernemers geleverd worden verminderen. Dit staat een efficiënte allocatie van goederen en diensten in de weg.

Werking

Het instrument BvGD bestaat uit twee fases. Fase 1 betreft een meting waarbij diverse ondernemersdossiers op tien door de gemeente zelf geselecteerde producten worden geaudit. De audit valt uiteen in drie sub-audits: een dossier-audit, een cliënt-audit en een issue-audit. In de dossier-audit worden dossiers van de geselecteerde producten beoordeeld op meetbare feiten. Bij de cliënt-audit wordt naar concrete ervaringen van bedrijven met de dienstverlening van de gemeente gevraagd. De issue-audit vormt een algemeen oordeel over een bepaald onderwerp voor geselecteerde producten op basis van informatie die bij de gemeente wordt opgevraagd. Deze gehele meting omvat een set van tien normen⁵⁰ die in kaart worden gebracht met een spindigram. Na afronding van de meting wordt, op basis van de scores van de betreffende gemeente op de tien normen, een verbeterplan opgesteld. Na het doorlopen van de meting en het opstellen van een verbeterplan kan de gemeente het BvGD ontvangen, deze is twee jaar geldig. Het verbeterplan hoeft hiervoor dus nog niet uitgevoerd te zijn.

Fase 2 betreft het ter uitvoering brengen van het in fase 1 opgestelde verbeterplan. Hiervoor wordt een periode van twee jaar genoemd, maar deze is niet bindend. Of en in welke mate gemeenten fase 2 uitvoeren kan dus van geval tot geval verschillen.

Om het gebruik van BvGD te stimuleren, bestaat er sinds 2009 een *voucherregeling*. Hiermee geeft het Ministerie van EZ financiële ondersteuning aan gemeenten, provincies en waterschappen om het BvGD in te voeren met behulp van een onafhankelijk adviseur. De voucherregeling bestaat uit twee vouchers op basis van minimaal 50% cofinanciering door gemeente. De eerste voucher is bedoeld voor uitvoering van fase 1 en heeft een maximale waarde van €5.000. Voor fase 2 is er een voucher van €10.000 beschikbaar. Het aantal gemeenten dat zonder voucherregeling de meting en het verbeterplan uitvoert is zeer beperkt. Het geld van de voucherregeling gaat rechtstreeks naar de adviesbureaus.

Uitvoering

De top 10 van meest aangevraagde vergunningen worden in principe geselecteerd voor de audit. Gemeenten maken de selectie uiteindelijk zelf. De vergunningen die onderdeel van de meting uitmaken kunnen zowel probleemdossiers als goedlopende dossiers zijn. De adviesbureaus checken steekproefsgewijs dossiers op vaste criteria, die in de praktijk wel verschillend door adviesbureaus worden geïnterpreteerd. Voor grote gemeenten bestaat deze steekproef uit ongeveer 200 dossiers, voor kleine gemeenten zijn dit er ongeveer 55.

Sinds begin 2012 is KING (Kwaliteitsinstituut Nederlandse Gemeenten) de uitvoerende organisatie voor en mede-eigenaar

⁵⁰Deze tien normen betreffen: 1.) Voldoen aan aanvraagtermijnen, 2.) Hersteltermijnen bij termijnoverschrijdingen, 3.) Volledigheid verzoeken en aanvragen, 4.) Inhoudelijke kennis en deskundigheid, 5.) Beleving toezicht, 6.) Toegankelijkheid van de gemeente, 7.) Actualiteit gemeentelijke informatie, 8.) Klanttevredenheid, 9.) Deugdelijke besluitvorming, 10.) Administratieve Lasten. Dit normenkader is i.s.m. VNO-NCW, MKB Nederland en VNG opgesteld.

van het BvGD-gemeenten.

Budget

Alleen de voucherregeling als onderdeel van het BvGD heeft een budgettair beslag. Het beschikbare bedrag voor deze regeling is €800.000 in 2012. De looptijd van de regeling is twee maal verlengd (in 2010 en 2011), maar zal in 2012 uitgefaseerd worden (na 2012 zullen gemeenten de kosten van fase 1 en fase 2 dus volledig zelf moeten dragen). Tot eind 2012 worden aanvragen nog gehonoreerd, waarna uitfinanciering tot eind 2013 plaatsvindt. Na 2013 zal het Rijk niet meer financieel bijdragen. De €800.000 geldt voor de gehele looptijd van de verlengde voucherregeling (dus 30-12-2011 tot en met 31-12-2013). Als het beschikbare budget op is, kan er geen gebruik meer worden gemaakt van de regeling.

Looptijd evaluatieperiode

Evaluatie staat oorspronkelijk nog gepland voor dit jaar (2012). Vanwege kosten- en responstechnische voordelen, wordt er voor gekozen om de evaluatie een jaar uit te stellen vanwege de mogelijkheid om dan resultaten van de enquête “MKB-vriendelijkste gemeente van Nederland” voor evaluatie te kunnen gebruiken. De evaluatie beslaat de periode 2009-2012. Daadwerkelijke evaluatie zal nu in 2013 plaatsvinden.

Hypotheses

Kernhypothese

- 1) Het BvGD leidt tot grotere tevredenheid bij ondernemers over de gemeentelijke dienstverlening.

Hypotheses ter ondersteuning kernhypothese

- 2) Het BvGD leidt tot betere informatievervalschaafing aan ondernemers over de gemeentelijke dienstverlening.
- 3) Door het BvGD worden ondernemers meer betrokken bij de gemeentelijke dienstverlening.
- 4) Het BvGD leidt tot verhoogde objectieve kwaliteit van de gemeentelijke dienstverlening.

Optionele hypothese

- 5) Van het BvGD gaat een reputatie-effect uit, waardoor gemeenten die slecht scoren qua ondernemerstevredenheid overgehaald worden hun prestaties te verbeteren.

Indicatoren

Het BvGD kent tien normen, zie bijlage. Drie van deze normen zijn gebaseerd op de mening (rapportcijfer) van ondernemers en daarvan komen er twee (deels) terug in de MKBVG enquête:

- Norm 4. Deskundigheid ambtenaar (Professioneel en capabel): Dit kan aansluiten bij rapportcijfer voor “Deskundigheid” (onder vraag 4 MKBVG)
- Norm 8. Klanttevredenheid (Klantgerichtheid en betrokkenheid): Dit kan aansluiten bij rapportcijfer voor “Begrip voor ondernemers” (onder vraag 4 MKBVG).

De werkgroep adviseert om extra vragen op te nemen in de enquête MKB-vriendelijkste gemeente van Nederland⁵¹, om zo een uitgebreider beeld te krijgen. Met name norm 5 en norm 7 lenen zich ervoor om zo te worden onderzocht.

- Norm 5. Beleving van toezicht (Beperkte toezichtlasten): bijvoorbeeld nieuwe subvraag/vragen bij vraag 4 MKBVG: rapportcijfer “Toezicht op naleving regels”
- Norm 7. Actualiteit gemeentelijke informatie: bijvoorbeeld nieuwe subvraag/vragen bij vraag 4 MKBVG: rapportcijfer “Actualiteit website” (sluit ook aan bij rapportcijfer “Communicatie”)

Voor de evaluatieontwerpen die hieronder besproken worden is het noodzakelijk om zo veel mogelijk voor de heterogeniteit tussen gemeenten controleren. Hiervoor zijn voor elke gemeente achtergrondvariabelen nodig die kunnen worden meegenomen in de empirische analyses. Enkele voorbeelden van variabelen die meegenomen kunnen worden zijn: urbanisatiegraad, omvang gemeente, gemiddeld per capita inkomen gemeente, percentage MKB/groot-bedrijven, aantal banen niet-publieke sector, geografische ligging, aantal ondernemers per 1000 inwoners, werkgelegenheidsgraad etc.

⁵¹ Nadeel is dat met deze enquête geen oordeel van grootbedrijf wordt meegewogen.

Het CBS beschikt over veel financiële cijfers van gemeenten. Op Statline staan bijvoorbeeld per gemeente heffingen en belastingen onderscheiden naar verschillende soorten. Meer gedetailleerde cijfers dan op Statline staan zijn in beginsel bij het CBS beschikbaar voor een dergelijke analyse.

Tabel 4 Overzicht indicatoren bij evaluatieontwerp BvGD

Hypotheses	Indicatoren		Databronnen
	Afhankelijk	Onafhankelijk	
1) Het BvGD leidt tot grotere tevredenheid bij ondernemers over de gemeentelijke dienstverlening.	<ul style="list-style-type: none"> Ondernemers-tevredenheid uitgesplitst naar subscores op normen 2 t/m 4 van BvGD [optioneel] Ondernemers-tevredenheid op alle 10 normen BvGD 	<ul style="list-style-type: none"> Dummy BvGD Ja/Nee 	<ul style="list-style-type: none"> MKBVG-enquête [optioneel] Aanvullende enquête onder ondernemers
2) Het BvGD leidt tot betere informatieverschaffing aan ondernemers over de gemeentelijke dienstverlening.	<ul style="list-style-type: none"> Mate van informatie-verschaffing aan ondernemers over de gemeentelijke dienstverlening 	<ul style="list-style-type: none"> Dummy BvGD Ja/Nee 	<ul style="list-style-type: none"> Aanvullende enquête onder gemeenten
3) Door het BvGD worden ondernemers meer betrokken bij de gemeentelijke dienstverlening.	<ul style="list-style-type: none"> Mate waarin ondernemers worden betrokken bij de gemeentelijke dienstverlening 	<ul style="list-style-type: none"> Dummy BvGD Ja/Nee 	<ul style="list-style-type: none"> Aanvullende enquête onder gemeenten
4) Het BvGD leidt tot verhoogde objectieve kwaliteit van de gemeentelijke dienstverlening.	<ul style="list-style-type: none"> Prestatie op tien normen BvGD 	<ul style="list-style-type: none"> Dummy BvGD Ja/Nee 	<ul style="list-style-type: none"> Aanvullende enquête onder gemeenten
5) Van het BvGD gaat een reputatie-effect, uit waardoor gemeenten die slecht scoren qua ondernemerstevredenheid overgehaald worden hun prestaties te verbeteren.	<ul style="list-style-type: none"> Ondernemers-tevredenheid 	<ul style="list-style-type: none"> Dummy BvGD Ja/Nee Dummy slecht scorende gemeenten Ja/Nee 	<ul style="list-style-type: none"> MKBVG-enquête Nulmeting Administratieve lasten Interbestuurlijk 2009 Informatie uit BvGD-gegevens
Controlevariabelen [niet uitputtend]	<ul style="list-style-type: none"> Urbanisatiegraad Omvang gemeente Gemiddeld per capita inkomen gemeente Percentage MKB/grootbedrijf Aantal banen niet-publieke sector Geografische ligging, Aantal ondernemers per 1000 inwoners Werkgelegenheidsgraad 		<ul style="list-style-type: none"> CBS

Evaluatieontwerp

De groep van gemeenten die meedoet aan het BvGD is niet willekeurig tot stand gekomen. Dit betekent dat er geen natuurlijke controlegroep van gemeenten is waarmee bovenstaande hypothesen getoetst kunnen worden. Het ideale evaluatieontwerp van een gerandomiseerd experiment kan dus niet worden toegepast. Deelnemende gemeenten verschillen sterk van gemeenten die niet meedoen. Dit blijkt uit het feit dat 30 van de 36 grootste gemeenten meedoen, terwijl het deelnamepercentage onder kleinere gemeenten veel lager is. Het vinden van vergelijkbare gemeenten met en zonder BvGD is daarom erg lastig, wat een kwantitatieve evaluatie moeilijk maakt. Een tweede beperkende factor is het relatief geringe budget dat beschikbaar is voor de evaluatie. Hieronder worden een aantal suggesties gedaan voor de evaluatie, waarbij rekening gehouden wordt met de budgetbeperking.

Optie 1 Panelregressie

Experimentele en controlegroep

Niet van toepassing.

Identificatiestrategie

Bij het beantwoorden van de kernhypothese moet worden bekeken hoe ondernemers aankijken tegen de gemeentelijke dienstverlening. Het ondervragen van ondernemers is duur en een grootschalige enquête onder ondernemers zal niet binnen het beschikbare budget kunnen. Daarom moet aansluiting gezocht worden bij bestaande databronnen. De meest logische kandidaat is de enquête "MKB-vriendelijkste gemeente van Nederland" (MKBVG)⁵². Het project MKB-vriendelijkste gemeente is voor het eerst in 2008 uitgevoerd, daarna voor een tweede keer in 2010/11 en binnenkort start de voorbereiding van de derde meting die in de eerste helft van 2013 wordt opgeleverd. Nadeel van deze enquête is dat geen zicht wordt gekregen op hoe grotere ondernemingen tegen de dienstverlening van gemeenten aankijken. Bovendien komen niet alle normen van het BvGD in deze enquête aan bod. De eerste drie hypothesen kunnen met deze gegevens deels beantwoord worden, maar dat is niet mogelijk voor de vierde hypothese over objectieve kwaliteit.

Daarom moet in aanvulling op de MKBVG-enquête een aanvullende enquête onder gemeenten worden gehouden om zicht te krijgen op wat gemeenten daadwerkelijk gedaan hebben (hypothese 2, 3 en 4). Hierbij worden gemeenten ondervraagd over de objectieve kwaliteit van de dienstverlening en ook of zij ondernemers hebben betrokken dan wel geïnformeerd. Waarbij een onderscheid wordt gemaakt tussen het moment van meting en de situatie 3 jaar geleden. Dus gemeenten worden geacht dezelfde vragen twee keer te beantwoorden, zowel voor de huidige periode als betrekking hebbende op de situatie 3 jaar geleden. Dit maakt het mogelijk een retrospectief panel te creëren.⁵³

De kosten van aanvullende enquêtering onder 100 gemeenten worden geschat op 40.000 euro⁵⁴ als wordt uitgegaan van face to face interviews bij de gemeenten. Telefonische enquêtering of enquêtering via post of internet is (veel) goedkoper, maar brengen het risico van een lagere respons met zich mee. Ook is minder diepgang voor achterliggende redenen/gemaakte keuzes (door beperkte of geen interactiviteit) mogelijk.

Econometrische specificatie en methoden

Met uitzondering van de extra vragen die worden toegevoegd aan de MKBVG-enquête van 2012/2013, biedt de MKBVG-enquête de mogelijkheid om een panel te creëren. De aanvullende enquête onder gemeenten geeft ook observaties op twee momenten in de tijd.⁵⁵ Een panel kan de robuustheid van de resultaten om verschillende redenen verhogen. Met name geldt dat met behulp van paneldata kan worden gecorrigeerd voor gemeentespecifieke effecten. Dit kan als de

⁵² De MKBVG-enquête wordt uitgevoerd door Lex Nova. Dit bureau voert ook de metingen voor BvGD uit bij gemeenten. Het ministerie van EZ moet waakzaam zijn voor mogelijke belangenverstrengeling in geval zij besluit met Lex Nova in zee te gaan bij een evaluatie.

⁵³ Idealiter zou je gemeenten en bedrijven op 3 punten in de tijd willen ondervragen (1. nulmeting, 2. BvGD ontvangen, 3. verbeterplan al dan niet uitgevoerd, dan wel bedrijven actief betrokken of geïnformeerd over de gemeentelijke dienstverlening). Aangezien resultaten in fase 2 niet zijn bijgehouden, zou je dus naast gemeenten zonder BvGD en bedrijven in die gemeente ook nog een keer gemeenten met BvGD (en bedrijven) in fase 2 willen ondervragen.

⁵⁴ Dit is als volgt berekend: 100 gemeenten keer een halve dag per interview (inclusief reistijd en tijd voor verslaglegging) keer een dagtarief van een onderzoeker van 800 euro (ex btw) per dag

⁵⁵ De schattingen op basis van een retrospectief panel zijn minder robuust in vergelijking tot een standaard panel, echter een retrospectief panel biedt meer analysemogelijkheden dan een cross-sectie, terwijl de meerprijs nihil is.

observatieperiode van het panel zodanig is dat er in deze periode voldoende gemeenten het traject van aanvraag van BvGD in zijn gegaan. Als dit het geval is, dan zijn twee tijdsperiodes voldoende om te kunnen corrigeren voor gemeentespecifieke effecten (zie de discussie over difference-in-difference methoden in hoofdstuk 2)⁵⁶. Het model dat geschat moet worden is een lineair panel data model, dat de volgende specificatie heeft:

$$Uitkomst_{it} = \alpha_i + \beta BvGD_{it} + \gamma X_{it} + \theta_t + \xi_{it}$$

De parameter waarin we geïnteresseerd zijn is β , die het effect weergeeft van het hebben van een bewijs van goede diensten op de uitkomstvariabele. Voor elke uitkomstvariabele wordt er een apart model geschat. Zo weten we hoe het hebben van een bewijs van goede dienst voor een gemeente elk van de uitkomsten gemeten in de MKBVG-enquête en de enquête onder gemeenten beïnvloedt. De variabele $BvGD_{it}$ geeft aan of gemeente i op moment t een bewijs van goede dienst had. Daarnaast zijn gemeentespecifieke effecten α_i opgenomen, evenals kalendertijdeffecten θ_t en wordt gecorrigeerd voor andere geobserveerde kenmerken van gemeenten die zijn opgenomen in X_{it} . Het lineaire paneldatamodel moet worden geschat met fixed effects methoden (zie hoofdstuk 2). Als de paneldata twee periodes beschrijven, dan reduceert het model tot het standaard difference-in-difference model zoals dat in hoofdstuk 2 is besproken.

Om hypothese 1 te toetsen moeten de 2 tot 4 normen die uit de MKBVG-enquête kunnen worden verkregen als uitkomstvariabelen gebruikt worden. De tweede hypothese wordt op dezelfde manier getoetst als de vorige, alleen is hier de afhankelijke variabele de mate van informatieverschaffing aan ondernemers over de gemeentelijke dienstverlening. Voor het toetsen van de derde hypothese is de afhankelijke variabele de mate waarin ondernemers worden betrokken bij de gemeentelijke dienstverlening. De vierde hypothese heeft betrekking op de objectieve kwaliteit, in kaart gebracht door de tien normen (of in elk geval een aantal daarvan). Voor een aantal normen is dit relatief makkelijk (bijvoorbeeld doorlooptijden, aantallen klachten en het percentage dat voldoet aan de aanvraagtermijnen) omdat zij via een enquête onder gemeenten kunnen worden achterhaald. Voor andere normen (bijvoorbeeld klanttevredenheid) ligt dit lastiger en kan alleen informatie verkregen worden door middel van een enquête onder ondernemers.

De vijfde hypothese heeft betrekking op gemeenten die slecht scoren. Deze hypothese stelt dat het aanvragen van BvGD een ander effect heeft voor gemeenten die minder goed scoren. Om deze hypothese te toetsen wordt een nieuwe variabele geconstrueerd die slecht scorende gemeenten identificeert. Slecht scorende gemeenten zijn dan gemeenten die in het laagste kwartiel van de kwaliteitsverdeling zitten. De nieuwe variabele is een dummyvariabele (slecht scorende gemeenten=1, overige gemeenten=0). Het regressie model wordt vervolgens uitgebreid naar

$$Uitkomst_{it} = \alpha_i + \beta BvGD_{it} + \delta BvGD_{it} \times Slecht_{it-1} + \lambda Slecht_{it-1} + \gamma X_{it} + \theta_t + \xi_{it}$$

Als het geschatte interactie-effect δ positief (en significant) is, dan geeft dat aan dat het BvGD inderdaad ervoor zorgt dat slecht scorende gemeenten hun uitkomsten sterk verbeteren als gevolg van BvGD. Het verkrijgen van een BvGD op tijdstip t heeft dan een extra positief effect als er sprake was van een slechte prestatie op $t-1$.

Er kunnen ook reputatie-effecten optreden die interessant zouden kunnen zijn voor de evaluatie. Reputatie-effecten kunnen van belang zijn als er binnen een bepaalde regio veel gemeenten deelnemen aan BvGD. Een gemeente die niet meedoet in zo'n regio kan druk voelen om ook mee te doen aan BvGD of om onafhankelijk daarvan ook haar dienstverlening te verbeteren. Dit kan onderzocht worden door in de set van achtergrondvariabelen X een extra variabele op te nemen die meet welke fractie van de gemeenten in de buurt deelnemen aan BvGD. Hierbij wordt verondersteld dat gemeenten zich vooral met hun "local peers" vergelijken en reputatie dus vooral lokaal uitstraalt. In de analyse kan nog worden bekeken of er een reputatie-effect van BvGD over de tijd heen te vinden is, zodat in de latere jaren gemeenten hun prestatie verbeteren ongeacht het feit dat zij niet deelnemen.

⁵⁶Het gaat hier dus niet om paneldata in de zuivere zin van het woord, maar om data verkregen uit antwoorden op vragen die op hetzelfde moment gesteld zijn voor twee momenten in de tijd.

Het grote risico van een paneldatamethode is dat het moment van aanvraag van BvGD gerelateerd is aan uitkomsten in vorige perioden (in de economische literatuur staat dit bekend als de Ashenfelter-dip). Het zou zo kunnen zijn dat gemeenten besluiten een aanvraag voor een BvGD te doen als ondernemers hun dienstverlening als goed ervaren, of juist als de dienstverlening als slecht te boek staat. In het eerste geval wordt het effect van het BvGD overschat, in het tweede geval is er sprake van onderschatting. De onzuiverheid is het gevolg van het verschil in trend tussen gemeenten die wel en niet aanvragen. Om dit te onderzoeken moeten worden bekeken hoe het moment van aanvraag gerelateerd is aan eerdere uitkomsten van de gemeentelijke dienstverlening in de gemeenten. Als de data maar twee perioden hebben, dan kan formeel niet onderzocht worden of er sprake is van een Ashenfelter-dip. Toch kan enig inzicht gekregen worden uit de volgende regressie

$$BvGD_{it} = \alpha + \beta Uitkomst_{it-1} + \gamma X_{i,t-1} + \xi_i$$

Als β significant is, dan kan dat een indicatie zijn dat de keuze voor een aanvraag van BvGD afhankelijk is van eerdere uitkomsten. Echter dat hoeft niet noodzakelijkerwijs zo te zijn, omdat het ook zo zou kunnen zijn dat de aanvraag voor BvGD gerelateerd zou kunnen zijn aan het gemeentespecifieke effect. Als de paneldata minimaal drie perioden bevatten, dan kan een betrouwbaardere test gedaan worden door middel van de volgende regressie:

$$BvGD_{it} = \alpha + \beta (Uitkomst_{it-1} - Uitkomst_{it-2}) + \gamma X_{i,t-1} + \xi_i$$

Als β nu significant is, dan geeft dit aan dat de keuze om een aanvraag voor BvGD te doen gerelateerd is aan de trend in de uitkomsten. Dit is problematisch omdat een verandering in de uitkomsten van een deelnemende gemeente nu niet zondermeer aan het BvGD kan worden toegeschreven. Als vooral gemeenten met verbeterende uitkomsten een BvGD aanvragen dan wordt het effect van BvGD overschat, als daarentegen vooral gemeenten met verslechterende prestaties een BvGD aanvragen, dan vindt er onderschatting van het effect plaats.

Gevoeligheden

Zoals hierboven al is opgemerkt verschillen de gemeenten die deelnemen aan BvGD van gemeenten die niet deelnemen. Om gevoel te krijgen voor de robuustheid van de resultaten is het van belang het probleem van *zelfselectie* zo veel mogelijk te ondervangen. Naast de hierboven beschreven analyses om deelname aan BvGD te verklaren uit vertraagde uitkomsten, kunnen ook nog andere manieren bedacht worden. Ten eerste kan worden gekeken naar de volgorde in tijd waarin gemeenten BvGD hebben aangevraagd. Het idee hierbij is dat gemeenten die als eerste een BvGD aanvragen sowieso gemeenten zijn die veel aandacht hebben voor het ondernemersklimaat en daardoor per definitie beter scoren dan andere gemeenten. Als dit het geval is wordt het effect van BvGD overschat. Om dit idee verder te formaliseren: in de econometrische regressieanalyse zou naast de dummy voor BvGD ook een interactieterm tussen dummy BvGD en het moment van tijd van aanvraag kunnen worden opgenomen. Als er geen zelfselectie-effecten zijn, heeft de interactieterm geen statistisch significant effect op de afhankelijke variabele (de prestatievariabele).

Ten tweede zou gebruik gemaakt kunnen worden van de “Nulmeting Administratieve Lasten Interbestuurlijk 2009”⁵⁷. Hoewel de score op Administratieve Lasten slechts één van de tien normen betreft, geeft deze nulmeting wel een indicatie van gemeenten die mogelijk beter zullen scoren. Ook deze variabele zou aan de regressievergelijking moeten worden toegevoegd. Als selectie-effecten niet opgetreden zijn, zou deze variabele geen effect moeten hebben op de afhankelijke variabele.

Eventueel kan de robuustheid van de paneldatamethode verbeterd worden door vooraf op basis van geobserveerde kenmerken een selectie te maken van welke gemeenten mee genomen kunnen worden in de paneldata-analyse. Leidend hierbij is dat gemeenten dezelfde ontwikkeling in de kwaliteit van de dienstverlening laten zien voor de invoering van het BvGD. Dit is per slot van rekening de belangrijkste veronderstelling van de analyse. Dus dat betekent dat gemeenten die een aanvraag voor BvGD deden op basis van de variabele ($Uitkomst_{it-1} - Uitkomst_{it-2}$) gekoppeld worden aan gemeenten die geen aanvraag deden.

⁵⁷ <http://www.rijksoverheid.nl/documenten-en-publicaties/kamerstukken/2009/06/11/brief-aan-de-tweede-kamer-over-de-nulmeting-administratieve-lasten-interbestuurlijk.html>

Optie 2 Surveymethode

Tijdens de instrumentensessie werd door de instrumentbeheerders de wens uitgesproken om dit jaar nog te evalueren. Indien wachten op de resultaten van de MKBVG-enquête geen optie is, zal tot enquêtering van bedrijven moeten worden overgegaan in zowel gemeenten die een BvGD hebben als gemeenten zonder BvGD. Dit zal naar mening van de werkgroep echter wel een substantiële verhoging van het beschikbare budget vereisen. Uitgaande van een steekproef van 50 gemeenten met BvGD en 50 gemeenten zonder BvGD en het ondervragen van 10 bedrijven per gemeente betekent dit dat 1000 bedrijven moeten worden geënquêteerd. Dit betekent een veelvoud van het hierboven genoemde budget voor aanvullende enquêtering van gemeenten. Een voldoende hoge respons onder bedrijven kan alleen gekregen worden door telefonische enquêtering en dit is een kostbare aangelegenheid.

Een voordeel van een aanvullende enquête is wel dat ook zicht kan worden gekregen op de andere normen van het BvGD die niet in de MKBVG-enquête staan. Ook bij aanvullende enquêtering onder ondernemers verdient het aanbeveling zowel te vragen naar de kwaliteit van de dienstverlening nu als die van 3 jaar geleden in verband met het vormen van een retrospectief panel. In dat geval kunnen alle empirische analyses zoals hierboven besproken toegepast worden.

Bij deze methode kunnen gemeenten conditionele vragen voorgelegd krijgen. Bijvoorbeeld had u de dienstverlening van uw gemeenten voor ondernemers in meer of mindere mate verbeterd zonder de aanvraag van een BvGD? Aan ondernemers kan daarentegen bijvoorbeeld de vraag worden gesteld in hoeverre zij de dienstverlening in hun gemeente verbeterd vinden ten opzichte van de situatie voordat hun gemeente een BvGD heeft ontvangen. Gezien de beperkte omvang van het budget zal de steekproef beperkt zijn.

Conclusie en advies

Omdat er sprake is van selectiviteit bij de gemeenten die meedoen aan BvGD, is het niet mogelijk tot een goede controlegroep te komen en om daarmee de hypothesen te toetsen. Het ideale evaluatieontwerp van een gerandomiseerd experiment is niet mogelijk. Dit betekent dat extra veronderstellingen gemaakt moeten worden om de causaliteit tussen instrument en waargenomen resultaat vast te stellen. Een vaak gebruikte terugvaloptie is de surveymethode met conditionele vragen (optie 2). Aan deze optie kleeft echter het risico van sociaal wenselijke antwoorden, wat de betrouwbaarheid van de uitkomsten schaadt.

De werkgroep adviseert daarom te kiezen voor een tussenoplossing waarbij wordt overgegaan tot econometrische analyse van enquêteresultaten (optie 1). Bij de analyse moeten zoveel mogelijk retrospectieve gegevens verzameld worden om te kunnen controleren voor bestaande verschillen tussen gemeenten. Daarnaast moet worden gekeken naar de mate en reden van zelfselectie. De resultaten van deze analyse kunnen een indruk geven van de effectiviteit van het BvGD, maar de betrouwbaarheid van de resultaten is afhankelijk van het soort zelfselectie van BvGD-aanvragers. Met deze methode kunnen alle vijf de hypothesen worden getoetst, mits naast de MKBVG-enquête aanvullend ook onder gemeenten wordt geënquêteerd.

Het budget dat momenteel voor de evaluatie beschikbaar is gesteld, is beperkt in omvang. Dit botst met de behoefte om snel tot een evaluatie te komen. Dit beperkt de mogelijkheden om inzicht in het effect van het instrument te krijgen. De werkgroep adviseert het budget voor de evaluatie te verhogen, indien er behoefte bestaat om sneller een indruk te krijgen over de werking van het BvGD en/of het gewenst is het BvGD te evalueren op alle 10 de normen. Indien wordt gekozen voor aanvullende enquêtering, kunnen optie 1 en 2 parallel aan elkaar worden uitgevoerd.

Bijlage Normen Bewijs van Goede Dienst

I Termijnen	Minimumnorm	Plusnorm
1. Voldoen aan aanvraagtermijnen > Eenduidige afhandeltermijnen	100% van de aanvragen wordt binnen de wettelijke aanvraagtermijn verleend	70% van de aanvragen wordt sneller dan de wettelijke aanvraagtermijn verleend
2. Hersteltermijn bij termijnoverschrijding > Anticipatie op termijnoverschrijding	60% van de termijnoverschrijdingen wordt binnen 5 werkdagen hersteld	80% van de termijnoverschrijdingen wordt binnen 5 werkdagen hersteld
3. Volledigheid verzoeken en aanvragen > Ontvankelijkheid van eerste aanvraag.	85% van de aanvragen is bij eerste indiening ontvankelijk	95% van de aanvragen is bij eerste indiening ontvankelijk

II Professionaliteit	Minimumnorm	Plusnorm
4. Deskundigheid ambtenaar > Professioneel en capabel	7	9
5. Beleving van toezicht > Beperkte toezichtslasten	7	9

III Ondernemersgerichtheid	Minimumnorm	Plusnorm
6. Responsnelheid > Tijd tussen de ontvangstbevestiging en het inhoudelijke antwoord van de gemeente	Gemiddelde responstijd is 3 dagen	Gemiddelde responstijd is een dag
7. Actualiteit gemeentelijke informatie	90% van de informatie via website en documentatie is actueel	100% van de informatie via website en documentatie is actueel
8. Klanttevredenheid > Klantgerichtheid en betrokkenheid	7	9

IV Betrouwbaarheid	Minimumnorm	Plusnorm
9. Deugdelijke besluitvorming	10% bezwaar- en beroepsprocedures zijn gegrond	5% bezwaar- en beroepsprocedures zijn gegrond
10. Administratieve lasten ondernemers > Beperkte administratieve lasten	25% lager dan landelijk gemiddelde administratieve lasten	35% lager dan landelijk gemiddelde administratieve lasten

4.5 Centra voor Innovatief Vakmanschap (CIV)

Doelstelling

Centra voor Innovatief Vakmanschap (CIV) beogen meer MBO-bètatechnici te laten afstuderen aansluitend bij de behoefte van het bedrijfsleven in de topsectoren, een grotere betrokkenheid van bedrijven bij praktijkopdrachten te stimuleren en meer werknemers uit bedrijfsleven van de topsectoren te laten deelnemen aan cursussen bij CIV. Het idee is dat hierdoor de kloof tussen aanbod en vraag op de arbeidsmarkt voor bètatechnici wordt verkleind.

Marktfalen

Onderwijsinstellingen worden momenteel onvoldoende geprikkeld opleidingen aan te bieden die goed aansluiten op de wensen van bedrijven. Hierdoor ontstaat een suboptimale allocatie van arbeidskrachten en een suboptimaal aanbod van (naschoolse) opleidingen. Het gebrek aan prikkels voor een goede aansluiting tussen opleidingen en banen kan het gevolg zijn van informatieasymmetrie tussen onderwijsinstellingen en bedrijven. CIV dienen de informatieasymmetrie tussen onderwijsinstellingen en bedrijven te verminderen door er voor zorgen dat zowel onderwijsinstellingen als studenten eerder op de hoogte zijn van veranderingen in de vraag naar arbeid waardoor de arbeidsmarkt beter functioneert en overheidsmiddelen efficiënter worden besteed.

Werking

CIV richten zich op nieuwe vormen van onderwijs en op het verhogen van de kwaliteit van het reguliere onderwijs door actieve inzet van het bedrijfsleven bij onderwijsprogramma's. Denk hierbij aan stages, praktijkopdrachten en bijvoorbeeld onderwijs op de werkvloer. CIV worden vorm gegeven als Publiek Private Samenwerking (PPSen)⁵⁸ en staan dus buiten de school of het bedrijfsleven. Veelal zijn ze wel gekoppeld aan onderwijsinstellingen.

De onderwijsinstellingen stellen met actieve betrokkenheid van een netwerk van bedrijfsleven in de regio uit de topsectoren businessplannen op. Op basis van een tender worden de beste voorstellen door de rijksoverheid financieel ondersteund. Subsidiering vindt plaats op basis van cofinanciering. De rijksoverheid verlangt cofinanciering van minimaal 50% van de investeringskosten voor een CIV. De rijksoverheid stelt een bedrag van €2 miljoen beschikbaar gedurende een periode van vijf jaar (€400.000 per jaar). Na deze periode van vijf jaar moeten CIV op eigen benen kunnen staan. De andere helft moet worden opgebracht door de onderwijsinstelling (25%) en aangesloten bedrijven (25%). In totaal is er dus minimaal €4 miljoen beschikbaar voor een CIV over een periode van vijf jaar. De groep bedrijven die in een CIV investeren (de primaire partners), omvat meestal een groep van 6-20.

Wanneer de bereidheid van bedrijven en lokale overheden voor financiering groot genoeg is, kan het voorkomen dat MBO-instellingen die niet een CIV subsidie krijgen, soms toch proberen hun business case te verwezenlijken door elders financiering te vinden, zeker wanneer deze wel goed beoordeeld is.

Uitvoering

Elke topsector kan aanspraak maken op één CIV. Middelen die voor één topsector zijn gereserveerd, zullen pas voor een andere topsector worden ingezet als de oorspronkelijke aanvrager tot twee keer toe plannen van onvoldoende kwaliteit heeft ingediend.

Binnenkomende voorstellen in de vorm van businesscases worden via een pitch voor een selectiecommissie geselecteerd en gerangschikt. Als het budget onvoldoende is om alle voorstellen te honoreren, dan wordt de rangschikking gebruikt om alleen de betere voorstellen te selecteren. Beoordeling vindt plaats op basis van de volgende criteria:

- Randvoorwaardelijke criteria: het centrum profileert zich op en binnen de topsector, er is sprake van gelijkwaardige betrokkenheid van relevante instellingen en andere partners (blijkend uit de kosten-batenanalyses), cofinanciering is geregeld, het centrum heeft toegang tot en inbreng vanuit (wetenschappelijke) kennis.
- Kerncriteria: ambitie, partners, haalbaarheid en duurzaamheid. Uiteindelijk wordt de ambitie van het plan afgewogen tegen de concreetheid en haalbaarheid van het plan.

Het selectieproces bestaat uit twee rondes. In de eerste ronde beoordeelt een expertcommissie of er aan de randvoorwaardelijke criteria wordt voldaan. In de tweede ronde beoordeelt de expertcommissie het businessplan. Wanneer aan één van de randvoorwaardelijke criteria niet wordt voldaan, betekent dit dat de aanvraag automatisch als

⁵⁸Bijvoorbeeld in de vorm van een stichting, BV of (coöperatieve) vereniging.

onvoldoende wordt beoordeeld. Als wel is voldaan, worden de vier kerncriteria afzonderlijk beoordeeld op basis van een tienpuntsschaal (dus maximaal 40 punten). Wanneer een aanvraag op één van de kerncriteria 0-3 punten scoort, wordt de aanvraag automatisch als onvoldoende beschouwd.

Het Platform voor Bètatechniek voert de tender uit, doet de toekenningen van de te honoreren voorstellen en monitort ze vanaf het moment van toekenning via het volgen van prestatieafspraken, via bezoeken van auditcommissies, gezamenlijke workshops en samenwerk- en uitwisselingsdagen.

Continuïteit van de subsidie is afhankelijk van tussentijdse resultaten die na twee jaar op basis van met het Platform voor Bètatechniek (PBT) overeengekomen prestatiecontracten worden gemonitord. Bij onvoldoende resultaten na twee jaar wordt de subsidie voor de resterende drie jaar stopgezet.

Budget

Voor de centra op MBO-niveau is door OCW een budget van € 11 mln. ter beschikking gesteld voor 6 CIV. Daarnaast levert EZ budget voor 2 centra in de groene sectoren. Deze gelden volgen echter een eigen procedure en proces los van de andere CIV. Vanwege deze reden wordt in onderstaande evaluatieontwerpen alleen de CIV meegenomen die worden beoordeeld en begeleid door PBT.

Looptijd evaluatieperiode

De beoogde evaluatie voor de CIV staat gepland voor 2017. De evaluatie zal de periode 2012-2017 beslaan.

Hypotheses

Kernhypothese

- 1) CIV leiden tot een grotere uitstroom van bètatechnici op MBO-niveau richting topsectoren (kwantitatief effect).

Hypothese ter ondersteuning kernhypothese

- 2) MBO-bètatechnici die een opleiding hebben genoten aan een CIV vinden na het afstuderen sneller een baan dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).
- 3) MBO-bètatechnici die een opleiding hebben genoten aan een CIV verdienen na het afstuderen een hoger salaris dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).

De CIV-subsidie kan andere MBO-instellingen zonder CIV er toe aanzetten om vergelijkbare initiatieven te ontplooiën. Om dit effect mee te nemen moeten hypothese 1, 2 en 3 dus ook worden getoetst voor projectplannen die tijdens de tender als kansrijk zijn beoordeeld, maar niet zijn gehonoreerd (hypotheses 4, 5, 6 & 7):

Hypotheses over aanjagende werking CIV

- 4) Als kansrijk beoordeelde projectplannen leiden tot een grotere uitstroom van bètatechnici op MBO-niveau richting topsectoren (kwantitatief effect).
- 5) MBO-bètatechnici die een opleiding hebben genoten aan een instelling waarvan het projectplan als kansrijk is beoordeeld, vinden sneller een baan dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).
- 6) MBO-bètatechnici die een opleiding hebben genoten aan een instelling waarvan het projectplan als kansrijk is beoordeeld, verdienen een hoger salaris dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).

Optionele hypothesen

- 7) CIV en/of instellingen waarvan het projectplan als kansrijk is beoordeeld, bieden meer nascholing gericht op bètatechnici aan.

Benodigde gegevens

Beschikbaarheid van data

Er zal een koppeling moeten worden gemaakt tussen OCW/DUO-data en polisgegevens (loonaangifte). Dit is mogelijk bij het CBS via het Sociaal Statistisch Bestand (SSB). Er zijn hierin gegevens beschikbaar van deelnemers/geslaagden van MBO-instellingen, onderscheiden naar instelling, opleidingsrichting et cetera over het tijdvak 2005-2010. Via diverse baan- en

Tabel 5 Overzicht indicatoren bij evaluatieontwerp CIV

Hypotheses	Indicatoren		Databronnen
	Afhankelijk	Onafhankelijk	
1) CIV leiden tot een grotere uitstroom van bètatechnici op MBO-niveau richting topsectoren (kwantitatief effect).	<ul style="list-style-type: none"> Afgestudeerde is werkzaam in topsector 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee 	<ul style="list-style-type: none"> OCW/DUO SSB
2) MBO-bètatechnici die een opleiding hebben genoten aan een CIV, vinden na het afstuderen sneller een baan dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).	<ul style="list-style-type: none"> Duur tussen afstuderen en moment van instroom 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee 	<ul style="list-style-type: none"> OCW/DUO SSB
3) MBO-bètatechnici die een opleiding hebben genoten aan een CIV, verdienen na het afstuderen een hoger salaris dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).	<ul style="list-style-type: none"> (Start)salaris van afgestudeerde 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee 	<ul style="list-style-type: none"> OCW/DUO SSB
4) Als kansrijk beoordeelde projectplannen leiden tot meer uitstroom bètatechnici op MBO-niveau richting topsectoren (kwantitatief effect).	<ul style="list-style-type: none"> Afgestudeerde is werkzaam in topsector 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee Indicator CIV afgewezen 	<ul style="list-style-type: none"> OCW/DUO SSB
5) MBO-bètatechnici die een opleiding hebben genoten aan een instelling waarvan het projectplan als kansrijk is beoordeeld vinden sneller een baan dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).	<ul style="list-style-type: none"> Duur tussen afstuderen en moment van instroom 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee Indicator CIV afgewezen 	<ul style="list-style-type: none"> OCW/DUO SSB
6) MBO-bètatechnici die een opleiding hebben genoten aan een instelling waarvan het projectplan als kansrijk is beoordeeld verdienen een hoger salaris dan MBO-bètatechnici die dit niet hebben gedaan (kwalitatief effect).	<ul style="list-style-type: none"> (Start)salaris van afgestudeerde 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee Indicator CIV afgewezen 	<ul style="list-style-type: none"> OCW/DUO SSB
7) CIV en/of instellingen waarvan het project als kansrijk is beoordeeld, bieden meer nascholing gericht op bètatechnici aan.	<ul style="list-style-type: none"> Indicator voor nascholing 	<ul style="list-style-type: none"> Indicator CIV Ja/Nee Indicator CIV afgewezen 	<ul style="list-style-type: none"> enquête
Controlevariabelen [niet uitputtend]	<ul style="list-style-type: none"> Indicatorvariabelen opleiding Indicatorvariabelen ROC Observeerbare kenmerken afgestudeerder 		<ul style="list-style-type: none"> CBS OCW/DUO

uitkeringsbestanden kan op persoonsniveau onderzocht worden welke carrière een afgestudeerde op de arbeidsmarkt doorloopt. De betreffende baangegevens zijn grotendeels gebaseerd op de loonaangifte. Bij een baan kunnen bijvoorbeeld ook de Standaard Bedrijfsindeling (SBI)⁵⁹ en de grootteklasse van de werkgever worden meegenomen. Ook wanneer betrokkenen werkzaam zijn als zelfstandige is dat in het SSB terug te vinden. De betreffende bestanden omvatten gegevens over periode van de baan/uitkering/zelfstandigheid, beloning, maar niet over de aard van de werkzaamheden (in het bijzonder of die aansluiten bij de opleidingsrichting – vooral een punt van aandacht als de werkgever een groot bedrijf is met een andere kernactiviteit (SBI) waarbij de afgestudeerde wellicht toch in zijn eigen vak werkzaam is).

Om de koppeling tussen OCW/DUO-data en het SSB mogelijk te maken zullen de BRINs van de onderwijsinstellingen en KvK-nummers van de bedrijven moeten worden aangeleverd. Bedrijven kunnen aan topsectoren worden toegekend via de SBI. Opleidingen kunnen aan topsectoren worden toegekend via de Standaard Onderwijs Indeling (SOI)⁶⁰.

Indicatoren

Belangrijke variabelen zijn op het niveau van afstudeerders van experimentele en controlegroep(en):

- Afstudeerdatum en datum begin eerste baan
- Gegevens over carrièrepad: opeenvolgende baanduren, werkloosheidsduren
- Gegevens over arbeidsinkomen over de tijd
- Sector waarin opeenvolgende banen worden gevonden

Deze variabelen zijn waarschijnlijk beschikbaar in het SSB. Op basis van een vergelijking tussen de resultaten voor de verschillende groepen kunnen (bij voldoende aantallen) de hypothesen 1, 2, 3, 4, 5 en 6 worden getoetst. Ook voor de langere termijn kan het carrièrepad van de CIV en ROC niet-CIV afstudeerders in het SSB worden gevolgd en kan worden bekeken hoe omvangrijk de verschillen zijn in de eventuele arbeidsmobiliteit tussen topsectoren en andere (technische) sectoren. Daarmee zou onderzocht kunnen worden hoe groot het spillovereffect is van de CIV technische opleidingen voor de gehele technische arbeidsmarkt. Voor het toetsen van de hypothese 7 is informatie nodig op het niveau van de ROC. Dat moet waarschijnlijk apart worden bevraagd.

Evaluatieontwerp

Voor het toetsen van de hypothesen zijn twee complementaire evaluatieontwerpen nodig. Het eerste ontwerp (Optie 1) kan gebruikt worden om de kernhypothese te toetsen. Bij dit ontwerp worden de prestaties van opleidingen vergeleken. Het tweede ontwerp (Optie 2) heeft als doel om de ondersteunende hypothesen te verifiëren. Hierbij worden afstudeerders met elkaar vergeleken.

Optie 1 Vergelijking van opleidingen

Experimentele en controlegroep

Bij de eerste optie wordt de statistische analyse gedaan op het niveau van de opleiding. Er wordt een onderscheid gemaakt tussen opleidingen die wel een CIV hebben en opleidingen zonder CIV. Daarnaast worden de prestaties van een opleiding voordat een CIV is ingevoerd vergeleken met de prestaties van dezelfde opleiding na invoering. Deze optie is primair bedoeld voor het toetsen van de kernhypothese en hypothese 4, maar kan ook voor de andere hypothesen gebruikt worden.

Identificatiestrategie

Het effect van CIV kan bepaald worden door na te gaan of er een verandering in de trend van de prestaties van opleidingen optreedt na invoering van een CIV. Hierbij wordt de trend in de prestaties van opleidingen die zonder CIV blijven als referentie gebruikt. Deze aanpak is een toepassing van de *difference-in-difference* methode die in hoofdstuk 2 besproken is.

⁵⁹De Standaard Bedrijfsindeling (SBI) is een hiërarchische indeling van economische activiteiten. Het CBS gebruikt de SBI onder meer om bedrijfseenheden in te delen naar hun hoofdactiviteit. Zie CBS voor meer informatie.

⁶⁰De SOI is een classificatie van opleidingen naar niveau en richting. De SOI is ontwikkeld voor gebruik bij statistiek en onderzoek en voor administratieve doeleinden in Nederland. Zie CBS voor meer informatie.

Econometrische specificatie en methoden

De afhankelijke variabele die het resultaat van de opleiding meet (bij de kernhypothese is dat het deel van de leerlingen van een opleiding die in een topsector gaan werken), dient te worden geregresseerd op een indicator die aangeeft of opleiding j in regio r een CIV heeft op tijdstip t . De geschatte coëfficiënt van de indicator meet het causale effect van CIV op de afhankelijke variabele.

Er kan in drie dimensies worden gecorrigeerd voor (ongeobserveerde) effecten die de geschatte invloed van CIV kunnen verstoren. Als eerste kan er gecorrigeerd worden voor veranderingen over de tijd die op alle opleidingen betrekking hebben. Daarnaast kan er bij voldoende waarnemingen gecontroleerd worden voor (ongeobserveerde) eigenschappen van opleidingen die niet over de tijd veranderen. Een derde dimensie is de regio waarin een opleiding gegeven wordt. De regressievergelijking is dan:

$$Y_{jrt} = \beta_0 + \beta_1 CIV_{jrt} + \eta_j + \rho_r + \theta_t + u_{jrt}$$

In deze vergelijking is Y de prestatie van de opleiding, geeft CIV aan of een er CIV is en zijn er “fixed effects” voor opleiding, η , regio ρ en tijd θ . De β is de te schatten coëfficiënt en u is het residu.

Gevoeligheden

Het is interessant om een verdere vergelijking te maken met de andere groepen. Zo is het nuttig om de analyse uit te breiden met informatie over opleidingen waarvoor de CIV wel was aangevraagd, maar waarvoor het niet was toegekend. Daarnaast kan er ook gekeken worden naar opleidingen zonder CIV die gegeven worden aan een ROC die voor een andere opleiding wel een CIV heeft. Zijn er verschillen binnen een ROC tussen opleidingen met en zonder CIV?

Optie 2 Vergelijking van afgestudeerden

Experimentele en controlegroep

Voor de toetsing van de ondersteunende hypothesen (hypothese 2 en 3) kan het instrument worden geëvalueerd op het niveau van studenten/afstudeerders. Drie primaire groepen kunnen worden onderscheiden: 1) Studenten afgestudeerd bij een opleiding waar een instelling een subsidie voor heeft gekregen, 2) studenten afgestudeerd bij een opleiding waarvoor een subsidie is aangevraagd die niet is toegekend, maar waarvoor alternatieve financiering is verkregen 3) studenten die bij een opleiding waarvoor een subsidie is aangevraagd die niet is toegekend en waarvoor geen alternatieve financiering is verkregen. Deze optie is ook van toepassing op hypothesen 5 en 6.

Identificatiestrategie

Er wordt een vergelijking gemaakt tussen enerzijds een groep van personen die zijn afgestudeerd aan een CIV en anderzijds personen van vergelijkbare opleidingen die niet bij een CIV zijn afgestudeerd. Er wordt hierbij onderzocht of er een verschil is in latere arbeidsmarktprestaties tussen vergelijkbare groepen afstudeerders van beide opleidingen. Voor wat betreft de vergelijkingsgroepen zijn er meerdere mogelijkheden: a) afstudeerders uit dezelfde opleidingsrichting in een andere ROC; b) afstudeerders van eerdere cohorten van deze opleiding (in de periode voordat CIV was gestart); c) afstudeerders uit andere opleidingen binnen het ROC die geen CIV hebben gekregen.

Econometrische specificatie en methoden

Er kunnen twee econometrische strategieën worden gebruikt. Bij de eerste strategie is de afhankelijke variabele de arbeidsmarktprestatie van afgestudeerde i aan opleiding j van ROC s met t als moment van instroom. Deze variabele kan worden geregresseerd op een indicator voor afstuderen via een CIV. Omdat de meeste studenten maar één keer afstuderen, is dit geen panelanalyse maar een cross-sectieanalyse. Er dient rekening gehouden te worden met (ongeobserveerde) verschillen tussen opleidingen en ROC's en cohorten studenten. Omdat ROC's regionaal actief zijn, hoeft er bij deze aanpak niet expliciet rekening gehouden te worden met regio's. Verder dienen er zoveel als mogelijk observeerbare kenmerken van de afstudeerder te worden opgenomen. De regressievergelijking is dan:

$$Y_{ijst} = \beta_0 + \beta_1 CIV_{ijst} + \gamma X_{ijst} + \eta_j + \sigma_s + \theta_t + u_{ijst}$$

In deze vergelijking is Y een prestatie-indicator voor de afgestudeerde, is CIV een indicator die aangeeft of de student bij een CIV is afgestudeerd, is X een reeks van controlevariabelen over eigenschappen van studenten en zijn er “fixed effects” voor opleiding, η , ROC σ en tijd θ . De β en γ zijn de te schatten coëfficiënten en u is het residu.

Bij de tweede strategie wordt een econometrisch duurmodel geschat, waarbij de afhankelijke variabele gebaseerd is op de tijd tussen het moment van afstuderen en het moment van instroom in de eerste baan. De regresoren zijn identiek aan die bij de eerste strategie. Een deel van de studenten zal al een baan hebben op het moment dat zij afstuderen. Daarom moet er bij het schatten van een duurmodel gecorrigeerd worden voor censoring. De proportie afgestudeerders met een baan direct na afstuderen kan ook als afhankelijke variabele in een probit of logit regressie fungeren. Eventueel kan deze regressie gecombineerd worden met het duurmodel.

Gevoeligheden

Probleem van deze identificatiestrategie is dat het niet mogelijk is om te controleren voor niet-waargenomen verschillen tussen afgestudeerders, omdat een afgestudeerder gekoppeld is aan slechts één opleiding. Met andere woorden, in de regressievergelijking kunnen geen “fixed effects” voor het individu worden opgenomen. Hierdoor is het belangrijk om zo goed als mogelijk te corrigeren voor observeerbare achtergrondkenmerken van de personen.

Optie 3 Enquête onder afgestudeerden, onderwijsinstellingen en bedrijven

Zowel aan studenten die hun opleiding op een CIV hebben voltooid als aan studenten die op een reguliere MBO-instelling hun opleiding hebben afgerond, kan gevraagd worden hoe snel zij een baan gevonden hebben en hoe zij de aansluiting op de arbeidsmarkt beleven. Aan MBO-instellingen kan gevraagd worden of zij zonder CIV subsidie vergelijkbare acties hadden ondernomen om de aansluiting bij het bedrijfsleven te verbeteren. Aan bedrijven kan gevraagd worden of CIV helpen om aan beter geschoold personeel te komen.

Conclusies en advies

De hypothesen over het effect van CIV kunnen getoetst worden op het niveau van opleidingen en op het niveau van afgestudeerders. Hierbij kunnen (studenten van) opleidingen met een CIV vergeleken worden met (studenten van) opleidingen zonder CIV. Bij de opleidingen met een afgewezen CIV-subsidieaanvraag kan verder nog een onderscheid gemaakt worden tussen opleidingen waarvoor alternatieve financiering is gevonden en opleidingen waarvoor dit niet het geval is.

Optie 1 kan gebruikt worden voor het toetsen van de kernhypothese (“CIV leiden tot een grotere uitstroom van bètatechnici richting topsectoren”) en voor het toetsen van hypothese 4 (“Als kansrijk beoordeelde projectplannen leiden tot een grotere uitstroom van bètatechnici richting topsectoren”). Optie 2, toetsen op opleidingsniveau, is nauwkeuriger dan optie 1 omdat er beter rekening gehouden kan worden met verschillen tussen studenten. Omdat het aantal studenten op een opleiding al geobserveerd kan worden voordat er afgestudeerders zijn, heeft de eerste strategie (moment van instroom) wel het voordeel dat de effectiviteit van CIV eerder kan worden bepaald dan bij de tweede strategie (duurmodel). Welke optie te prefereren is voor hypothesen 2, 3, 5 en 6 hangt daarmee ondermeer af van het moment waarop geëvalueerd wordt. Voor het toetsen van hypothese 7 zal een aanvullende enquête moeten worden uitgevoerd onder ROC's.

4.6 Innovatieprestatiecontracten (IPC)

Doelstelling

De regeling Innovatieprestatiecontracten (IPC) heeft als doel MKB-bedrijven aan te zetten tot meer en langdurig samenwerken om te innoveren.

Marktfalen

De MKB-bedrijven waar het IPC zich op richt zijn vaak te veel met “de waan van de dag” bezig, zijn weinig extern gericht en hebben onvoldoende organiserend vermogen en informatie (wie kan ik vertrouwen als ontwikkelpartner?) om innovatieallianties voor gezamenlijk risico en rekening aan te gaan. Er is dus zowel sprake van coördinatiefalen als van informatieasymmetrie. Bovendien zijn er in het geval van innovatieve projecten potentiële kennisspillovereffecten wanneer projecten met een laag privaat rendement zonder IPC-stimulering niet van de grond zouden kunnen komen.

Werking

IPC richt zich op het segment “toepassers” en “volgers”⁶¹ binnen het MKB. Deze groep doet niet of slechts incidenteel aan innovatie. Om deze groep toch aan te zetten tot samenwerking ten behoeve van innovatie biedt de IPC-regeling een subsidie zonder terugbetalingsverplichting maar met verplichte cofinanciering. De ratio achter deze subsidie is dat samenwerking van doorslaggevend belang kan zijn voor innovatie.

IPC kent drie onderdelen, in volgorde van ingezette middelen per onderdeel: 1) IPC-projecten, 2) middelen voor verkenning naar brancheoverstijgende samenwerking 3) middelen voor verkenning naar internationale samenwerking.

Voor IPC-projecten schrijft AgNL een à twee keer per jaar een tender uit, waarbij groepen van 10 à 20 bedrijven (al dan niet met een kennisinstelling) gezamenlijk voorstellen voor tweejarige innovatieprojecten kunnen indienen. Een penvoerder (vaak een brancheorganisatie) begeleidt de ondernemers daarbij en vraagt de subsidie aan. Ook kan kosteloos ondersteuning door Syntens worden aangevraagd. Binnen de IPC-aanvraag zijn bedrijven thematisch aan elkaar gelinkt (samenwerkende bedrijven uit het hele land) of regionaal (bedrijven uit dezelfde regio) en is er sprake van dwarsverbanden tussen de subprojecten. Deelnemers werken samen in kleinere samenwerkingsverbanden binnen zo’n IPC. Iedere ondernemer moet samenwerken met ten minste één ander bedrijf uit het IPC. IPC-projecten betreffen een innovatie van product of productieproces dat nieuw is voor de ondernemer. Het IPC staat naast industriële innovatie ook open voor diensteninnovatie.

De subsidie dekt 40% van de gemaakte projectkosten, tot een maximum van €25.000 per ondernemer. Daarnaast krijgt de penvoerder een vergoeding van €3.000 per ondernemer. De subsidie kan ingezet worden voor projectkosten als materiaal, loonkosten voor uren gemaakt in het kader van IPC-project, inhuur van externen (buiten IPC-consortium) et cetera. Eigen bedrijfskosten, zoals reiskosten, huisvesting en managementoverhead, zijn niet subsidiabel.

Onderdelen 2 en 3 betreffen het verkennen van samenwerking, zowel nationaal als internationaal. Een penvoerder onderzoekt dan de mogelijkheden voor potentiële samenwerkingsverbanden van MKB-bedrijven met één of meer ondernemingen of publieke kennisinstellingen binnen één of meer topsectoren of onderzoekt de mogelijkheid om samen met een soortgelijke (branche)organisatie over de grens te komen tot een collectief onderzoek ten voordele van de gehele branche. Aangezien het budgetair beslag van deze twee onderdelen van het IPC beperkt is, heeft de werkgroep ervoor gekozen om in het evaluatieontwerp vooral bij onderdeel 1 stil te staan.

Uitvoering

Onderstaande tekst heeft alleen betrekking op de IPC-projecten, voor de twee andere onderdelen van IPC’s wordt voor het budget het “wie het eerst komt, wie het eerst maalt” principe gehanteerd.

Huidige praktijk

In de huidige (gewijzigde) situatie rangschikt AgNL aanvragen voor IPC-projecten op drie criteria: innovatie, samenwerking (waaronder met kennisinstelling) en aansluiting bij topsector. De weging van de factoren betreft respectievelijk in procenten 40/30/30. AgNL honoreert vervolgens de hoogst geplaatste aanvragen voor zover het budget van de tender reikt. De feitelijke score op de criteria is van ondergeschikt belang aan de ranking. Er bestaat geen drempel in de vorm van

⁶¹ Dit is echter geen expliciete voorwaarde, “koplopers” worden niet per definitie uitgesloten.

een minimumscore. Zolang er geld is, wordt het weggezet in volgorde van de ranking. Hoe beter aan een criterium wordt voldaan, des te hoger de score is. De minimale eis is 20% samenwerking⁶². Bij innovatie gaat het om “nieuw voor het bedrijf”. Is het “echt nieuw” dan is de score op dit criterium hoger. Circa 2/3 deel van de 50 aanvragen door 800 bedrijven werd voor de eerste tranche in 2012 gehonoreerd. Bij de tender in 2011 kwam er voor 45 mln. aan aanvragen binnen terwijl er budget was voor 15 miljoen.

Ontwikkeling vormgeving

De selectieprocedure van IPC is door de jaren heen een aantal keer veranderd. Het budget voor IPC werd in het begin op een first-come, first-serve basis verdeeld (2008 en 2009); in 2008 werd het budget volledig uitgeput en in 2009 en 2010 is vanwege het grote aantal aanvragen direct in het begin, op dag één was het budget al overschreven, noodgedwongen gekozen voor een loting. Later is gekozen voor een generieke tender, eerst met 7 criteria (2011 en begin 2012), daarna volgens de hierboven beschreven methodiek. De 7 criteria zijn samengevat in 2 criteria en hieraan is het nieuwe criterium “aansluiting bij topsector” toegevoegd. De oorspronkelijke criteria worden dus nog wel steeds meegewogen. Bijvoorbeeld “mate van onderlinge samenwerking” en “samenwerking met een publieke kennisinstelling” waren eerst afzonderlijke criteria. Deze zijn nu ondergebracht in het criterium “kwaliteit van samenwerking”.

In 2013 zal een deel van het IPC-budget worden aangewend voor een meerjarig MKB-versterkingsplan in de topsectoren. De juridische basis voor deze regeling is de nieuwe regeling MKB Topsectoren. Het is een regeling aanvullend op de TKI-toeslag, waarvoor TKI's aanvragen indienen. Er blijft een bedrag van ruim € 7 miljoen beschikbaar voor de IPC-regeling. Deze (resterende) IPC-middelen worden via een tender (één per jaar) ter beschikking gesteld voor IPC-projecten. Rangschikking zal gebeuren op basis van de criteria “innovatie” en “samenwerking”. Het criterium “aansluiting bij topsector” vervalt, omdat er voor MKB-bedrijven in de topsectoren nu de nieuwe regeling MKB Topsectoren is. Daarnaast zal in 2013 de verkenning naar brancheoverstijgende samenwerking vervallen als onderdeel van de IPC.

Box 4.6.1 Advies toekomstige opzet IPC

De werkgroep is om advies gevraagd voor de toekomstige vormgeving van het instrument. Vanuit het oogpunt van effectmeting is het optimaal om bij selectie te kiezen voor loting omdat zo een gerandomiseerde groep ontstaat. Dit vergemakkelijkt een robuuste effectmeting. Vanuit beleidsoogpunt is een tender echter te prefereren. Om een brug te slaan tussen beide belangen adviseert de werkgroep om in de toekomst het selectieproces als volgt vorm te geven:

- Organiseer een tender die resulteert in twee groepen: de groep afgewezenen en de groep potentiële toekenningen. Binnen de tweede groep wordt voor een loting gekozen ten behoeve van effectmeting.
- Om de ‘oneerlijkheid’ van loting te verkleinen en prikkels voor het afleveren van goede plannen in stand te houden, kan gekozen worden voor een gewogen loting, waar voorstellen een hogere kans maken om ingeloot te worden naarmate hun voorstel beter is beoordeeld in de tender. De weegkansen in de loting vormen dan de propensity score.

Budget

Het budgettaire beslag van IPC bedraagt in 2012 €41,2 miljoen. In 2013 zal het budget dalen naar ruim € 7 miljoen, vanwege de aanwending van een bedrag van € 15 miljoen voor de nieuwe regeling MKB Topsectoren. In het verleden zijn de budgetten steeds uitgeput.

Looptijd evaluatieperiode

De IPC-regeling zal medio 2015 geëvalueerd worden.

⁶²Iedere deelnemer aan IPC stelt een eigen begroting voor zijn eigen IPC-project op. Minimaal 20% van deze begroting (in uren of geld) moet gereserveerd zijn voor samenwerking met andere bedrijven.

Box 4.6.2 Aanvullende observaties werkgroep over evaluatiemogelijkheden IPC

De IPC-regeling is vaak van opzet veranderd. Dit biedt extra mogelijkheden vanuit het oogpunt van evaluatie. Zo kan bijvoorbeeld de vraag worden beantwoord welk verdelingsmechanisme het beste werkt. Een tender heeft voor- en nadelen. Enerzijds dwingt het partijen betere voorstellen in te dienen, anderzijds leidt het tot een kostbaar uitvoeringsproces. Terwijl loting qua uitvoering goedkoop is, bestaat er geen prikkel om kwalitatief goede voorstellen in te dienen. Met aanvullend onderzoek, waarin data over de gehele looptijd van het IPC worden gebruikt, kan de toegevoegde waarde van een beoordeling van een IPC-aanvraag door AgNL worden geanalyseerd.

Hypotheses

In de hypotheses wordt een onderscheid gemaakt tussen toe- en afgewezen IPC-aanvragen om het effect van het ontvangen subsidiegeld te kunnen onderscheiden van het effect van de aanvraagprocedure.

Kernhypotheses

- 1a) Een toegewezen IPC-aanvraag leidt tot meer en langdurigere samenwerking tussen de betrokken bedrijven in vergelijking tot bedrijven waarvan de IPC-aanvraag is afgewezen.
- 1b) Een afgewezen IPC-aanvraag leidt tot meer en langduriger samenwerking tussen betrokken bedrijven in vergelijking tot bedrijven die geen IPC-aanvraag hebben ingediend.

Hypotheses tweede- en derde-orde effecten

- 2a) Een toegewezen IPC-aanvraag leidt tot meer innovatie bij de betrokken bedrijven in vergelijking tot bedrijven waarvan de IPC-aanvraag is afgewezen.
- 2b) Een afgewezen IPC-aanvraag leidt tot meer innovatie tussen betrokken bedrijven in vergelijking tot bedrijven die geen IPC-aanvraag hebben ingediend.
- 3a) Bedrijven betrokken bij een toegewezen IPC-aanvraag groeien sneller en hebben een hogere productiviteit dan bedrijven waarvan de IPC-aanvraag is afgewezen.
- 3b) Bedrijven betrokken bij een afgewezen IPC-aanvraag groeien sneller en hebben een hogere productiviteit dan bedrijven die geen IPC-aanvraag hebben ingediend.

Hypotheses voor toetsing selectieprocedure

- 4) Een tender leidt tot een hogere effectiviteit van de IPC-regeling dan loting (meer kansrijke projecten door competitie-element).

Optionele hypothese voor toetsing effect verkenning

- 5) IPC-aanvragen met behulp van verkenning scores beter dan aanvragen zonder deze verkenning.

Indicatoren

1. Indicatoren over samenwerking tussen bedrijven. Waarschijnlijk zullen deze indicatoren geënkquêteerd moeten worden.
2. (Arbeids-)productiviteit per bedrijf
3. Uitgaven aan S&O per bedrijf per jaar
4. Aantal WBSO-uren per bedrijf per jaar
5. Indicatoren voor innovatie per bedrijf: nieuwe producten, processen of patentaanvragen
6. Overige bedrijfskarakteristieken: omzet, aantal werknemers, omvang kapitaal, leeftijd, etc.
7. Topsector per bedrijf

Tabel 6 Overzicht indicatoren bij evaluatieontwerp IPC

Hypotheses	Indicatoren		Databronnen
	Afhankelijk	Onafhankelijk	
1a) Een toegewezen IPC-aanvraag leidt tot meer en langdurigere samenwerking tussen de betrokken bedrijven in vergelijking tot bedrijven waarvan de IPC-aanvraag is afgewezen.	<ul style="list-style-type: none"> Mate van samenwerking 	<ul style="list-style-type: none"> Dummy IPC Ja/Nee 	<ul style="list-style-type: none"> Enquête AgNL
1b) Een afgewezen IPC-aanvraag leidt tot meer en langduriger samenwerking tussen betrokken bedrijven in vergelijking tot bedrijven die geen IPC-aanvraag hebben ingediend.	<ul style="list-style-type: none"> Mate van samenwerking 	<ul style="list-style-type: none"> Dummy IPC Ja/Nee Dummy afgewezen IPC Ja/Nee 	<ul style="list-style-type: none"> Enquête AgNL
2a) Een toegewezen IPC-aanvraag leidt tot meer innovatie bij de betrokken bedrijven in vergelijking tot bedrijven waarvan de IPC-aanvraag is afgewezen.	<ul style="list-style-type: none"> Nieuwe producten & processen Octrooien en patenten Uitgaven aan R&D per bedrijf 	<ul style="list-style-type: none"> Dummy IPC Ja/Nee 	<ul style="list-style-type: none"> AgNL WBSO-data NL Octrooiencentrum CBS
2b) Een afgewezen IPC-aanvraag leidt tot meer innovatie tussen betrokken bedrijven in vergelijking tot bedrijven die geen IPC-aanvraag hebben ingediend.	<ul style="list-style-type: none"> Nieuwe producten & processen Octrooien en patenten Uitgaven aan R&D per bedrijf 	<ul style="list-style-type: none"> Dummy IPC Ja/Nee Dummy afgewezen IPC Ja/Nee 	<ul style="list-style-type: none"> AgNL WBSO-data NL Octrooiencentrum CBS
3a) Bedrijven betrokken bij een toegewezen IPC-aanvraag groeien sneller en hebben een hogere productiviteit dan bedrijven waarvan de IPC-aanvraag is afgewezen.	<ul style="list-style-type: none"> (Arbeids-) productiviteit Groei van bedrijven 	<ul style="list-style-type: none"> Dummy IPC Ja/Nee 	<ul style="list-style-type: none"> AgNL WBSO-data NL octrooiencentrum CBS
3b) Bedrijven betrokken bij een afgewezen IPC-aanvraag groeien sneller en hebben een hogere productiviteit dan bedrijven die geen IPC-aanvraag hebben ingediend.	<ul style="list-style-type: none"> (Arbeids-) productiviteit Groei van bedrijven 	<ul style="list-style-type: none"> Dummy IPC Ja/Nee Dummy afgewezen IPC Ja/Nee 	<ul style="list-style-type: none"> AgNL WBSO-data NL octrooiencentrum CBS
4) Een tender leidt tot een hogere effectiviteit van de IPC-regeling dan loting (meer kansrijke projecten door competitie-element).	<ul style="list-style-type: none"> Mate van kansrijkheid project 	<ul style="list-style-type: none"> Dummy tender Ja/Nee 	<ul style="list-style-type: none"> AgNL
Controlevariabelen	<ul style="list-style-type: none"> WBSO-uren per bedrijf Bedrijfskenmerken (omzet, aantal werknemers, leeftijd bedrijf etc.) 		<ul style="list-style-type: none"> AgNL CBS

Optie 1: (Gewogen) loting

Experimentele en controlegroep

In 2009 en 2010 is er geloot bij de toekenning van IPC's. Bij een loting bestaat de experimentele groep uit de ingelote bedrijven. De controlegroep wordt gevormd door de uitgelote bedrijven. Als de loting gewogen is, dan worden observaties gewogen met de kans op inloting. Met een regressieanalyse kan eventueel gecontroleerd worden of de trekkingen van de loting willekeurig zijn geweest. Deze optie is van toepassing voor hypothesen 1a, 2a en 3a.

Identificatiestrategie

Het effect van IPC's op de doelvariabelen volgt direct uit het verschil tussen het (gewogen) gemiddelde van de experimentele groep en het (gewogen) gemiddelde van de controlegroep.

Econometrische specificatie en methoden

Doordat loting de toekenning tot experimentele groep en controlegroep bepaalt, is een eerste vergelijking van gemiddelden van de doelvariabele tussen beide groepen al een goede schatting van het effect van IPC's op doelvariabelen. In termen van regressieanalyse betekent dit een regressie van de doelvariabele op een indicator voor IPC, waarbij geconditioneerd kan worden op karakteristieken van bedrijven zoals omzet, aantal werknemers, leeftijd. Deze variabelen zouden echter geen effect mogen hebben, omdat de bedrijven sterk vergelijkbaar zijn vanwege de toekenning aan experimentele en controlegroep door middel van loting. Hiervoor kan OLS worden gebruikt door de doelvariabele te regresseren op een indicator voor IPC en alle overige achtergrondkenmerken..

Eventueel kan dit worden gespecificeerd als een difference-in-difference vergelijking, waarbij informatie voorafgaande aan de IPC-toekenning kan worden gebruikt om tot een trendmatige vergelijking te komen tussen de controlegroep en de interventiegroep. Vanwege de loterij is dit echter niet strikt noodzakelijk.

Gevoeligheden

Anticipatie: Een aangekondigde ongewogen loting geeft bedrijven een prikkel om ook slechte voorstellen in te dienen. Bij een gewogen loting zal dit anticipatie-effect minder sterk zijn.

Optie 2: Toekenning op volgorde van aanvraag

Experimentele en controlegroep

In 2008 werden IPC's toegekend op volgorde van aanvraag totdat het budget was uitgeput. In dit geval bestaat de experimentele groep uit bedrijven waaraan een IPC is toegekend. De controlegroep bestaat uit afgewezen bedrijven. Zowel de experimentele als de controlegroep kunnen beperkt worden tot aanvragen rond de datum waarop het budget uitgeput raakte. Deze optie is van toepassing voor hypothesen 1a, 2a en 3a.

Identificatiestrategie

Identificatie is gebaseerd op het idee dat beide groepen bedrijven onderling sterk vergelijkbaar zijn, behalve dat sommige bedrijven een IPC hadden toegekend gekregen en andere vergelijkbare bedrijven niet. Het is goed mogelijk dat bedrijven verwachtingen hebben gehad over de datum waarop het budget uitgeput zou kunnen raken. Hierdoor zou het kunnen zijn dat aanvragen die in het begin van het jaar gedaan zijn, systematisch verschillen van aanvragen, die later in het jaar zijn ingediend. Het zou kunnen zijn dat vroege aanvragen van een hogere kwaliteit zijn dan latere aanvragen omdat vroege aanvragen afkomstig zijn van een slagvaardige groep bedrijven. Er zijn ook andere redenen denkbaar voor een verschil tussen vroege en latere aanvragen.

Om het effect van verwachtingen van bedrijven over de sluitingsdatum te beperken, kunnen de experimentele groep en de controlegroep beperkt worden tot aanvragen rond de sluitingsdatum. Deze methode is in hoofdstuk 2 omschreven als *regression discontinuity*.

Econometrische specificatie en methoden

Vergelijking van het gemiddelde van de experimentele groep met het gemiddelde van de controlegroep geeft al een goede schatting van het effect van IPC's op doelvariabelen. Bij het selecteren van de experimentele en controlegroep is er een afweging te maken: het effect van verwachtingen wordt kleiner naarmate de steekproef dichter op de sluitingsdatum

zit, maar tegelijkertijd neemt het aantal observaties, en daarmee de nauwkeurigheid van het geschatte effect, af. Er kan verder worden geconditioneerd op karakteristieken van bedrijven zoals omzet, aantal werknemers, leeftijd, maar deze variabelen zouden juist geen effect mogen hebben, omdat de bedrijven sterk vergelijkbaar zouden zijn. Mochten deze variabelen toch effect hebben, dan valt het te betwijfelen of regression discontinuity wel een geschikte schattingsmethode is voor dit specifieke probleem.

Gevoeligheden

Selectie: Sterke verwachtingen over de sluitingsdatum bij (een deel van) de IPC-aanvragers kunnen tot een over- of onderschatting van het effect van IPC leiden. Bovendien kunnen de bedrijven voor en na toekenning sterk heterogeen zijn. Als de bedrijven voor de sluitingsdatum meer (minder) productief zijn, dan zal het leiden tot een overschatting (onderschatting) van het effect van IPC.

Optie 3: Rangorde

Experimentele en controlegroep

In 2011 en 2012 zijn IPC-aanvragen beoordeeld door Agentschap NL. De experimentele groep bestaat uit gehonoreerde aanvragen met een relatief lage (maar voldoende) beoordeling en de controlegroep bestaat uit afgewezen aanvragen met een relatief goede (maar onvoldoende) beoordeling. Deze optie is van toepassing voor hypothesen 1a, 2a en 3a.

Identificatiestrategie

Om het effect van IPC's op doelvariabelen te schatten, worden gehonoreerde en afgewezen aanvragen met elkaar vergeleken. Vanwege de beoordeling van Agentschap NL zullen gehonoreerde aanvragen van een hogere kwaliteit zijn dan afgewezen aanvragen. Om het verschil in kwaliteit tussen de experimentele groep en de controlegroep zo klein mogelijk te maken, wordt de steekproef beperkt tot aanvragen die net wel of net niet een voldoende oordeel hebben gekregen. Deze methode is in hoofdstuk 2 omschreven als *regression discontinuity*, waarbij de effecten lokaal worden gemeten.

Een alternatieve identificatiestrategie is om gebruik te maken van de informatie omtrent toekenning en hier een *propensity score* methode toe te passen. Voor 2011 en 2012 zijn er verschillende criteria voor toekenning, die gebruikt kunnen worden om de propensity score te schatten. Het is hier echter onduidelijk of de criteria onderscheidend genoeg zijn om de propensities nauwkeurig genoeg te kunnen schatten.

Econometrische specificatie en methoden

Bij de regression discontinuity is bij het selecteren van de experimentele en controlegroep een afweging te maken: het effect van kwaliteitsverschillen wordt kleiner naarmate een waarneming dichterbij de minimale beoordeling ("rapportcijfer 5,5") zit, maar tegelijkertijd neemt het aantal observaties, en dus de nauwkeurigheid van de geschatte effecten, af. Rondom de discontinuïteit mag er binnen de selectie voor wat betreft de overige geobserveerde variabelen geen substantieel verschil zijn tussen de geselecteerde en niet-geselecteerde bedrijven. Om de verschillen in karakteristieken tussen aanvragen in de experimentele groep en de controlegroep (verder) te beperken, kan er geconditioneerd worden op omzet, aantal werknemers, leeftijd, topsector en andere niet-doelvariabelen. Nadeel hiervan is echter dat de precisie van de schattingen (de t-waarden) minder wordt door opname van deze variabelen. Alleen als deze variabelen een statistisch significant effect hebben dienen ze te worden opgenomen. Het kan zijn dat het eindoordeel van Agentschap NL over een aanvraag dichotoom is, dat wil zeggen "voldoende" of "onvoldoende". In dit geval is een extra stap vereist om aanvragen te selecteren die net wel of net niet zijn toegekend. Het kan zijn dat er een "rapportcijfer" te reconstrueren is op basis van subcriteria en hun gewichten.

Bij de propensity score kan de waarschijnlijkheid van een toekenning (ten opzichte van een afgewezen aanvraag) geschat worden met behulp van scores op subcriteria.

Gevoeligheden

Bezien moet worden of er voldoende observaties voor vergelijkbare bedrijven beschikbaar zijn rondom de minimale beoordeling.

Optie 4: Meta-optie selectiemethoden: Dummy's voor selectiemethoden

Experimentele en controlegroep

In hypothese 4 wordt de vraag gesteld of een tender effectiever is dan loting. Om deze hypothese te onderzoeken wordt de effectiviteit van IPC's vergeleken tussen 2009/2010 en 2011/2012. In 2009/2010 is gebruik gemaakt van loting terwijl de toekenning in 2011/2012 is gebeurd op basis van rangorde. De groep bestaat uit de toegekende of afgewezen aanvragen. In principe worden de propensities bepaald zoals onder optie 3 beschreven is. Vervolgens worden de geschatte propensities bij elkaar genomen in een bestand van alle aanvragen van 2009/2010 en 2011/2012 gezamenlijk.

Identificatiestrategie

Variatie van de propensities over de beide periodes wordt geschat door een indicatorvariabele voor periodes in de propensity score mee te nemen. Hiermee kan het patroon over de periodes heen worden beoordeeld. Het idee is dat bedrijven worden vergeleken over de jaren heen. Dus dat toegekende bedrijven die hebben meegedaan aan de loting worden vergeleken met de toegekende bedrijven die hebben meegedaan aan een tender. Ook kunnen de afgewezen bedrijven (van loting) worden vergeleken met die van een tender.

Econometrische specificatie en methoden

In de eerste ronde worden de propensity scores geschat. Vervolgens worden in de tweede ronde de propensity scores vergeleken door een indicatorvariabele voor aanvraagronde (loting/tender) op te nemen.

Gevoeligheden

De mogelijkheid om propensities te bepalen hangt af van de beschikbaarheid van vergelijkbare scores op subcriteria in de verschillende aanvraagrondes (met lotingen en tenders).

Optie 5: Meta-optie selectiemethoden: Ordening selectiemethoden

Experimentele en controlegroep

Bij deze optie worden de geselecteerde aanvragen voor 2011 en 2012 onderling vergeleken, zonder dat gebruik wordt gemaakt van de afgewezen aanvragen. Het achterliggende idee is dat een geselecteerd IPC in een minder strenge selectieronde wel geselecteerd kan zijn, maar dat dit IPC in een strenge selectieronde niet geselecteerd zou kunnen worden. Dus de geselecteerde bedrijven worden naar mate van strengheid beoordeeld. Deze optie is een verdieping binnen hypothese 4.

Identificatiestrategie

Per cut-off point een propensity score toepassen. De geschatte propensity scores worden gepooled voor de verschillende aanvraagrondes. De verschillende rondes op mate van strengheid ordenen. Vervolgens worden aanvragen boven het cut-off point van de strengere selectie vergeleken met de aanvragen van de iets minder strenge selectie (die bij de strengere selectie niet geselecteerd zouden worden en dus onder het cut-off point vallen van die strengere selectie).

Econometrische specificatie en methoden

Er moet een propensity score worden uitgevoerd door een geselecteerd bedrijf uit jaar 2012 te vergelijken met geselecteerd bedrijf uit het jaar 2011. Hiervoor wordt dus een propensity score bepaald op basis van informatie van geselecteerde bedrijven. Vervolgens worden de propensities uit 2012 vergeleken met die van 2011. Indien er geen verschillen zijn, dan heeft een strengere selectie niet geleid tot een beter resultaat van de afhankelijke variabele.

Gevoeligheden

Bedrijven kunnen besluiten om geen aanvraag te doen als bekend is dat de selectie in een specifieke aanvraagronde strenger is geworden.

Optie 6: Surveymethode

Identificatiestrategie

IPC-aanvragers kunnen rechtstreeks gevraagd worden naar hun mening over het IPC en naar veranderingen in hun gedrag door het IPC. Alle hypothesen met uitzondering van hypothese 4 kunnen met een enquête getoetst worden. In het bijzonder voor de toetsing van hypothesen 1b, 2b en 3b lijkt enquêtering belangrijk te zijn. Bij die hypothesen worden bedrijven met een afgewezen IPC-aanvraag vergeleken met bedrijven die geen IPC-aanvraag hebben ingediend.

Er kunnen vragen worden gesteld over zowel het verleden, het heden als de toekomstverwachting. In de enquête zal een vraag worden opgenomen over de winstverwachting van bedrijven in hun sector, waarbij moet worden aangegeven voor welke periode dit is.

Econometrische specificatie en methoden

De uitkomsten van een enquête hangen deels af van de manier waarop vragen geformuleerd zijn. Een pilotstudie met open interviews is nodig om de vragenlijst te valideren. De consistentie van een antwoord dient getoetst te worden door belangrijke vragen in een iets andere vorm te herhalen. “Markers” dienen te worden gebruikt om de uitkomsten te normaliseren. Een werknemer zal bijvoorbeeld mogelijk anders antwoorden dan een directeur.

Gevoeligheden

Uitkomsten: Antwoorden geven niet altijd een realistisch beeld van de werkelijkheid. Met name het geven van sociaal wenselijke antwoorden kan de uitkomst verstoren.

Conclusies en advies

Optie 1 is de te prefereren strategie, omdat het gebaseerd is op loting. Bovendien geeft het geen lokale schatting, maar een schatting over de hele populatie. De opties gebaseerd op de propensity score methode lijken ook veelbelovend te zijn, mede omdat ontwikkelingen over jaren heen kunnen worden beoordeeld. De discontinuity methode biedt slechts een lokale schatting. Bij de hypothesen 1, 2 en 3 kunnen telkens twee groepen bedrijven worden vergeleken. De eerste vergelijking betreft bedrijven met een toegewezen IPC-aanvraag versus bedrijven waarvan de IPC-aanvraag is afgewezen (hypothesen 1a, 2a en 3a). De tweede vergelijking gaat over bedrijven betrokken bij een afgewezen IPC-aanvraag versus bedrijven die geen IPC-aanvraag hebben ingediend (hypothesen 1b, 2b, 3b). Voor de eerste vergelijking lijkt het datamateriaal eenvoudiger beschikbaar te zijn. Bij de tweede vergelijking lijkt een aanvullende survey onder bedrijven zonder IPC-aanvraag nodig te zijn.

Ten behoeve van de toekomstige vormgeving van het instrument adviseert de werkgroep om in de toekomst het selectieproces als volgt vorm te geven:

- Organiseer een tender die resulteert in twee groepen: de groep afgewezenen en de groep toekenningen. Binnen de tweede groep wordt voor een loting gekozen ten behoeve van effectmeting.
- Om de ‘oneerlijkheid’ van loting te verkleinen en prikkels voor het afleveren van goede plannen in stand te houden kan gekozen worden voor een gewogen loting, waar voorstellen een hogere kans maken om ingeloot te worden naarmate hun voorstel beter is beoordeeld in de tender. De weegkansen in de loting vormen dan de propensity score.

4.7 Topconsortia voor Kennis en Innovatie (TKI)

Doelstelling

Een TKI heeft als doel het verbeteren van de aansluiting van publieke R&D-investeringen op de innovatie-inspanningen in de topsectoren. Door de samenhang in de kennisketen te verhogen, wordt private R&D uitgelokt die uiteindelijk zal resulteren in meer nieuwe (innovatieve) producten en diensten. De idee is dat een TKI zorgt voor een hoger rendement op publieke onderzoeksmiddelen doordat meer vraagsturing in het kennissysteem leidt tot meer commerciële en maatschappelijke toepassing van onderzoeksresultaten. De doelgroep van TKI's bestaat uit bedrijven en kennisinstellingen (universiteiten en hogescholen, TNO, GTI's, KNAW-instituten) eventueel aangevuld met meebetalende overheden.

Marktfalen

De legitimatie van de TKI's komt ten eerste voort uit het bestaan van informatieasymmetrie. In de keten van kennis, kunde, kassa ontbreekt vaak voldoende informatie over wie welk onderzoek het best kan uitvoeren en wie de resultaten van dit onderzoek het beste naar de markt kan brengen. Een TKI beoogt deze informatiekloof te overbruggen. Ten tweede adresseert een TKI coördinatieproblemen. Door onderzoek binnen een TKI te beleggen worden onderzoeksactiviteiten niet gedupliceerd, wat tot een efficiëntere allocatie van middelen leidt. Als laatste wordt samenwerking in pre-concurrentieel onderzoek gestimuleerd door de TKI-toeslag. Deze samenwerking kan worden gehinderd door de aanwezigheid van positieve externe effecten in de vorm van kennisillovers.

Werking

Een TKI is een meerjarig samenwerkingsverband tussen private en publieke partijen, waarbij op programmatische basis wordt samengewerkt op het gebied van onderzoek en innovatie op terreinen zoals belegd binnen innovatiecontracten van de topsectoren. Bij een TKI zijn minimaal 3 ondernemingen en 3 publiek gefinancierde onderzoeksorganisaties betrokken. Bedrijven committeren zich in een TKI financieel aan een onderzoeksprogramma. Per topsector zijn er één of meerdere TKI's. Per 1 september 2012 zijn er voor de 9 topsectoren 19 TKI's opgericht. Deze zullen als paraplu dienen, waar diverse samenwerkingsverbanden onder komen te hangen. Een TKI biedt maatwerk per topsector, waardoor TKI's een verschillende invulling per topsector kunnen krijgen. TKI's staan open voor iedereen die zich wil aansluiten.

Essentie van de samenwerking tussen bedrijven en publieke kennisinstellingen binnen een TKI is dat voor gezamenlijke rekening en risico onderzoek- en innovatieactiviteiten worden verricht. Bij een project is minimaal één bedrijf en één kennisinstelling betrokken. Dit gebeurt over de hele innovatieketen van fundamenteel onderzoek tot en met experimentele ontwikkeling, steeds in precompetitieve samenwerking tussen de private en publieke partijen. TKI's maken het mogelijk voor bedrijven om met een ex ante inleg en duidelijke intellectueel eigendomsregeling een deel van het risico te poolen op het investeren in precompetitieve R&D in samenwerking met een kennisinstelling. Het onderzoek wordt uitgevoerd door universiteiten, technologische instituten, hogescholen en bedrijven (waaronder private onderzoeksinstellingen). Een TKI kan voor het uitzetten van onderzoek gebruik maken van NWO en TNO. Een TKI zorgt voor de regie, netwerkvorming en kennis(ver)deling van de onderzoeks- en innovatieactiviteiten.

TKI-toeslag

Om bedrijven aan te zetten tot publiek-private samenwerking en zoveel mogelijk van hun gezamenlijke activiteiten in een TKI onder te brengen bestaat hiervoor een financiële prikkel, namelijk de TKI-toeslag. Voor de TKI-toeslag komen uitsluitend TKI's in aanmerking die zijn opgenomen in de begrotingswet van het jaar waarop de aanvraag voor TKI-toeslag betrekking heeft. De toeslag wordt aangevraagd door het TKI zelf, niet door afzonderlijke bedrijven en kennisinstellingen. De toeslag gaat dan ook niet naar individuele bedrijven, maar naar de TKI. De TKI-toeslag moet vervolgens voor een nieuw project gebruikt worden dan voor het project waarvoor de private cashbijdrage is verleend, het TKI beslist hierover. Een TKI gebruikt de toeslag samen met de andere bijdragen voor activiteiten in het werkprogramma over de gehele keten, bijvoorbeeld voor aio-onderzoek bij universiteiten, toegepast onderzoek of valorisatie. In kosten kunnen zaken worden meegenomen als: personeelskosten, apparatuur, kosten gebouwen, kosten contractonderzoek en extra algemene vaste kosten, mits deze toe te rekenen zijn aan het onderzoeksprogramma. De hoogte van de grondslag wordt bepaald door de jaarlijkse cash-bijdragen van bedrijven aan het onderzoeksprogramma van de TKI.

Uitvoering

Een TKI heeft een onderzoeksprogramma met een minimale omvang van €5 miljoen per jaar. De kwantitatieve inzet van private partijen binnen de TKI (cash en in kind) bedraagt in 2015 ten minste 40% (2014: 35%, 2013: 30%). Een TKI zorgt ervoor dat het intellectuele eigendom van privaat publieke onderzoeksprojecten naar rato van de private bijdrage aan een onderzoeksproject wordt verdeeld. Als een partij het exclusieve recht wil verwerven, dan kunnen bedrijven elkaar uitkopen.

Een TKI kan bij AgNL TKI-toeslag aanvragen. Uitsluitend de hoogte van de private cashbijdrage voor samenwerkingsprojecten van het TKI-programma aan een onderzoeksorganisatie bepaalt de hoogte van de toeslag. Over de private cashbijdrage aan een onderzoeksorganisatie krijgt een bedrijf geen RDA of WBSO (er is hier sprake van uitbesteed onderzoek), terwijl een bedrijf voor intern uitgevoerde R&D wel in aanmerking komt voor RDA en WBSO. Aanvragen voor TKI-toeslag worden altijd gehonoreerd mits daarbij aan de voorwaarden van het O&O&I-steunkader is voldaan. De TKI-toeslag betreft een openeinderegeling. Dat wil zeggen dat wanneer de private bijdrage hoger uitvalt het percentage in het opvolgende jaar naar beneden kan worden bijgesteld.

Budget

Voor de TKI-toeslag is vanaf 2013 €90 miljoen beschikbaar⁶³. Het toeslagpercentage is 25%. Voor de eerste € 20.000 cashbijdrage per bedrijf per jaar aan het onderzoeksprogramma van een TKI is het toeslagpercentage 40%. Hiermee wordt de TKI gestimuleerd veel, veelal kleinere bedrijven te betrekken bij het onderzoeksprogramma. Om budget uit te putten is ongeveer €360 miljoen aan private bijdragen nodig.

Looptijd evaluatieperiode

In 2015 staat een tussenevaluatie van de TKI's gepland. Ook is dan een eerste evaluatie van de TKI-toeslag voorzien.

Hypotheses

Kernhypotheses

- 1) Een TKI zorgt voor meer publiek-private samenwerking in de kennisketen

Hypothese ter ondersteuning kernhypothese

- 2) Een TKI zorgt voor betere aansluiting van publieke R&D-investeringen op de innovatie-inspanningen in de topsectoren.
- 3) Een TKI leidt tot meer private R&D.

Hypothese tweede-orde effect

- 4) Inzet van publieke onderzoeksmiddelen via een TKI leidt tot meer commerciële en maatschappelijke toepassing van onderzoeksresultaten.

Hypothese voor toetsing TKI-toeslag

- 5) De TKI-toeslag stimuleert de private R&D-investeringen in de TKI's.

Benodigde gegevens

Beschikbaarheid van data

De TKI zelf (of Agentschap NL) zijn een belangrijke bron van data als het gaat om de uitgaven die bedrijven en instellingen doen binnen een TKI-project. In principe zouden de TKI ook de gegevens over octrooien en ander intellectueel eigendom dat voortkomt uit TKI-projecten moeten kunnen aanleveren.

⁶³Volgens het regeerakkoord Rutte/Samson komt hier nog eens €110 miljoen bij. Op het moment van schrijven was de uitwerking hiervan echter nog niet bekend. Mogelijk geeft deze uitbreiding nog verdere aanknopingspunten voor de constructie van een evaluatieontwerp.

Tabel 7 Overzicht indicatoren bij evaluatieontwerp TKI

Hypotheses	Indicatoren		Databronnen
	<i>Afhankelijk</i>	<i>Onafhankelijk</i>	
1) Een TKI zorgt voor meer publiek-private samenwerking in de kennisketen.	<ul style="list-style-type: none"> • Meer PPS-publicaties • Meer pps-octrooien 	<ul style="list-style-type: none"> • Indicator TKI Ja/Nee 	<ul style="list-style-type: none"> • TKI • PATSTAT • CWTS • AgNL/OCNL
2) Een TKI zorgt voor betere aansluiting van publieke R&D-investeringen op de innovatie-inspanningen in de topsectoren..	<ul style="list-style-type: none"> • Geen kwantitatieve indicator beschikbaar 	<ul style="list-style-type: none"> • Geen kwantitatieve indicator beschikbaar 	<ul style="list-style-type: none"> • Enquête
3) Een TKI leidt tot meer private R&D.	<ul style="list-style-type: none"> • Private R&D-investeringen 	<ul style="list-style-type: none"> • Indicator TKI Ja/Nee 	<ul style="list-style-type: none"> • TKI • AgNL/WBSO • PATSTAT
4) Inzet van publieke onderzoeksmiddelen via een TKI leidt tot meer commerciële en maatschappelijke toepassing van onderzoeksresultaten.	<ul style="list-style-type: none"> • Meer octrooien door kennis-instellingen 	<ul style="list-style-type: none"> • Indicator TKI Ja/Nee 	<ul style="list-style-type: none"> • TKI • AgNL/OCNL • PATSTAT
5) De TKI-toeslag stimuleert de private R&D-investeringen in de TKI's.	<ul style="list-style-type: none"> • Private R&D-investeringen 	<ul style="list-style-type: none"> • Indicator TKI Ja/Nee Dummy 	<ul style="list-style-type: none"> • TKI • AgNL/WBSO • CBS
Controlevariabelen [niet uitputtend]	<ul style="list-style-type: none"> • Bedrijfskenmerken • Kenmerken kennisinstellingen 		<ul style="list-style-type: none"> • PS/SFO • CIS

Het CBS heeft verschillende gegevensbestanden over bedrijven die gebruikt kunnen worden voor de evaluatie van de TKI. Gegevens van aanvragers kunnen via KvK-nummers gekoppeld worden aan het ABR en vervolgens met andere CBS-bronnen. Hier is (ruime) ervaring mee bij bijvoorbeeld de WBSO-aanvragen. De productiestatistieken (PS) zijn een bron voor gegevens over omzet, productiviteit, werkgelegenheid etcetera. Een beperking daarbij is het steekproefkarakter van de PS voor kleinere ondernemingen. Hetzelfde geldt voor de Statistieken Financiën Ondernemingen (SFO).

De innovatie-enquête (CIS) en het R&D-onderzoek van het CBS bevatten indicatoren over R&D- en innovatie-activiteit. Ook hier geldt de beperking dat het om steekproeven gaat. De enquêtes zijn tweejaarlijks en de microdata komen pas 18 maanden na dato beschikbaar. In de steekproef zit 0% van de bedrijven met minder dan 10 werknemers, ongeveer 10% van de bedrijven met 10-50 werknemers, ongeveer 20% van de bedrijven met 50-100 werknemers en 100% van de bedrijven met meer dan 100 werknemers. WBSO-data van AgNL kunnen worden gebruikt voor een brede dekking over bedrijven van gegevens over R&D (de komende jaren kunnen ook RDA-data van AgNL worden benut).

Agentschap NL en CBS beschikken over de patentaanvragen van Nederlandse ondernemingen binnen Nederland en binnen de EU. Deze beslaan voor een heel groot deel alle patenten zoals in PATSTAT en in tegenstelling tot PATSTAT is deze informatie aan het ABR gekoppeld.

Gegevens over kennisinstellingen zijn minder goed beschikbaar. Het CPB heeft een bestand met wetenschappelijk medewerkers tot zijn beschikking, afkomstig van de VSNU met aanvullende gegevens van de individuele universiteiten. Het tijdvak is plusminus 2002-2008. Voor hergebruik is toestemming van de VSNU nodig.

Er bestaat een NWO-bestand van aanvragers van een subsidie uit de Vernieuwingsimpuls. Dit bestand wordt (geanonimiseerd) binnen de beveiligde microdataomgeving van het CBS door het CPB gebruikt.

Het CWTS heeft wetenschappelijke publicaties op naam. Deze publicaties zijn moeilijk te koppelen aan CBS-data, omdat het CBS vooral koppelt op geslacht/geboortedatum/adres. CBS Gemeentelijke Basisadministratie persoonsgegevens (GBA) data bevatten geen namen.

In bestanden met octrooigegevens (zoals PATSTAT) staat zowel de naam van de uitvinder als van de instellingen en/of bedrijven die het octrooi aanvragen. Behalve voor bedrijven zijn deze gegevens op dit moment onvoldoende gestandaardiseerd en dus nog niet koppelbaar met andere data.

Optie 1 Vergelijking TKI-bedrijven met andere bedrijven

Experimentele en controlegroep

Bedrijven kunnen per TKI worden vergeleken met vergelijkbare bedrijven in dezelfde economische sector die niet in een TKI participeren. Beide groepen bedrijven moeten hierbij worden gevolgd vanaf de periode voorafgaande aan de TKI. Deze optie is van toepassing op hypothesen 1 en 3.

Identificatiestrategie

De trendmatige ontwikkeling van de TKI-bedrijven dient te worden vergeleken met die van de niet-TKI bedrijven. Dit lijkt op een difference-in-difference benadering, waarbij verondersteld wordt dat in de periode voorafgaande aan de TKI beide groepen van bedrijven dezelfde trendmatige ontwikkeling hadden.

Econometrische specificatie en methoden

Het gegevensbestand bevat informatie van de periode voorafgaande aan introductie van TKI. De afhankelijke variabele wordt geresseerd op een indicator, TKI_{it} , die gelijk is aan 1 vanaf het moment dat een bedrijf deel uitmaakt van een TKI. Het effect van TKI kan worden vastgesteld door de coëfficiënt van deze indicator. Verder dient er te worden gecorrigeerd voor tijd, zodat voor elk jaar een afzonderlijke indicatorvariabele θ wordt opgenomen. Voor verschillen tussen bedrijven wordt rekening gehouden door voor elk bedrijf een indicatorvariabele op te nemen. Dit wordt ook wel aangeduid als de “fixed effects” η). De regressievergelijking is dan:

$$Y_{it} = \beta_0 + \beta_1 TKI_{it} + \eta_i + \theta_t + u_{it}$$

In deze vergelijking is Y de prestatie van bedrijf i op tijdstip t , de β 's zijn de te schatten coëfficiënten en u is de storingsterm.

Gevoeligheden

Er is zelfselectie bij bedrijven en kennisinstellingen die een TKI vormen. Daardoor zou het wel eens kunnen zijn dat de veronderstelling van een gemeenschappelijke trend onjuist is. Verder kan worden onderzocht of de gevonden effecten over de loop van de tijd sterker worden, omdat de bijdrage van de private partijen binnen een TKI naar verwachting relatief groter worden (40% vanaf 2015).

Optie 2 Vergelijking TKI-bedrijven met en zonder projectdeelname

Experimentele en controlegroep

Vervolgens kan worden ingezoomd op informatie van bedrijven die deel uitmaken van een TKI. Binnen een TKI kunnen verschillende projecten worden uitgevoerd, waarbij niet noodzakelijkerwijs alle bedrijven en kennisinstellingen meedoen. Een optie is om de bedrijven binnen een TKI die deelnemen aan een specifiek project te vergelijken met bedrijven uit het TKI die niet aan dat specifieke deelproject meedoen. Deze optie is van toepassing op hypothesen 1, 3 en 5.

Identificatiestrategie

Bedrijven die deelnemen aan een specifiek project worden vergeleken met bedrijven die niet in een specifiek project participeren. De vergelijking is op basis van een verschil in trendmatige ontwikkeling in de periode voorafgaande aan een specifiek project en de periode nadat het project is ingevoerd.

Econometrische specificatie en methoden

Een difference-in-difference specificatie, waarbij de afhankelijke variabele wordt geregresseerd op indicatorvariabelen die 1 zijn vanaf het moment dat een bedrijf meedoet aan een specifiek project. Deze variabele is 1 voor alle bedrijven die meedoen aan dit specifieke project. Hierbij krijgt elk project een afzonderlijke indicatorvariabele. Verder dient er te worden gecorrigeerd voor tijd, zodat voor elk jaar een afzonderlijke indicatorvariabele θ wordt opgenomen. Voor verschillen tussen bedrijven wordt rekening gehouden door voor elk bedrijf een indicatorvariabele op te nemen. Dit wordt ook wel aangeduid als de "fixed effects" η . De regressievergelijking is dan:

$$Y_{it} = \beta_0 + \sum_j \beta_j \text{PROJECT}_{ijt} + \eta_i + \theta_t + u_{it}$$

In deze vergelijking is Y de prestatie van bedrijf i op tijdstip t , de β 's zijn de te schatten coëfficiënten en u is de storingsterm.

Verder kan worden onderzocht of de gevonden effecten over de loop van de tijd sterker worden, omdat de bijdrage van de private partijen binnen een TKI naar verwachting relatief groter worden (40% vanaf 2015). In dit geval kunnen de β 's dus variëren over de tijd. Voor het toetsen van hypothese 5 kan een indicator worden toegevoegd die aangeeft of het project gebruik maakt van de TKI-toeslag.

Gevoeligheden

Er is zelfselectie bij bedrijven en kennisinstellingen die aan een TKI-project deelnemen. Daardoor zou het wel eens kunnen zijn dat de veronderstelling van een gemeenschappelijke trend onjuist is. Het is toegestaan dat bedrijven aan een of meer projecten meedoet.

Box 4.7.1 Variatie in toeslagpercentages

Indien in de toekomst de toeslagpercentages wijzigen kan bij een evaluatie mogelijk gebruik van worden gemaakt van variatie die hierdoor over de jaren ontstaat. Bij een verschuiving van de grenzen van de toeslagpercentage kunnen bedrijven worden vergeleken die juist onder de nieuwe grens uitkomen met bedrijven die er juist boven vallen. Probleem van deze empirische strategie is overigens wel dat bedrijven invloed kunnen uitoefenen of ze boven of onder de grens uitkomen. Deze variatie is dus niet geheel exogeen.

Optie 3 Vergelijking van kennisinstellingen

Er kan ook worden ingezoomd op de kennisinstellingen. In principe kunnen hiervoor dezelfde specificaties worden gebruikt als hierboven beschreven bij opties 1 en 2. Hierbij wordt dan als afhankelijke variabele de publicatierecord of het aantal octrooien van een kennisinstelling gebruikt. Deze optie is van toepassing op hypothese 4.

Experimentele en controlegroep

Er worden deelnemende en niet-deelnemende wetenschappelijke instellingen vergeleken met betrekking tot de publicaties en het verwerven van 2^e en 3^e geldstroommiddelen. De vergelijking is op basis van de wetenschappelijke productie in de periode voorafgaande aan de TKI ten opzichte van die in de TKI.

Identificatiestrategie

Wetenschappers die deel uitmaken van een specifiek project worden vergeleken met bedrijven die niet in een specifiek project participeren. De vergelijking is op basis van een verschil in trendmatige ontwikkeling in de periode voorafgaande aan een specifiek project en de periode nadat het project is ingevoerd.

Econometrische specificatie en methoden

In principe wordt dezelfde identificatie strategie gehanteerd als bij optie 2. Een difference-in-difference specificatie, waarbij de afhankelijke variabele wordt geregresseerd op indicatorvariabelen die 1 zijn vanaf het moment dat een onderzoeker meedoet aan een specifiek project. Deze variabele is 1 voor alle onderzoekers die meedoen aan dit specifieke project. Hierbij krijgt elk project een afzonderlijke indicatorvariabele. Verder dient er te worden gecorrigeerd voor tijd, zodat voor elk jaar een afzonderlijke indicatorvariabele θ wordt opgenomen. Voor verschillen tussen bedrijven wordt rekening gehouden door voor elk bedrijf een indicatorvariabele op te nemen. Dit wordt ook wel aangeduid als de “fixed effects” η . De regressievergelijking is dan:

$$Y_{it} = \beta_0 + \sum_j \beta_j \text{PROJECT}_{ijt} + \eta_i + \theta_t + u_{it}$$

Gevoeligheden

Complicatie hierbij is dat publicaties een substantiële vertraging van meerdere jaren hebben, zodat deze optie weinig kansrijk lijkt te zijn.

Optie 4 Enquête onder bedrijven en kennisinstellingen

Hierin worden willekeurige ondernemingen binnen een TKI vergeleken met ondernemingen die geen deel uitmaken van TKI. Ondervraag de ondernemingen in hoeverre opbrengsten van innovatie en het gebruik van publieke onderzoeksresultaten zijn toegenomen in vergelijking met de periode voordat TKI werd aangevraagd en in hoeverre deze verandering door de TKI komt. Deze optie is van toepassing op alle hypothesen. Een dergelijke enquête kan ook worden uitgevoerd onder kennisinstellingen, waarbij dan zou moeten worden ingezoomd op publicaties en het verwerven van 2^e en 3^e geldstroommiddelen.

Conclusies

Per topsector zijn er één of meerdere TKI's. De opbrengsten van een TKI kunnen terecht komen bij de participerende bedrijven en bij de kennisinstellingen. Om de kernhypothese te toetsen zouden de bedrijven moeten worden vergeleken met een groep van bedrijven die niet deelneemt in een TKI, waarbij toewijzing van het TKI gerandomiseerd zou zijn. Hetzelfde design zou ook moeten gelden voor de kennisinstellingen. Willekeurige toewijzing is echter niet mogelijk, zodat er alternatieve schattingsmethoden worden voorgesteld. De plannen van de opzet van een TKI lijken het niet mogelijk te maken dat binnen een topsector meerdere clusters van bedrijven en kennisinstellingen dingen naar een deelname aan een TKI. Het is daardoor niet mogelijk om een TKI te vergelijken met een andere TKI binnen de topsector.

Van alle gepresenteerde opties lijken optie 1 en 2 het beste perspectief te bieden op inzicht in het effect van TKI's op innovatieresultaat. Beide opties zijn gebaseerd op een difference-in-difference specificatie. Optie 3 betrekking hebbend op kennisinstellingen, lijkt minder kansrijk te zijn vanwege substantiële vertragingen van publicaties. Deze opties geven echter geen antwoord op hypothese 2 over de aansluiting van publieke R&D-investeringen op de innovatie-inspanningen in de topsectoren. Een enquête kan als terugvaloptie worden beschouwd, waarmee inzicht kan worden gekregen in alle hypothesen, zonder dat hiermee causale effecten kunnen worden geschat.

Hoofdstuk 5 Conclusie en aanbevelingen

5.1 Conclusie

De roep om meer zicht op de effecten van beleid wordt steeds luider. In reactie hierop zijn steeds meer ministeries actief bezig met evidence-based beleid. De inspanningen van het ministerie van EZ staan niet op zich. Zo is bij het ministerie van Onderwijs, Cultuur en Wetenschap (OCW) in 2010 een traject in gang gezet dat vergelijkbaar is met dat in dit rapport. Voor ongeveer tweederde van de 50 in het programma betrokken beleidsinterventies bleek een natuurlijk experimenteel design mogelijk. Van de 50 projecten zijn er inmiddels ruim 10 ook echt aanbesteed. Van een aantal interventies is de effectiviteit aangetoond. Een bemoedigend resultaat.

Het DG B&I heeft ruime ervaring met beleidsevaluaties. De onderzoeksmethodes die in het verleden zijn toegepast variëren van enquêtes tot experimenten met loting. Voorbeelden van evaluaties waarin de directe effecten van beleid kwantitatief in beeld worden gebracht zijn de evaluaties van de Innovatievouchers en de WBSO. Het DG B&I wil zijn bestaande evaluatiepraktijk verder verbeteren, om zo te komen tot een heldere verantwoording en een zo effectief mogelijke besteding van publieke middelen. Dit rapport geeft hiervoor een aantal suggesties.

De hoofdpodracht van de Expertwerkgroep Effectmeting is het gestructureerd in beeld brengen van de verschillende evaluatiemogelijkheden voor de zes beleidsinstrumenten die in dit rapport worden behandeld. Tijdens de sessies waarin de verschillende instrumenten zijn besproken, heeft de werkgroep getracht hierin zo innovatief mogelijk te zijn. Naar het oordeel van de werkgroep bevat dit rapport een serie evaluatieontwerpen die in potentie de mogelijkheid biedt de directe effecten van de zes beleidsinstrumenten te kwantificeren. Hierbij moet wel worden opgemerkt dat, aangezien de vormgeving en de werking van de instrumenten voor de werkgroep een gegeven zijn, de sterkte van de aannames die aan de verschillende evaluatieontwerpen ten grondslag liggen uiteenlopen. Daarom geeft de werkgroep per instrument een advies (zie hoofdstuk 4) welk(e) ontwerp(en) zij het meest kansrijk acht voor evaluatie.

De Expertwerkgroep is voorts gevraagd om een oordeel over de meeste recente WBSO-evaluatie. De WBSO is een dusdanig grootschalig en generiek instrument dat het lastig is om een vergelijkbare controlegroep te vinden. Bij de WBSO-evaluatie is gebruik gemaakt van een dynamisch paneldatamodel dat zoveel mogelijk probeert te corrigeren voor geobserveerde en ook niet-geobserveerde bedrijfskenmerken. Omdat het nooit zeker is of dit volledig lukt, blijft er altijd enige onzekerheid of het geschatte effect ook echt het causale effect is van de WBSO. In het algemeen geldt dat een betere onderzoeksstrategie is om beleidsvariatie te zoeken die ervoor zorgt dat er vergelijkbare groepen bedrijven zijn die wel en niet meegedaan hebben aan een instrument. Bij de WBSO is zo iets lastig vanwege haar generieke karakter. Dit beklemtoont het belang om reeds in de ontwikkelingsfase van instrumenten veel meer aandacht te besteden aan toekomstige kwantitatieve evaluatiemogelijkheden.

Meting van de indirecte effecten van beleid, de zogenaamde tweede- en derde-orde-effecten, blijft een lastige zaak. Dit komt in de eerste plaats doordat hogere beleidsdoelen, zoals economische groei of werkgelegenheid, worden beïnvloed door veel factoren. Deze factoren maken het lastig de bijdrage van een instrument te isoleren van alle andere effecten. Hiernaast speelt tijd een belangrijke rol. Om tweede- en derde-orde-effecten te kunnen waarnemen moet meer tijd zijn verstreken dan bij het meten van directe effecten. Dit is niet altijd het geval bij het moment van evaluatie. Tot slot is het belangrijk dat voldoende data voorhanden is en databestanden kunnen worden gekoppeld.

5.2 Aanbevelingen

De werkgroep heeft voor de instrumenten Innovatiekrediet en Bewijs van Goede Dienst een tussentijds advies uitgebracht, aangezien voor deze instrumenten al op korte termijn een evaluatie was voorzien. Het feit dat DG B&I op voortvarende wijze aan de slag is gegaan met de adviezen, is voor de werkgroep een indicatie dat het ministerie serieus werk maakt van effectmeting. De werkgroep komt graag met nadere aanbevelingen, waardoor de effecten van beleid in de toekomst nog beter zichtbaar kunnen worden gemaakt.

1. Ex post evalueren begint ex ante met een heldere analyse. Wat is het maatschappelijk probleem? Moet de overheid dit probleem oplossen? Zo ja, welk beleidsinstrument is daarvoor het meest geschikt? Wanneer niet helder is wat het probleem is of hoe een beleidsinstrument bij moet dragen aan de oplossing van een probleem, wordt het lastig evalueren. In hoofdstuk 1 van dit rapport is aangegeven waarom bij innovatie maatschappelijke problemen kunnen ontstaan en waarom voor de overheid daarom eventueel een rol is weggelegd. *Werk voor elk beleidsinstrument ex ante de beleidstheorie uit, waarin per instrument één meetbaar doel wordt geformuleerd en wordt ingegaan op de economische legitimatie van de overheidsrol.*
2. Bij de opzet van beleidsinstrumenten is vaak weinig aandacht voor de toekomstige evaluatie van het instrument. Dit heeft als consequentie dat een goede empirische beleidsevaluatie vaak lastig is. Goed vergelijkbare groepen zijn dan bijvoorbeeld niet meer te vinden of slechts op lokale schaal rondom de beoordelingsgrens die voor een instrument bepaalt in welke mate bedrijven aanspraak kunnen maken op een interventie. In dit laatste geval kunnen slechts voor een deel van de gebruikers de effecten robuust in kaart worden gebracht. Extrapolatie van lokale effecten naar de gehele gebruikersgroep is niet zonder risico en kan tot foute conclusies leiden. Vaak moet daarom naar evaluatiemethoden worden teruggegrepen waaraan sterke veronderstellingen ten grondslag liggen. Hierbij is niet altijd helder vast te stellen of aan alle veronderstellingen is voldaan, zodat geen harde uitspraken over de effecten van beleid kunnen worden gedaan. *De werkgroep beveelt aan om reeds in de ontwikkelingsfase van nieuwe beleidsinstrumenten een evaluatieontwerp op te stellen waarin wordt aangegeven hoe de effecten van het instrument in de toekomst kunnen worden geëvalueerd.*
3. Zoals aangegeven in hoofdstuk 2 is het causale effect van een interventie het meest zuiver te schatten via een sociaal experiment op basis van willekeurige toewijzing. Zo'n experiment heeft als bijkomend voordeel dat het geschatte effect betrekking heeft op de gehele gebruikersgroep. Door voor een sociaal experiment te kiezen hoeven minder aannames te worden gemaakt, zodat ook tweede- en derde-orde-effecten beter in kaart zijn te brengen. De werkgroep beseft dat willekeurige toewijzing niet altijd mogelijk of wenselijk is. Bij gewogen loting worden subsidie- of kredietaanvragen eerst beoordeeld op kwaliteit, waarbij binnen de groep die boven een te bepalen grens (de streep) komt een loting plaatsvindt. Dit verhoogt de eerlijkheid van een loting en kan zo het draagvlak van een dergelijke maatregel vergroten. *Om het inzicht in de effectiviteit van instrumenten verder te vergroten beveelt de werkgroep aan om bij de introductie van nieuwe beleidsinstrumenten waar mogelijk te werken met sociale experimenten op basis van gewogen loting.*
4. Er is tijd nodig om de effecten van beleid te kunnen meten. Zelfs wanneer een goed evaluatieontwerp voorhanden is, moet er voldoende tijd verstrijken om observaties te kunnen registreren. Subsidies moeten volgens de Algemene Wet Bestuursrecht elke vijf jaar worden geëvalueerd. Het aantal jaren waarover observaties kunnen worden waargenomen is noodzakelijkerwijs dikwijls beperkt, omdat rekening moet worden gehouden met een aanlooperperiode voor inwerkingtreding van een instrument en met het feit dat een evaluatietraject vaak eerder wordt ingezet om vijf jaar na invoering over een evaluatie te kunnen beschikken. Bovendien kunnen, gegeven het doel van het instrument, niet altijd binnen de periode van vijf jaar alle effecten in kaart worden gebracht. Terwijl bij de meeste innovatie-instrumenten binnen de periode van vijf jaar een eerste-orde-effect mag worden verwacht, zoals een toename van de private R&D-investeringen, is deze periode voor tweede- en derde-orde-effecten (nieuwe producten, economische groei), gewoonweg te kort. *Voordat instrumenten grootschalig worden uitgerold is het aan te bevelen om een kleinschalige pilot met experimentele opzet te overwegen. Daarbij moet een realistisch evaluatiemoment worden gekozen, zodat in ieder geval de directe effecten in kaart kunnen worden gebracht.*
5. Empirische beleidsevaluaties leveren meer inzichten op wanneer de beschikbare data van hogere kwaliteit is. Ten aanzien van dataverzameling beveelt de werkgroep daarom concreet aan:
 - *Databestanden van AgNL (zoals Innovatiekrediet en WBSO), CBS en Belastingdienst zoveel mogelijk te koppelen en de resultaten daarvan zoveel mogelijk te ontsluiten via de website www.volginnovatie.nl. Dit is in lijn met het Open Data-beleid en draagt bij aan een betere verantwoording en transparantie van het beleid.*
 - *Dat AgNL niet alleen toegewezen maar ook afgewezen aanvragen actief gaat volgen met betrekking tot het verdere verloop van ingediende voorstellen/ projecten.*
 - *Tijdens evaluaties zowel onder toegewezen als afgewezen aanvragers enquêtes uit te zetten. Zonder gegevens over de controlegroep van niet-gebruikers is het lastig overtuigend bewijs voor de effectiviteit te vinden.*
 - *Met een positieve financiële prikkel te werken om de respons onder (afgewezen) aanvragers te vergroten (bijvoorbeeld een kleine geldelijke beloning voor het aanleveren van een ingevulde enquête).*

6. De Expertwerkgroep Effectmeting geeft in dit rapport slechts advies over zes beleidsinstrumenten. Het bedrijvenbeleid en de overige beleidsterreinen van het ministerie omvatten veel meer instrumenten waarvan de effecten nog onduidelijk zijn. *De werkgroep beveelt aan om op basis van de aanpak gehanteerd in dit rapport de effecten in kaart te brengen van de overige (grote) beleidsinstrumenten van zowel het bedrijvenbeleid (zoals de Innovatiebox) als van de andere beleidsterreinen van het ministerie.*

7. Het komt voor dat beleidsinstrumenten een vergelijkbare doelstelling hebben. Denk bijvoorbeeld aan de WBSO en RDA. *De werkgroep beveelt aan om instrumenten met een vergelijkbaar beleidsdoel zoveel mogelijk gezamenlijk te evalueren.*

De Expertwerkgroep Effectmeting beseft dat de implementatie van bovenstaande aanbevelingen een serieuze inspanning vergt van het ministerie van EZ. Deze inspanning moet worden afgezet tegen de omvang van beleidinstrumenten. Beleidsinstrumenten met een grotere financiële omvang⁶⁴ verdienen echter een deugdelijke evaluatie. Een ministerie moet kunnen rechtvaardigen dat belastingmiddelen effectief worden besteed.

⁶⁴€5 miljoen of meer.

