



Kennisspillovers van R&D in Nederland:

Een eerste stap in de micro-analyse van het publieke rendement van R&D

Brenda Bos (CBS), Gerben de Jong (SEO), Erik Brouwer (SEO) & Sander Sleijpen-Snoek (Min. EZ)
Maart 2026

Dit rapport is tot stand gekomen in het kader van het BAT-lab, het beleidsanalyse-laboratorium van het Directoraat-Generaal voor Bedrijfsleven en Innovatie van het ministerie van Economische Zaken. Daar wordt nauw samengewerkt door het Centraal Bureau voor de Statistiek (CBS), de Rijksdienst voor Ondernemend Nederland (RVO) en het Beleidsanalyseteam (BAT) van DG B&I. Centraal in deze samenwerking staan beleidsgedreven analyses op basis van microdata. Het CBS draagt geen medeverantwoordelijkheid voor beleidsmatige conclusies die worden getrokken op basis van deze analyses.

Managementsamenvatting

In hoeverre leveren investeringen in onderzoek en ontwikkeling (R&D) opbrengsten op voor het investerende bedrijf zelf? En in hoeverre vloeien deze baten door naar de economie als geheel doordat nieuwe kennis zich verspreidt naar andere bedrijven? Via welke kanalen vinden deze kennisspillovers dan plaats? Dit onderzoek brengt het private en publieke rendement van R&D-investeringen in Nederland in beeld, met specifieke aandacht voor kennisspillovers. Het private rendement geeft weer wat R&D oplevert voor het bedrijf dat investeert, terwijl het publieke rendement daarnaast ook de bredere baten voor de economie omvat.

Uit internationale studies blijkt veelvuldig dat het publieke rendement van R&D – dankzij het bestaan van spillovers – meestal aanzienlijk hoger is dan het private rendement. Voor Nederland ontbreken echter recente econometrisch inzichten over de omvang van spillovereffecten en de manier waarop deze zich door de economie verspreiden. Dat is een belangrijke lacune, want kennisspillovers vormen één van de voornaamste economische redenen voor overheidsbeleid dat R&D stimuleert, zoals de WBSO en de Innovatiebox (samen goed voor een budget van circa 4 miljard euro, bijna 0,4 procent van het bbp). Beter inzicht in deze spillovers helpt beleidsmakers om R&D-stimulering effectiever in te richten, bijvoorbeeld door ondersteuning te richten op bedrijven of sectoren waar het publiek rendement het grootst is.

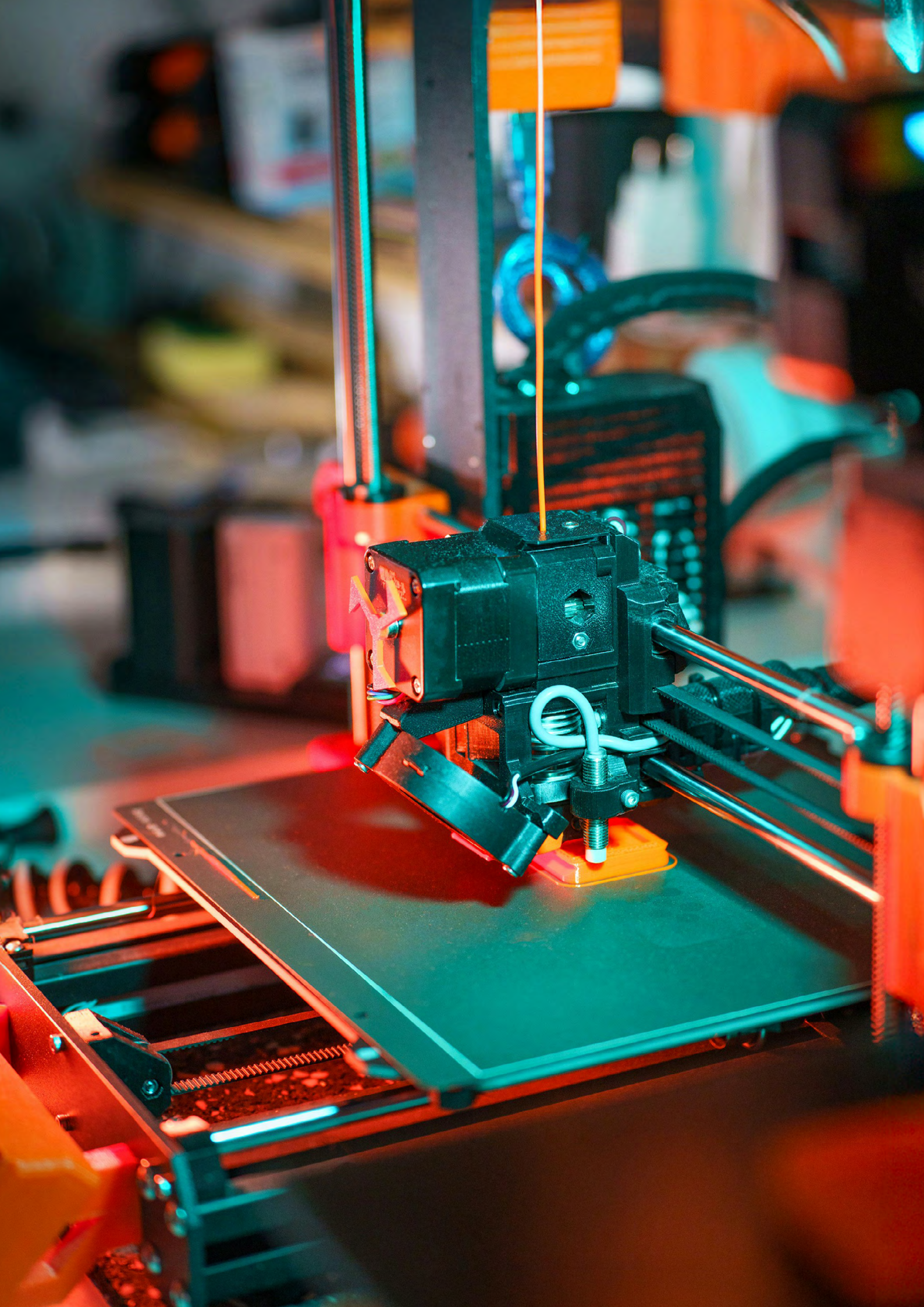
Kennis verspreidt zich vooral tussen bedrijven die aan elkaar gerelateerd zijn. In dit onderzoek brengen we gerelateerdheid tussen bedrijven – en daarmee de kans op kennisspillovers – op twee manieren in kaart. De eerste methode is generiek en gaat ervan uit dat kennis zich binnen sectoren en/of regio's verspreidt. De tweede methode is specifiek en bepaalt de mate van gerelateerdheid op basis van het gebruik van vergelijkbare technologieën, de toepassing van vergelijkbare vaardigheden en/of de geografische afstand tussen de vestigingen van bedrijven, waarbij sector en regiogrenzen niet langer leidend zijn. De omvang van de spillovers schatten we vervolgens met een productiefunctiemodel op microniveau, waarin zowel interne als externe kennisvoorraden een rol spelen (naast de gebruikelijke productiefactoren arbeid en kapitaal). We doen dit voor de gehele markteconomie in Nederland over de periode 2014–2022.

Op basis van onze modelschattingen berekenen we de zogenoemde 'R&D-multiplier': de waarde die op de lange termijn ontstaat door één extra euro aan R&D-investeringen, zowel voor het investerende bedrijf zelf (privaat) als voor de economie als geheel (publiek). De generieke modellen schatten het publieke rendement op circa 2,00 euro per geïnvesteerde euro. We beschouwen dit als een ondergrens, omdat kennisspillovers buiten de eigen sector en/of gemeente (en het buitenland) niet worden meegenomen. De specifieke modellen komen uit op een relatief hoog publiek rendement van ongeveer 6,50 euro per geïnvesteerde euro. Dit interpreteren we als een bovengrens, aangezien deze methode mogelijk ook andere regionale of economische effecten oppikt die de werkelijke spillovers overschatten (endogeniteit). Het private rendement ligt in beide benaderingen stabiel rond de 0,82 euro. Uit beide methoden blijkt daarmee dat er aanzienlijke kennisspillovers plaatsvinden, waarbij de baten uit spillovers groter zijn dan het private rendement. Dat het private rendement gemiddeld onder de één euro blijft, onderstreept bovendien dat bedrijven zonder overheidssteun onvoldoende prikkels hebben om extra R&D-investeringen te doen.

Ons onderzoek vormt een eerste stap richting een model waarmee de maatschappelijke baten van verschillende vormen van R&D-beleid kunnen worden gekwantificeerd en doorgerekend voor een groot deel van de Nederlandse economie. Vervolgonderzoek is nodig om beter zicht te krijgen op verschillen in spillovereffecten tussen bedrijven en sectoren, en om de schattingen van het publieke rendement verder aan te scherpen en de bandbreedte van de effecten te verkleinen.

Inhoud

Managementsamenvatting	3
1 Inleiding	6
1.1 Aanleiding	6
1.2 Onderzoeksopzet	6
1.3 Leeswijzer	8
2 Literatuur	9
2.1 Internationale empirische literatuur over de private en publieke baten van R&D	9
2.2 Onderzoek in de Nederlandse context	10
2.3 Andere onderwerpen in de literatuur	11
3 Data	12
3.1 Databronnen en afbakening populatie	12
3.2 Samenstellen interne en externe kennisvoorraden	13
4 Descriptieve statistieken	19
4.1 Relevante variabelen in de onderzoekspopulatie	19
5 Econometrische methodologie	24
6 Resultaten	27
6.1 Modellen met generieke externe kennisvoorraden	27
6.2 Modellen met specifieke externe kennisvoorraden	28
6.3 Gevoeligheidsanalyses	29
6.4 Langetermijneffecten en R&D-multipliers	32
6.5 Interpretatie en discussie	33
7 Conclusies	34
Literatuur	36
Bijlage A: Achtergrondinformatie samenstelling dataset	38
Bijlage B: Additionele figuren en tabellen	40



1 Inleiding

1.1 Aanleiding

Bedrijven investeren in onderzoek en ontwikkeling (R&D) met de verwachting dat de kennis die ze hiermee opdoen leidt tot betere producten en/of lagere productiekosten en daarmee tot verbeterde bedrijfsprestaties (bijvoorbeeld een hogere productiviteit, omzet of winst). Kennis wordt beschouwd als een van de belangrijkste bronnen voor bedrijven om een duurzaam concurrentievoordeel te behalen (Grant, 1996). Deze zoektocht naar nieuwe ideeën door winstmaximaliserende ondernemers vormt de kern van de endogene groeitheorie (Romer, 1990), waarin economische groei niet vanzelf ontstaat, maar voortkomt uit investeringen van bedrijven in kennis en innovatie.

De opbrengsten van R&D reiken verder dan het bedrijf dat de investering doet, omdat andere bedrijven via *spillovereffecten* ook profiteren van de ontwikkelde kennis en innovaties (Griliches, 1979; Chesbrough, 2003).¹ In dit licht wordt onderscheid gemaakt tussen het *private rendement* van R&D voor de investerende partij en de bredere economische baten die via kennisspillovers bij andere bedrijven terechtkomen. Het *private rendement* en de spillovers vormen samen het *publiek rendement*. Omdat bedrijven de baten voor andere bedrijven niet meenemen in hun investeringsbeslissingen, wordt er mogelijk minder in R&D geïnvesteerd dan vanuit maatschappelijk oogpunt wenselijk is. Het bestaan van kennisspillovers is daarmee een belangrijke economische rechtvaardiging van overheidsbeleid gericht op het stimuleren van R&D-investeringen (Becker, 2015; Bloom et al., 2019).^{2, 3}

In dit onderzoek analyseren we het *private* en *publieke rendement* van R&D-investeringen voor de gehele markteconomie in Nederland over de periode 2014–2022, met specifieke aandacht voor de rol van kennisspillovers. Hoewel het belang van spillovereffecten breed wordt erkend, ontbreken voor Nederland recente (econometrische) inzichten in de daadwerkelijke omvang en verdeling hiervan. Daarbij focust veel literatuur – zowel in de internationale als de Nederlandse in context – zich vaak op een subset van bedrijven en/of sectoren (bijvoorbeeld bedrijven die patenteren). In deze studie maken we gebruik van integrale registerdata van het CBS om een zo volledig mogelijk en actueel beeld te schetsen van de kennisspillovers in Nederland. Vragen die centraal staan zijn: in hoeverre is de kennis die voortkomt uit R&D *privaat* of *publiek*? Wat is de omvang van kennisspillovers en via welke kanalen verspreidt kennis zich? Hoe verhouden de prikkels voor bedrijven om in R&D te investeren zich tot de bredere economische opbrengsten?

1.2 Onderzoeksopzet

Bestaande literatuur laat zien dat kennisspillovers via uiteenlopende transmissiekanalen tot stand kunnen komen, waarbij de kans op kennisverspreiding groter is naarmate bedrijven sterker aan elkaar gerelateerd zijn (zie, o.a., Griliches, 1992; Hall et al., 2010; Bloom et al., 2013; Lucking et al., 2019). De gerelateerdheid tussen bedrijven kan op verschillende manieren worden vastgesteld. In dit onderzoek gebruiken we twee methodes. We beginnen met een *generieke* methode waarbij wordt aangenomen dat alle bedrijven binnen dezelfde sector of regio aan elkaar gerelateerd zijn en dus in theorie van elkaars kennis kunnen profiteren (zie o.a. Griliches, 1992; Soete & Ter Weel, 1999). Deze methode veronderstelt dat bedrijven ofwel volledig aan elkaar gerelateerd zijn of helemaal niet, waardoor bedrijven buiten de eigen sector of regio per definitie niet van elkaars kennis profiteren.

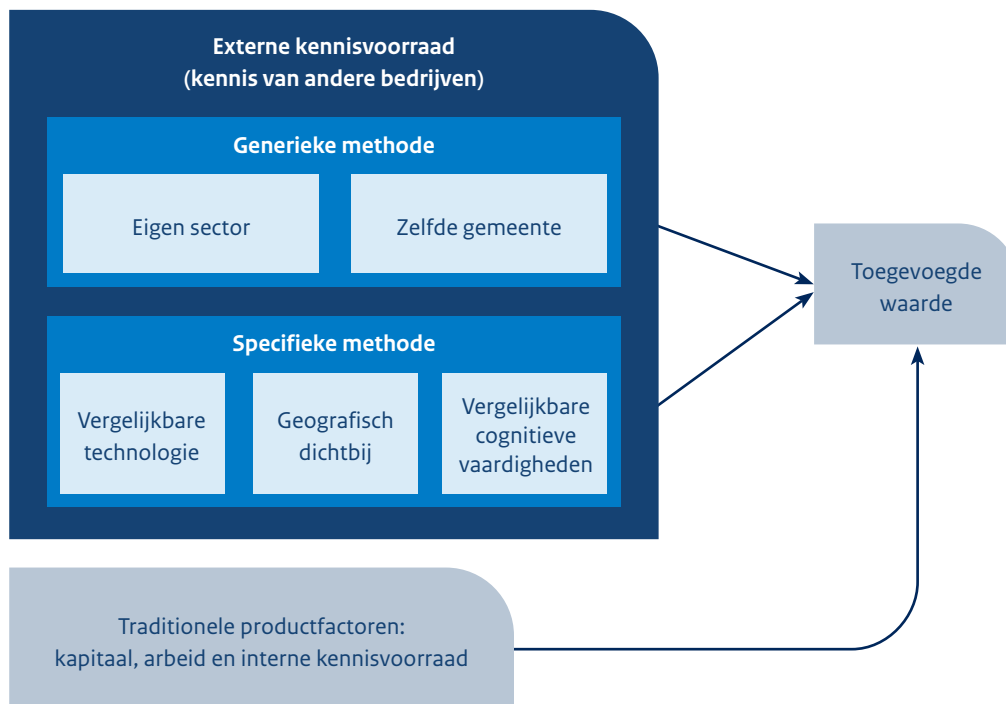
1 Kennis wordt hierdoor vaak beschouwd als een gedeeltelijk publiek goed (zie Nelson, 1959; Arrow, 1962, die hier als eerste op wezen), omdat het gebruik ervan door anderen de waarde niet noodzakelijk vermindert en het bovendien moeilijk is om anderen volledig van gebruik uit te sluiten.

2 Nederlandse voorbeelden van publieke regelingen gericht op het stimuleren van R&D-investeringen zijn de Wet Bevordering Speur- en Ontwikkelingswerk en de Innovatiebox. In 2022 hadden deze twee regelingen samen een budget van ongeveer 4 miljard euro (bijna 0,4% van het bbp).

3 Andere onderbouwingen zijn dat de innovaties die uit R&D voortkomen, kunnen helpen bij het versnellen van maatschappelijk wenselijke transities, bijvoorbeeld op het gebied van verduurzaming, digitalisering en/of het betaalbaar houden van de zorg (Mazzucato, 2018).

Daarnaast hanteren we een *specifiekere* methode, waarbij de mate van gerelateerdheid tussen bedrijven op een nauwkeurigere manier wordt vastgesteld. Hierbij onderscheiden we drie verschillende kanalen voor kennispillovers: bedrijven kunnen van elkaars kennis profiteren omdat ze (i) gebruikmaken van vergelijkbare technologieën (Jaffe, 1986; Bloom et al., 2013), (ii) geografisch gezien dichtbij elkaar liggen (Audretsch & Feldman, 1996; Lychagin et al., 2016) en/of (iii) overeenkomen in het type cognitieve vaardigheden dat in de bedrijven wordt toegepast (Neffke et al., 2017).⁴ Deze methode staat toe dat bedrijven gedeeltelijk aan elkaar gerelateerd kunnen zijn, ook wanneer zij in verschillende sector of gemeente actief zijn. Zie Figuur 1 voor een schematisch overzicht van de twee methoden.

Figuur 1: Schematisch overzicht van de twee verschillende methoden (generiek versus specifiek)



Voor elke methode en elk transmissiekanaal brengen we in kaart in welke mate bedrijven aan elkaar gerelateerd zijn. Vervolgens wordt voor elk kanaal een aparte *externe kennisvoorraad* berekend door de mate van gerelateerdheid te combineren met de R&D-investeringen, zodat per bedrijf duidelijk wordt hoeveel externe kennis via elk kanaal beschikbaar is. Daarna schatten we op basis van economische modellen wat het effect van deze externe kennisvoorraden is op de toegevoegde waarde van bedrijven. Concreet betekent dit dat we bij de generieke methode de toegevoegde waarde van bedrijven relateren aan de externe kennisvoorraden van (i) andere bedrijven in de eigen sector en (ii) van bedrijven in dezelfde gemeente. Bij de specifieke methode relateren we de toegevoegde waarde aan de externe kennisvoorraden van (i) technologisch gerelateerde bedrijven, (ii) geografisch gerelateerde bedrijven en (iii) cognitief gerelateerde bedrijven.

Voor het schatten van het effect van externe kennis op toegevoegde waarde gebruiken we – in lijn met een lange lijst aan internationale literatuur over de impact van R&D op economische groei (zie Hall et al., 2010 voor een overzicht) – een productiefunctiekader waarin, naast arbeid en kapitaal, ook interne en externe kennis als productiefactoren worden meegenomen (zie onderste blok in Figuur 1). De modellen worden geschat op microdata van het Centraal Bureau voor de Statistiek (CBS) voor de periode 2014–2022. Als benadering voor R&D-investeringen gebruiken we tijdreeksen van S&O-uitgaven onder de WBSO-regeling.⁵ Deze zijn gekoppeld aan bedrijfsdemografische en financiële gegevens van vrijwel alle economisch actieve bedrijven met een commerciële rechtspersoon in het niet-financiële bedrijfsleven van Nederland.

4 Recentere literatuur benadert de verbondenheid tussen bedrijven vanuit hun positie in bedrijfsnetwerken (zie König et al., 2019; Hernandez et al., 2025). Deze benadering van gerelateerdheid stelt hoge eisen aan de data en schattingsmethoden en blijft buiten beschouwing in dit onderzoek.

5 De WBSO (Wet Bevordering Speur- en Ontwikkelingswerk) is een fiscale stimuleringsregeling voor R&D die sinds 1994 in Nederland bestaat (Brouwer et al., 2025). De regeling richt zich op Speur- en Ontwikkelingswerk (S&O), een subset van R&D die bestaat uit technisch-wetenschappelijk onderzoek en de ontwikkeling van technisch nieuwe producten, processen en programmatuur. In Bijlage A reflecteren we kort op het gebruik van S&O als proxy voor R&D.

De geschatte effecten voor interne en externe kennis vormen de basis voor het afleiden van het private en publieke rendement van R&D-investeringen. Hiermee berekenen we vervolgens de ‘R&D multiplier’: hoeveel economische waarde op de langere termijn wordt gegenereerd door één extra euro aan R&D-uitgaven, zowel voor het bedrijf zelf als voor de economie als geheel.

Dit onderzoek vormt een eerste stap richting een model waarmee de maatschappelijke baten van verschillende vormen van R&D-beleid kunnen worden gekwantificeerd en doorgerekend voor een groot deel van de Nederlandse economie. Deze inzichten kunnen beleidsmakers helpen R&D-stimuleringsmaatregelen effectiever vorm te geven, bijvoorbeeld door deze te richten op bedrijven en sectoren waarin het grootste publieke rendement behaald kan worden.

1.3 Leeswijzer

Hoofdstuk 2 bespreekt de literatuur over de publieke en private baten van R&D-investeringen en hoe deze empirisch geschat kunnen worden. Hoofdstuk 3 presenteert de dataset, de afbakening van de onderzoekspopulatie en onze methode om de verschillende transmissiekanalen en de bijbehorende externe kennisvoorraden te berekenen. Hoofdstuk 4 bevat relevante beschrijvende statistieken. Hoofdstuk 5 gaat in op de econometrische details van het productiefunctiemodel. Hoofdstuk 6 presenteert en bediscussieert de resultaten en hoofdstuk 7 concludeert.

2 Literatuur

In dit hoofdstuk geven we een kort overzicht van de belangrijkste academische en beleidsmatige literatuur. Eerst beschrijven we een aantal studies over de bijdrage van R&D en kennis aan bedrijfsprestaties en economische groei, en manieren om de private en publieke opbrengsten van investeringen hierin te meten. Vervolgens focussen we op enkele studies die specifiek betrekking hebben op Nederland. De meeste van deze studies zijn inmiddels relatief oud. Als laatste benoemen we nog enkele andere onderwerpen uit de literatuur die in dit onderzoek buiten beschouwing blijven maar mogelijk wel relevant zijn voor vervolgonderzoek op dit terrein.

2.1 Internationale empirische literatuur over de private en publieke baten van R&D

Zvi Griliches legde in de jaren zeventig de basis voor het schatten van de private en publieke baten van R&D, binnen een productiefunctieraamwerk (Griliches, 1973; 1979). Deze modellen beschouwen kennis als een derde productiefactor, naast arbeid en kapitaal, en veronderstellen dat R&D-investeringen de kennisvoorraad van een bedrijf, sector of economie kunnen vergroten. Daarbij wordt de productiefactor kennis opgedeeld in interne kennis afkomstig uit eigen R&D-investeringen, en externe kennis door investeringen van andere bedrijven in de sector of economie. Ook wordt rekening gehouden met het feit dat een deel van de investeringen bestaande, verouderde kennis vervangen, en een deel de kennisvoorraad daadwerkelijk vergroten. Hall et al. (2010) hebben een gezaghebbend overzichtsartikel geschreven over de geschiedenis van deze literatuur, de verschillende econometrische modellen die daar ingezet worden en een overzicht van de empirische resultaten die tot dat moment waren verkregen.⁶ Het artikel stelt dat deze literatuur laat zien dat het private rendement op R&D aanzienlijk is en dat het publieke rendement, door substantiële kennisspillovers, nog hoger ligt, al variëren de schattingen daarvan sterk en zijn ze vaak onnauwkeurig.

Vervolgonderzoek van Bloom et al. (2013) onderscheidt twee tegengestelde *mechanismen* waarlangs R&D andere bedrijven beïnvloedt: *technologische spillovers* en *productmarktconcurrentie*. Technologische spillovers zijn het resultaat van kennisverspreiding via technologische gerelateerdheid – bedrijven profiteren van elkaars innovaties doordat zij vergelijkbare technologieën gebruiken, patenten kunnen inzien of personeel overnemen. Dit kanaal heeft een positief effect op de productiviteit van andere bedrijven. Concurrentie op de productiemarkt daarentegen zorgt voor een negatief spillovereffect: succesvolle innovatie versterkt de concurrentiepositie van het innoverende bedrijf, wat leidt tot lagere prijzen, marktaandeel- of winstverliezen bij concurrenten. Bloom et al. ontwikkelen een methode om deze twee mechanismen afzonderlijk te identificeren. De technologische positie van een bedrijf wordt bepaald aan de hand van de verdeling van patentaanvragen over technologische domeinen (à la Jaffe, 1986), en in de positie in de productmarkt op basis van verkoopactiviteiten over viercijferige industrieën. Toegepast op een panel van Amerikaanse bedrijven (1981–2001) laten de resultaten zien dat beide effecten kwantitatief relevant zijn, maar dat de positieve technologische spillovers domineren. Het publieke rendement op R&D (circa 55 procent) is ongeveer tweeënhalf keer zo groot als het private rendement (circa 21 procent).⁷

Lucking et al. (2019) bouwen voort op deze analyse en passen de methode van Bloom et al. toe op een langere periode (1985–2015) voor Amerikaanse bedrijven. Zij vinden eveneens aanzienlijke positieve technologische spillovers, naast kleinere negatieve effecten via productmarktconcurrentie. Een globale welvaartsanalyse suggereert dat het marginale publieke rendement op R&D circa 58 procent bedraagt, terwijl het marginale private rendement rond de 14 procent ligt. Met andere woorden: het maatschappelijke rendement is ongeveer vier keer zo hoog als het private rendement. Bovendien laten zij zien dat zowel het private als het publieke rendement in 2015 vergelijkbaar zijn met hun niveaus dertig jaar eerder.

Lychagin et al. (2016) breiden de benadering van Bloom et al. (2013) verder uit door een tweede *kanaal* van kennisspillovers te analyseren: naast technologische gerelateerdheid kijken ze ook naar de

6 Ugur et al. (2016; 2020) en Moen & Thorsen (2017) zijn recentere meta-analyses, die vooral laten zien dat als gecorrigeerd wordt voor publicatie bias de baten van R&D, zowel privaat als publiek, waarschijnlijk lager uitvallen dan gevonden in de eerdere literatuur.

7 In ons onderzoek nemen we geen maatstaf mee voor *productmarktconcurrentie*. Hiervoor heb je gedetailleerde informatie nodig over verkoopactiviteiten van bedrijven. Deze informatie is niet beschikbaar bij het CBS.

geografische nabijheid van vestigingslocaties.⁸ Bedrijven die dichtbij elkaar gelegen zijn, hebben meer mogelijkheden om met elkaar te kunnen communiceren. Professionals ontmoeten elkaar bijvoorbeeld via lokale bedrijfsnetwerken of in de privésfeer. Geografische nabijheid kan daardoor ook voor kennisverspreiding zorgen. Technologische gerelateerdheid hoeft daarvoor geen voorwaarde te zijn, maar kan dit wel verder versterken. De uitbreiding met geografische nabijheid is belangrijk omdat het negeren hiervan kan leiden tot een overschatting van spillovers via technologische gerelateerdheid, bijvoorbeeld wanneer technologisch vergelijkbare bedrijven zich in dezelfde regio concentreren. Naast de beide transmissiekanalen voor spillovers nemen de onderzoekers in hun model ook nabijheid in de productmarkt mee om voor het productmarktconcurrentie-effect te controleren. De resultaten laten zien dat ook geografische nabijheid een significant kanaal vormt voor kennisverspreiding: bedrijven profiteren sterker van elkaars R&D wanneer zij zich dicht bij elkaar bevinden. Bovendien blijft dit effect in de tijd overeind, ondanks toenemende digitalisering. Deze bevinding sluit aan bij eerdere studies, zoals Aldieri en Cincera (2009), die eveneens positieve effecten vinden van zowel technologische als geografische gerelateerdheid op productiviteitsgroei. De algemene conclusie luidt dat geografische nabijheid een belangrijk kanaal voor kennisspillovers blijft, zelfs in een sterk gedigitaliseerde economie.

De bestaande literatuur meet kennisspillovers doorgaans met behulp van patenten. Niet alle sectoren doen patentaanvragen, voornamelijk omdat de aard van de activiteiten zich daar niet voor leent. Patentaanvragen in de dienstverlening zijn bijvoorbeeld erg zeldzaam. Dat wil niet zeggen dat er geen kennisspillovers plaats kunnen vinden in sectoren die niet of nauwelijks patenteren. Er is een separate stroom in de literatuur die arbeidsstromen tussen bedrijven gebruikt om (de kans op) kennisspillovers in kaart te brengen (Stoyanov & Zubanov, 2014). Bedrijven waartussen veel arbeidsstromen plaatsvinden zijn cognitief gerelateerd aan elkaar, wat wil zeggen dat hun cognitieve vaardigheden op elkaar lijken.⁹ Deze overeenkomsten kunnen kennisverspreiding bevorderen. In dit onderzoek gebruiken we *cognitieve gerelateerdheid* als een derde kanaal waarlangs spillovers kunnen optreden. Abolhassani (2024) vindt bijvoorbeeld voor de Nederlandse productiesector dat mobiliteit van werknemers binnen dezelfde sector sterker gerelateerd is aan de verspreiding van kennis en vaardigheden dan de mobiliteit van werknemers tussen sectoren. Neffke et al. (2017) vinden dat hoe groter de overlap in kennis tussen de oude en nieuwe werkgever, hoe meer deze nieuwe kennis bijdraagt aan de productiviteit van de nieuwe werkgever.

2.2 Onderzoek in de Nederlandse context

Er zijn enkele eerdere studies gedaan naar het private en publieke rendement op R&D in Nederland, al zijn veel van deze onderzoeken relatief lang geleden uitgevoerd. Bartelsman et al. (1996) schatten enkel het private rendement op R&D voor een steekproef van Nederlandse industriële bedrijven. De studie maakt gebruik van de gangbare productiefunctie, waarbij de eigen R&D-kennisvoorraad (of R&D-intensiteit) is toegevoegd als extra inputfactor, en waarbij de output zowel als de bruto productiewaarde en als toegevoegde waarde is gemeten. De gegevens zijn afkomstig van de vierjaarlijkse R&D-enquêtes en de jaarlijkse productiegegevens die het CBS heeft verzameld voor de jaren 1985, 1989 en 1993. Over verschillende modelspecificaties varieert het private rendement op R&D in de Nederlandse industrie van 12 procent (voor omzet) tot 30 procent (voor toegevoegde waarde). Dit ligt in lijn met internationale schattingen. Spillovereffecten zijn niet meegenomen in deze studie.

Soete en Ter Weel (1999) en Jacobs et al. (2002) schatten wél de omvang van kennisspillovers in Nederland. Beide studies modelleren spillovers op mesoniveau via generieke transmissiekanalen op het niveau van sectoren. De studies verschillen in de sectorafbakening – enkel industrie, of ook diensten – en of de schattingen zijn uitgevoerd op arbeidsproductiviteit of totale factorproductiviteit. In de twee studies zijn significante spilloverelasticiteiten geschat voor zowel de R&D uitgevoerd in de eigen sector, andere Nederlandse sectoren, als in het buitenland.

8 In Lychagin et al. (2016) blijken de geschatte geografische effecten sterker wanneer de locatie van uitvinders wordt gebruikt als maatstaf voor R&D-locatie, in plaats van de vestigingsplaats van het hoofdkantoor.

9 De redenering is dat wanneer werknemers van baan wisselen, de kennis en vaardigheden die zij nodig hadden bij hun vorige werkgever doorgaans lijken op wat hun nieuwe werkgever vraagt.

Koopmans en Donselaar (2015) hebben recenter een *metaregressie-analyse* uitgevoerd op basis van empirische studies naar het effect van R&D op productiviteit. Daarbij combineren zij resultaten uit meerdere bestaande onderzoeken — elk met eigen data en methoden — om een gewogen gemiddelde ('best guess') te schatten van het te verwachten effect voor Nederland. Alleen studies op macroniveau (landenvergelijkingen) zijn meegenomen. Op basis van deze analyse schatten zij dat een toename van 10 procent in privaat R&D-kapitaal leidt tot 0,58 procent hogere productiviteit. Witteman en Bijlsma (2024) komen op basis van een bredere set aan studies uit op een mediane outputelasticiteit van de R&D-kapitaalvoorraad van 0,50 procent hogere productie bij een toename van 10 procent in privaat R&D-kapitaal. Brouwer et al. (2025) rekenen op basis van deze metaschattingen uit dat één euro extra aan R&D-investeringen op termijn tussen de 2,50 en 3,00 euro aan bruto contant publiek rendement oplevert.

2.3 Andere onderwerpen in de literatuur

Naast de hierboven genoemde thema's bespreekt de literatuur ook andere onderwerpen die in dit onderzoek verder niet expliciet aan bod komen, maar wel relevant kunnen zijn voor eventueel vervolgonderzoek.

Ten eerste zijn er studies die de onderlinge gerelateerdheid tussen bedrijven analyseren op basis van R&D-samenwerkingsnetwerken (bijv. König et al., 2019; Hernandez et al., 2025). Deze benadering biedt een nog nauwkeurigere manier om te bepalen welke bedrijven met elkaar verbonden zijn, en waar kennisspillovers dus het meest waarschijnlijk zijn. Het toepassen van deze methode vereist echter gedetailleerde data over R&D-samenwerkingen, die voor Nederland momenteel niet beschikbaar zijn – zeker niet op het niveau van het volledige niet-financiële bedrijfsleven.

Daarnaast wordt in de literatuur veel aandacht besteed aan het feit dat R&D niet alleen nieuwe kennis genereert, maar bedrijven ook in staat stelt om externe kennis beter op te nemen ('absorptievermogen'). Dit wordt aangeduid als *the two faces of R&D* (Cohen & Levinthal, 1989; Griffith et al., 2003). Het impliceert dat bedrijven met eigen R&D-investeringen een grotere absorptiecapaciteit hebben en dus beter kunnen profiteren van spillovers. In ons onderzoek veronderstellen we echter generieke spillovers naar alle bedrijven, ongeacht of zij zelf R&D verrichten. Ook richten we ons op de afzonderlijke effecten van de verschillende transmissiekanalen om de reikwijdte van ons onderzoek te beperken, hoewel er aanwijzingen zijn dat, bijvoorbeeld, geografische en technologische gerelateerdheid elkaar kunnen versterken (Aldieri & Cincera, 2009).

In open economieën zoals de Nederlandse kunnen er ook aanzienlijke internationale kennisstromen zijn (Coe & Helpman, 1995).¹⁰ In ons onderzoek beschikken we niet over gegevens over buitenlandse R&D-investeringen om hier rekening mee te houden. Het publieke rendement dat we hier schatten moet daarom worden geïnterpreteerd als een *nationaal* publiek rendement. Onder de aanname van positieve internationale spillovers vormt dit een ondergrens van het *globale* publieke rendement van R&D.

Tot slot is het belangrijk om op te merken dat we ons in dit onderzoek richten op het effect van R&D op de toegevoegde waarde van bedrijven en, in het verlengde daarvan, op de bruto toegevoegde waarde. Voor de doorrekening van beleidsmaatregelen ter stimulering van R&D is echter óók het effect van dat beleid op de (private) R&D-investeringen van belang (zie Van Vuuren et al., te verschijnen). In de praktijk wordt dit vaak geschat via de zogenoemde *bang-for-the-buck*: hoeveel extra R&D-investeringen elke euro aan beleidsinzet oplevert.¹¹ Door deze schattingen te combineren met ons berekende rendement op R&D-investeringen kan vervolgens het totale rendement van R&D-beleid worden vastgesteld.

10 Soete & Ter Weel (1999) en Jacobs et al. (2002) nemen internationale spillovers wel mee in hun schattingen. Deze zijn doorgaans echter kleiner dan binnenlandse spillovers. Wij beschikken in dit onderzoek niet over microdatagegevens over buitenlandse R&D-investeringen.

11 Zie Rao (2016), Agrawal et al. (202) en Dechezleprêtre et al. (2023) voor enkele voorbeelden van dergelijke studies uit de internationale literatuur en Bijlsma et al. (2024), Brouwer et al. (2025) en De Jong et al. (2025) voor enkele voorbeelden in de Nederlandse context.

3 Data

Dit hoofdstuk beschrijft de opzet van de dataset en de afbakening van de onderzoekspopulatie. Vervolgens wordt ingegaan op het samenstellen van de interne kennisvoorraden en, tot slot, komen de verschillende externe kennisvoorraden aan bod. In bijlage A is op een aantal punten meer detail gegeven.

3.1 Databronnen en afbakening populatie

Gebruikte bronnen

Voor dit onderzoek is gebruikgemaakt van gegevens in de microdataomgeving van het CBS. De basis van het onderzoek is een databestand met een ongebalanceerde panelstructuur voor de periode 2014 – 2022, samengesteld op basis van het Algemeen Bedrijvenregister (ABR) en het Bedrijfsdemografisch Kader (BDK).¹² Deze twee microdatabestanden bevatten de bedrijfsdemografische kenmerken van alle economisch actieve bedrijven in Nederland.¹³ Aan deze basis zijn vervolgens financiële informatie uit de Statistiek financiën van niet-financiële ondernemingen (NFO) en uitgaven aan speur- en ontwikkelingswerk (S&O) onder de Wet Bevordering Speur- en Ontwikkelingswerk (WBSO) gekoppeld. We gebruiken S&O-uitgaven als benadering voor R&D-investeringen, waaruit we de interne en externe kennisvoorraden afleiden (zie paragraaf 3.2).¹⁴ Alle financiële gegevens zijn op basis van macro-indices gecorrigeerd voor inflatie en uitgedrukt in contante waarde tegen prijzen van 2022. Tabel 1 toont een overzicht van de gebruikte databronnen en variabelen.

Tabel 1: Overzicht gebruikte databronnen en variabelen

Bron	Variabelen	Bijzonderheden
Bedrijfsdemografisch Kader (BDK)	Bedrijfseenheid-identificatienummer, bedrijfstak (SBI-letter), sector (tweecijferige SBI-code), rechtsvorm, en aantal werkzame personen	- Aantal werkzame personen is gebruikt als proxy voor productiefactor arbeid; - Aantal werkzame personen is gecorrigeerd voor R&D-medewerkers, om dubbeltellingen met de productiefactor kennis te voorkomen¹⁵; - Data op bedrijfseenheidsniveau, geaggregeerd op ondernemingengroepsniveau.
Algemeen Bedrijvenregister (ABR)	Ondernemingengroep-identificatienummer, vestigingslocatie (gemeente)	Data op bedrijfseenheidsniveau, geaggregeerd op ondernemingengroepsniveau.
Financiën van niet-financiële ondernemingen (NFO)	Toegevoegde waarde, afschrijvingen	- Toegevoegde waarde is de uitkomstmaat in de productiefunctiemodellen; - Afschrijvingen is gebruikt als proxy voor de productiefactor kapitaal; - Data op ondernemingengroepsniveau.
Wet Bevordering Speur- en Ontwikkelingswerk (WBSO)	S&O-uitgaven (loon- en nietloonkosten) en S&O-uren	- S&O-uitgaven is gebruikt om interne en externe kennisvoorraden te bepalen; - Data op bedrijfseenheidsniveau, geaggregeerd op ondernemingengroepsniveau.

12 Over deze periode is een consistente tijdreeks beschikbaar voor alle belangrijke variabelen. In 2014 is het ABR grondig gereviseerd, door o.a. het gebruik van het Nieuw Handels Register als leidende bron en veranderingen in hoe statistische eenheden, zoals bedrijfseenheden en ondernemingengroepen, worden afgeleid, zie [cbs.nl](https://www.cbs.nl). Ten tijde van (de start van) het onderzoek was 2022 het laatste jaar waarover financiële gegevens uit de NFO beschikbaar waren.

13 Volgens de definitie van het CBS is een bedrijf economisch actief als het in een gegeven jaar omzet genereert of werknemers in dienst heeft.

14 Dit is in lijn met de gangbare praktijk in R&D-onderzoek in Nederland (zie, bijv., Sleijpen et al., 2024). Recent zijn er wel aanwijzingen dat de uitgaven aan S&O en R&D telkens verder uit elkaar zijn gaan lopen (Brouwer et al., 2025). Zie verdere toelichting over het gebruik van S&O als proxy voor R&D in Bijlage A.

15 Deze correctie komt erop neer dat we het aantal werkzame personen verlagen met het aantal S&O-arbeidsjaren (uitgaande van 1976 uur in een arbeidsjaar), waarbij nul als natuurlijke ondergrens geldt.

Analyseniveau

De analyses zijn uitgevoerd op het niveau van een ondernemingengroep, oftewel een onderneming/concern. Een onderneming kan bestaan uit één of meerdere statistische bedrijfseenheden, die op hun beurt kunnen bestaan uit één of meerdere juridische eenheden. Vooral bij grotere en complexere bedrijven is er sprake van meerdere bedrijfseenheden binnen dezelfde ondernemingengroep. We kiezen voor analyse op het niveau van de ondernemingengroep, omdat de financiële gegevens (waaronder de toegevoegde waarde, onze belangrijkste outputvariabele) van bedrijven op dit niveau zijn gedefinieerd. Bovendien worden beslissingen over R&D-investeringen doorgaans genomen op dit hogere concernniveau. Voor enkele variabelen (aantal werkzame personen, bedrijfstak, vestigingslocatie, etc.) zijn de data daarom geaggregeerd naar de ondernemingengroep. Voor continue variabelen sommeren we de data van verschillende bedrijfseenheden binnen dezelfde onderneming, en voor categorische variabelen nemen we de waarde van de bedrijfseenheid met de meeste werkzame personen. Voor de sector en vestigingslocatie van de onderneming is een aantal correcties uitgevoerd. Deze zijn in bijlage A verder toegelicht.

Onderzoekspopulatie

Voor de beschrijvende analyses en het schatten van de productiefunctiemodellen maken we verschillende selecties binnen de dataset. Deze selecties sluiten grotendeels aan bij de populatie van de NFO, de bron van de financiële gegevens in onze dataset. Allereerst filteren we op het niet-financiële bedrijfsleven. Dit betreft de marktgerichte bedrijfstakken zonder de overheidssector, de agrarische sector, de financiële dienstverlening, onderwijs, zorg, cultuur, sport en recreatie, belangen- en hobby-verenigingen en overige persoonlijke dienstverlening.¹⁶ Vervolgens beperken we de selectie tot commerciële rechtspersonen, ofwel besloten vennootschappen, naamloze vennootschappen en coöperaties. We verwijderen daarnaast enkele observaties met ontbrekende gegevens over de vestigingsplaats, omdat het voor deze groep niet mogelijk is om de geografische nabijheid ten opzichte van andere bedrijven te berekenen die nodig is voor het in kaart brengen van spillovers via geografische gerelateerdheid. Tot slot selecteren we enkel observaties met een positieve toegevoegde waarde, omdat de relatief kleine groep van observaties met een negatieve toegevoegde waarde de geschatte elasticiteiten sterk vertekent. De totale dataset die gebruikt wordt voor de analyses bevat 1.793.441 waarnemingen, over ongeveer 350.000 unieke ondernemingengroepen.^{17, 18} Deze waarnemingen vertegenwoordigen ongeveer 90 procent van de totale S&O-uitgaven en 50 procent van de totale toegevoegde waarde van alle bedrijven in Nederland over de gehele observatieperiode (2014 – 2022).

3.2 Samenstellen interne en externe kennisvoorraden

Interne kennisvoorraad

De interne kennisvoorraad van een bedrijf is gebaseerd op de cumulatieve S&O-uitgaven, oftewel de huidige eigen S&O-uitgaven en de S&O-uitgaven uit het verleden. In lijn met de literatuur (zie Hall et al., 2010) gebruiken we daarvoor de zogenoemde perpetuele voorraadmethode. In deze methode worden de huidige eigen S&O-uitgaven ieder jaar opgeteld bij de opgebouwde voorraad uit voorgaande jaren. Veroudering van kennis wordt meegenomen door de bestaande kennisvoorraad jaarlijks met een vast percentage af te schrijven. Hiervoor hanteren we het internationaal gangbare afschrijvingspercentage van 15 procent op jaarbasis (Hall et al., 2010). In wiskundige termen ziet dat er als volgt uit:

$$K_{i,t}^I = (1 - \delta)K_{i,t-1}^I + R_{i,t}$$

waarbij $K_{i,t}^I$ de interne kennisvoorraad is van onderneming i op tijdstip t ; $R_{i,t}$ de (reële) huidige S&O-uitgaven; en δ het afschrijvingspercentage.

16 Volgens de standaard bedrijven indeling (SBI) is het de samentelling van hoofdstuk B-N, exclusief K en inclusief 95. Deze selectie staat ook bekend als de *business economy* (zie [cbs.nl](https://www.cbs.nl)).

17 Er kunnen meerdere, elkaar niet-uitsluitende, redenen zijn waarom observaties wegvallen uit de dataset. 34,9 procent van de dataset viel buiten de business economy. 78,8 procent betrof niet-commerciële rechtspersonen (met name zelfstandigen). 0,1 procent van de observaties had geen informatie over de gemeente. 12,4 procent heeft geen positieve toegevoegde waarde. Tezamen reduceren deze selecties de dataset met 89,6 procent. Deze reductie zit met name in het wegvallen van zelfstandigen (geen rechtspersonen) en bedrijven uit sectoren die buiten het niet-financiële bedrijfsleven vallen.

18 In de modelschattingen worden soms nog additionele selecties gemaakt, deze worden dan telkens expliciet weergegeven.

Voor het vaststellen van de kennisvoorraad in het jaar voorafgaand aan de start van de onderzoeksperiode ($K_{i,t=2013}^I$) delen we de S&O-investeringen in het eerste jaar ($R_{i,t=2014}$) door de som van het afschrijvingspercentage en de groeivoet van de S&O-investeringen. De bijbehorende wiskundige formulering luidt:

$$K_{i,t=2013}^I \equiv \frac{R_{i,1}}{\delta + g}$$

waarbij g de (constante) groeivoet is en de andere termen hetzelfde zijn als in de voorgaande formule. Onder de aanname van een constante groeivoet (en afschrijvingspercentage), geeft dit namelijk een benadering van de (steady-state) kenniskapitaalvoorraad (zie pp. 1048-1049, Hall et al., 2010). We stellen de groeivoet g vast op basis van de groeivoet op macroniveau (d.w.z. van alle bedrijven tezamen) over de observatieperiode.¹⁹

We beperken ons bij het berekenen van de interne kennisvoorraden niet tot de onderzoekspopulatie maar nemen alle bedrijfseenheden mee die in de onderzochte periode S&O hebben uitgevoerd. Zodoende hebben we een volledig beeld van de interne kennisvoorraden van alle bedrijfseenheden in Nederland, geaggregeerd op ondernemingengroepsniveau. In de analyses onderzoeken we vervolgens alleen de relatie van interne en externe kennisvoorraden op de toegevoegde waarde van ondernemingen binnen het niet-financiële bedrijfsleven, zoals beschreven in de paragraaf ‘Onderzoekspopulatie’.

Externe kennisvoorraden

Externe kennisvoorraden zijn opgebouwd uit de interne kennisvoorraden van andere bedrijven. Er zijn verschillende methodes om externe kennisvoorraden te bepalen, die onderling verschillen in de aannames over welke bedrijven, in welke mate, en via welke transmissiekanalen gebruik kunnen maken van elkaars kennis. Alle methodes gaan ervan uit dat hoe groter de gerelateerdheid tussen twee bedrijven, hoe aannemelijker het is dat zij van elkaars kennis kunnen profiteren.

De mate van gerelateerdheid tussen twee bedrijven wordt doorgaans weergegeven met een gewicht tussen 0 en 1. Dat gewicht bepaalt welk aandeel van de kennis van het andere bedrijf wordt meegeteld in de berekening van de externe kennisvoorraad. Een bedrijf kan met het ene bedrijf dus een ander gewicht hebben dan met het andere bedrijf. De externe kennisvoorraad van elk bedrijf wordt vervolgens berekend door deze gewichten te vermenigvuldigen met de interne kennisvoorraden van alle andere bedrijven. In wiskundige termen betekent dit dat de externe kennisvoorraad van onderneming i op tijdstip t wordt weergegeven als een gewogen som van de interne kennisvoorraden van alle andere ondernemingen:

$$K_{i,t}^E = \sum_{j \neq i} \omega_{ij,t} K_{j,t}^I$$

waarbij $K_{i,t}^E$ de externe kennisvoorraad is van bedrijf i op tijdstip t , en $\omega_{ij,t}$ de gewichten vertegenwoordigen, die de gerelateerdheid tussen bedrijf i en elk ander bedrijf $j \neq i$ op tijdstip t weergeven. Elke methode om een externe kennisvoorraad samen te stellen kent zijn eigen manier om een set aan gewichten te berekenen.

We gebruiken twee methodes om de gewichten $\omega_{ij,t}$ te berekenen. Ten eerste gebruiken we een *generieke methode* die veronderstelt dat bedrijven ofwel volledig aan elkaar gerelateerd zijn (gewicht is gelijk aan 1) of helemaal niet (gewicht is gelijk aan 0), d.w.z. $\omega_{ij,t} = \begin{cases} 0 \\ 1 \end{cases}$. Ten tweede hanteren we een *specifieke methode* die toestaat dat bedrijven gedeeltelijk aan elkaar gerelateerd kunnen zijn. Hierbij ligt het gewicht 2x tussen 0 en 1, d.w.z. $\omega_{ij,t} \in (0,1)$.

De twee verschillende methodes worden hieronder verder uitgelegd. Ter illustratie zijn in Bijlage B matrices opgenomen met de gewichten die zijn gebruikt voor twee verschillende externe kennisvoorraden. Deze gewichten geven weer in welke mate bedrijven uit verschillende sectoren aan elkaar gerelateerd zijn.²⁰

19 Om de groeivoet te berekenen is een lineaire regressieanalyse uitgevoerd met de totale jaarlijkse S&O-uitgaven als afhankelijke variabele en enkel het jaar als onafhankelijke variabele. Hierbij is nog geen selectie gemaakt op de onderzoekspopulatie, waardoor alle bedrijven in Nederland zijn meegenomen. Ook zijn alle jaren in de onderzoeksperiode meegenomen. De coëfficiënt voor de jaarvariabele is 0,0049 en is statistisch significant. Hiermee stellen we de groeivoet vast op 0,5 procent.

20 De matrices in de bijlage tonen de gewichten voor de generieke sectorale kennisvoorraad en de specifieke cognitieve kennisvoorraad. De matrix met gewichten voor de specifieke technologische kennisvoorraad is niet opgenomen in het rapport, omdat een substantieel deel van de cellen niet kan worden getoond vanwege onthullingsrisico's. De matrix met gewichten voor de generieke gemeentelijke kennisvoorraad is een identiteitsmatrix (enen op de diagonaal en nullen daarbuiten) vergelijkbaar aan matrix voor de generieke sectorale kennisvoorraad, en is vanwege de omvang niet opgenomen in Bijlage B. De gewichten die zijn gebruikt voor de geografische kennisvoorraden zijn niet gebaseerd op CBS-microdata, maar publiek beschikbare informatie (aantal kilometers tussen gemeenten).

Generieke methode: kennisvoorraden binnen dezelfde sector of gemeente

Bij de *generieke methode* wordt aangenomen dat alle bedrijven ofwel aan elkaar gerelateerd zijn ofwel helemaal niet. We kijken binnen deze methode naar twee verschillende transmissiekanalen: bedrijven kunnen aan elkaar gerelateerd zijn omdat ze in dezelfde sector actief zijn en/of bedrijven kunnen aan elkaar gerelateerd zijn omdat ze in dezelfde regio actief zijn. Een belangrijke aanname in de *generieke methode* is dat bedrijven niet kunnen profiteren van kennis buiten de eigen sector of regio.

Een vaak gebruikte methode om de externe kennisvoorraad te meten is door de voorraad kennis te nemen die gegenereerd wordt door alle andere bedrijven in *dezelfde sector* (bijv., Griliches, 1979). De aanname is dat bedrijven alleen kunnen profiteren van kennis ontwikkeld door andere bedrijven in hun eigen sector. In dat geval dragen alle bedrijven binnen een sector volledig bij aan de externe kennisvoorraad in die sector. De gewichten worden in dit geval gelijk aan 1 als bedrijf i en j op tijdstip t in dezelfde sector s actief zijn, en aan 0 als dat niet het geval is. In wiskundige termen betekent dit dat de sectorale gerelateerdheid als volgt wordt bepaald:

$$\omega_{ij,t}^{sector} = \begin{cases} 1, & \text{als } s(i) = s(j), \\ 0, & \text{als } s(i) \neq s(j), \end{cases}$$

waarbij $s(i)$ betekent “de bedrijfstak s van bedrijf i ” en $s(j)$ betekent “de bedrijfstak s van bedrijf j ”.

Eenzelfde methode kan worden gebruikt om de externe kennisvoorraad in *dezelfde regio* te berekenen (bijv., Audretsch & Feldman, 1996). In deze variant kunnen bedrijven alleen gebruikmaken van de kennis van andere bedrijven in hun eigen regio en niet daarbuiten. We hanteren hiervoor de gemeente als regioniveau. Waar in bovenstaande formule de gewichten gelijk waren aan 1 voor bedrijven binnen dezelfde sector, zijn de gewichten in deze variant gelijk aan 1 indien bedrijf i en bedrijf j op tijdstip t in dezelfde gemeente gevestigd zijn. In wiskundige termen komt het bepalen van de regionale gerelateerdheid, $\omega_{ij,t}^{regio}$, neer op dezelfde manier berekening als hierboven, maar dan met een functie $g(i)$ welke staat voor “de gemeente g van bedrijf i ” en $g(j)$ welke staat voor “de gemeente g van bedrijf j ”.

Specifieke methode: kennisvoorraden op basis van technologische, cognitieve en geografische gerelateerdheid

Bij de specifieke methode kunnen bedrijven – anders dan bij de generieke methode – ook profiteren van kennis die is ontwikkeld door bedrijven in andere sectoren of gemeenten. Daarvoor moeten zij wel in enige mate aan elkaar gerelateerd zijn. In dit onderzoek onderscheiden we drie transmissiekanalen waarop kennisverspreiding kan plaatsvinden: bedrijven kunnen van elkaars kennis profiteren wanneer zij (i) vergelijkbare technologieën gebruiken, (ii) vergelijkbare cognitieve vaardigheden inzetten en/of (iii) geografisch dicht bij elkaar zijn gevestigd. Waar de generieke methode uitsluitend werkt met gewichten van 0 of 1, kunnen de gewichten bij de specifieke methode elk waarde tussen 0 en 1 aannemen, afhankelijk van de mate van gerelateerdheid. Net als bij de generieke methode worden deze gewichten en de daaruit voortkomende externe kennisvoorraden exogeen bepaald. Dat wil zeggen dat de gerelateerdheid niet simultaan wordt geschat binnen het productiefunctiemodel, maar vooraf wordt berekend op basis van aparte methoden. Deze methoden beschrijven we hieronder in meer detail.

Technologische gerelateerdheid

Van de drie transmissiekanalen die we binnen de specifieke methode meenemen is de berekening van de mate waarin bedrijven gebruikmaken van dezelfde (nieuwe) technologieën, oftewel hun technologische gerelateerdheid (bijv. Jaffe, 1986; Bloom et al., 2013), het meest gangbaar. In lijn met de literatuur meten wij technologische gerelateerdheid aan de hand van patentgegevens van Nederlandse ondernemingen die tussen 2013 en 2021 zijn ingediend bij Octrooiencentrum Nederland (OCNL) en het Europees Octrooibureau (EPO). Elk patent wordt door het octrooibureau toegewezen aan één of meerdere patentklassen, die aangeven tot welk technologisch domein de uitvinding behoort. Het CBS heeft deze patentklassen geordend in 35 bredere technologiegebieden, op basis van de indeling van de Wereldorganisatie voor de Intellectuele Eigendom (WIPO). Patenten die betrekking hebben op verwante technologieën vallen zo binnen hetzelfde technologiegebied. De dataset bevat circa 41.200 unieke patentfamilies, oftewel groepen van patenten die betrekking hebben op dezelfde uitvinding maar in verschillende landen kunnen zijn ingediend. Op basis van deze gegevens is voor 50 sectoren (op tweecijferig SBI-niveau) berekend welk aandeel van hun

patenten binnen elk van de 35 technologiegebieden valt. Dit levert per sector een technologisch profiel op: een verdeling van de patentactiviteit over de verschillende technologiegebieden.^{21, 22}

De technologische gerelateerdheid wordt vervolgens berekend volgens de methode van Jaffe (1986), wat neerkomt op een niet-gecentreerde correlatiecoëfficiënt van de aandelen patenten per technologie-klasse tussen de sectoren (SBI 2-digit) van bedrijf i en elk ander bedrijf $j \neq i$:

$$\omega_{ij,t}^{\text{technologisch}} = T_{s(i), s(j)} = \frac{F_s F'_s}{(F_{s(i)} F'_{s(i)})^{1/2} (F_{s(j)} F'_{s(j)})^{1/2}}$$

waarbij $F_{s(i)}$ bestaat uit een vector met het aandeel patenten van sector s in de verschillende technologiegebieden. Deze gewichten liggen tussen 0 en 1, omdat niet-gecentreerde correlatiecoëfficiënten altijd tussen 0 en 1 liggen als alle elementen in de vectoren niet-negatief zijn (zoals bij patenten). Vervolgens is getest of de berekende gewichten statistisch significant zijn. Enkel de gewichten die statistisch significant zijn op 5%-niveau²³ zijn vervolgens gebruikt bij het berekenen van de externe technologische kennisvoorraden. De gewichten die niet statistisch significant zijn, zijn op nul gezet.

Cognitieve gerelateerdheid

Niet alle sectoren dienen patenten in²⁴ en bedrijven kunnen ook van elkaar leren als ze dezelfde type vaardigheden gebruiken. Deze zogenoemde *cognitieve gerelateerdheid* dient als een tweede kanaal waarlangs spillovers kunnen plaatsvinden. Om dit te meten gebruiken we de ‘skill relatedness’ index uit Neffke et al. (2017). We brengen eerst in kaart welke personen op t_0 (december 2017) en t_1 (december 2019) werkzaam waren bij welk bedrijf.²⁵ Hierbij is gekeken naar zowel werknemers als zelfstandigen. Wanneer personen bij meerdere bedrijven werkzaam waren, zijn we bij de werknemers uitgegaan van de hoofdbaan op de peildatum en bij de zelfstandigen van hun belangrijkste bedrijf in het peiljaar. Vervolgens kijken we alleen naar personen die van baan tussen bedrijven zijn veranderd. Dit zijn de daadwerkelijke arbeidsstromen tussen de twee peilmomenten.²⁶

De cognitieve gerelateerdheid wordt vervolgens berekend volgens de methode van Neffke et al. (2017), aan de hand van de arbeidsstromen tussen de sectoren (SBI 2-digit) van bedrijf i en elk ander bedrijf $j \neq i$:

$$R_{s(i), s(j)} = \frac{A_{s(i), s(j)} A_{s(j), s(i)}}{A_{s(i), \dots} A_{\dots, s(j)}}$$

waar $A_{s(i), \dots}$ de totale uitstroom uit sector $s(i)$ aangeeft, $A_{\dots, s(j)}$ de totale instroom in sector $s(j)$, $A_{s(i), s(j)}$ de stroom van sector $s(i)$ naar sector $s(j)$, en $A_{\dots, \dots}$ geeft het totaal aantal arbeidsstromen weer. Aangezien de waardes $R_{s(i), s(j)}$ scheef zijn verdeeld, passen we net als Neffke e.a. (2017) een transformatie toe. Deze getransformeerde waardes gebruiken we als gewichten:

$$\omega_{ij,t}^{\text{cognitief}} = \bar{R}_{s(i), s(j)} = \frac{R_{s(i), s(j)} - 1}{R_{s(i), s(j)} + 1}$$

Na deze transformatie liggen alle waarden tussen -1 en 1. Een positieve waarde betekent dat de vaardigheden in beide sectoren aan elkaar gerelateerd zijn. Net als Neffke e.a. (2017) zetten we de negatieve waarden op nul, zodat ze niet meetellen in de berekening van de externe kennisvoorraden. Vervolgens toetsen we de positieve waarden op significantie met behulp van een binomiale toets.²⁷ Alleen de waardes die statistisch significant zijn op 5%-niveau en waar de daadwerkelijke stroom tussen beide sectoren groter dan 10 was, zijn als gewichten gebruikt bij het berekenen van de cognitieve externe kennisvoorraden. Bij de andere sectoren paren is het gewicht gelijk aan nul en wordt het niet meegenomen in de berekening van de voorraden.

21 In bijlage A staat meer achtergrondinformatie beschreven over het gebruik van de patentdata in dit onderzoek.

22 De kennisvoorraden op basis van technologische gerelateerdheid zijn berekend op het niveau van de ondernemingsgroep en niet op het niveau van bedrijfseenheden zoals het geval is bij de andere kennisvoorraden. Dit komt omdat patentgegevens bij het CBS alleen beschikbaar zijn op ondernemingsgroepsniveau. Deze maatstaf wijkt hierin af van de andere maatstaven.

23 Dit zijn de sectorparen met een correlatie groter dan 0,334.

24 Daarom richt veel van de bestaande literatuur zich doorgaans op steekproeven van bedrijven die wél patenteren (zoals in Bloom et al., 2013; Lucking et al., 2019).

25 We kozen voor twee peilmomenten vóór de start van de coronacrisis, omdat deze periode voor veel afwijkende arbeidsstromen zorgde.

26 Meer achtergrondinformatie over de arbeidsstromen staat beschreven in bijlage A.

27 Neffke e.a. (2017) voeren geen significantietoets uit, maar verwijderen enkel alle gerelateerdheidsscores kleiner dan 0 (waarde vóór transformatie), dat wil zeggen waarbij er minder stromen zijn dan verwacht op basis van willekeurigheid. Onze maatstaf is daarom wat strenger.

Geografische gerelateerdheid

De laatste transmissiekanaal is gebaseerd op geografische nabijheid (bijv. Lychagin et al., 2016). Het onderliggende idee is dat kennis zich gemakkelijker verspreidt tussen bedrijven die zich fysiek dicht bij elkaar bevinden – bijvoorbeeld via informele contacten, regionale arbeidsmobiliteit of lokale samenwerkingsverbanden. Dit kan binnen dezelfde gemeente zijn, maar ook daarbuiten in aanliggende gemeenten. Het is daarmee een bredere benadering van geografische gerelateerdheid dan de *generieke methode* die enkel kijkt naar bedrijven binnen dezelfde gemeente. Voor het meten van geografische nabijheid gebruiken we de hemelsbrede afstand in kilometers tussen de gemeenten van bedrijf i en elk ander bedrijf $j \neq i$.

Om te voorkomen dat verafgelegen gemeenten even zwaar meewegen als nabijgelegen gemeenten, is deze afstand getransformeerd met een negatieve exponentiële functie (in lijn met Aldieri en Cincera, 2009). Deze functie zorgt ervoor dat de gewichten sneller afnemen naarmate de afstand groter wordt: bedrijven op korte afstand krijgen dus een relatief hoog gewicht, terwijl de invloed van bedrijven op grotere afstand exponentieel kleiner wordt. Anders gezegd: de functie legt een *penalty* op grotere afstanden, zodat lokale kennisuitwisseling zwaarder meetelt dan landelijke. De formule luidt als volgt:

$$\omega_{ij,t}^{\text{geografisch}} = G_{g(i),g(j)} = \frac{1}{e^{a \cdot d_{g(i),g(j)}}}$$

waarbij $d_{g(i),g(j)}$ de hemelsbrede afstand in kilometers tussen gemeenten $g(i)$ en $g(j)$ weergeeft. Parameter a bepaalt hoe sterk de gerelateerdheid daalt als de afstand toeneemt (afstandsverval of *distance decay* in het Engels). In dit onderzoek hebben we gekozen voor de waarde 0,03. In de praktijk betekent dit dat de gerelateerdheid sterkt daalt na ongeveer 50 km en na 100 km erg klein is.²⁸ Dit vinden wij in een relatief klein land als Nederland aannemelijk. De getransformeerde gewichten liggen altijd tussen 0 en 1. Hoe hoger de waarde, hoe dichter twee gemeenten bij elkaar liggen en hoe zwaarder de kennis van bedrijven gevestigd in deze gemeenten meeweegt als gewicht in de bepaling van externe kennisvoorraad.²⁹

Box 3.1 Rekenvoorbeeld interne en externe kennisvoorraden

In dit simpele rekenvoorbeeld zijn er drie bedrijven. Bedrijf A is actief in SBI 24 heeft een interne kennisvoorraad van 21 miljoen. Bedrijf B is ook actief in SBI 24 en heeft een interne kennisvoorraad van 40 miljoen. Bedrijf C is actief in SBI 72 en heeft een interne kennisvoorraad van 50 miljoen.

De *interne kennisvoorraad* bestaat uit de huidige S&O-uitgaven en de bestaande interne kennisvoorraad. Bedrijf A had vorig jaar een voorraad van 20 miljoen. Deze voorraad wordt jaarlijks afgeschreven met een percentage van 15 procent. De afgeschreven voorraad bedraagt 20 miljoen $\times$ (1 – 0,15) = 17 miljoen. De huidige S&O-uitgaven zijn 4 miljoen. De huidige interne kennisvoorraad bedraagt dan 17 + 4 = 21 miljoen euro.

We berekenen de *externe kennisvoorraden* op basis van de *generieke methode*. Bedrijf A en B zijn volledig aan elkaar gerelateerd (gewicht is 1), want ze zijn actief in dezelfde sector. Bedrijf A en B zijn niet gerelateerd aan bedrijf C, want deze is actief in een andere sector (gewicht is 0). De externe kennisvoorraad van bedrijf A is: 1 $\times$ 40 miljoen (bedrijf B) + 0 $\times$ 50 miljoen (bedrijf C) = 40 miljoen.

Vervolgens berekenen we de *externe kennisvoorraad* op basis van de *specifieke methode*. Bedrijven kunnen hier ook gedeeltelijk aan elkaar gerelateerd zijn (ook buiten de eigen sector). De gerelateerdheid tussen bedrijf A en B heeft een gewicht van 0,90 en tussen bedrijf A en C geldt een gewicht van 0,20. De externe kennisvoorraad van bedrijf A is dan: 0,90 $\times$ 40 miljoen (bedrijf B) + 0,20 $\times$ 50 miljoen (bedrijf C) = 36 + 10 = 46 miljoen.

28 Veel internationaal onderzoek focust zich op grotere landen (bijv. VS, China, Duitsland). De perceptie van afstand is hier anders dan in Nederland. Dat wil zeggen dat een afstand van 50 kilometer in Nederland anders wordt waargenomen dan in een groot land als de VS. Er is in de literatuur geen eenduidig antwoord wat de afstandsverval parameter in Nederland zou moeten zijn. De huidige parameter is gekozen op basis van onze eigen inzichten.

29 Bij de andere twee specifieke externe kennisvoorraden is getest op statistische significantie, waarbij de niet-significante gewichten op nul zijn gezet. Bij specifieke geografische kennisvoorraden is een significantietoets niet mogelijk. Om consistent te blijven met de andere specifieke kennisvoorraden hebben we hier alle paren met een gewicht kleiner dan 0,2 op nul gezet om te voorkomen dat ze toch deels meetellen in elke externe kennisvoorraad. Het aandeel nulwaarden in de gewichtenmatrix is daardoor vergelijkbaar tussen de drie specifieke externe kennisvoorraden.

Berekenen van externe kennisvoorraad voor de hele ondernemingengroep

Voor vrijwel alle externe kennisvoorraden, zowel de generieke als specifieke, geldt dat we deze hebben opgebouwd vanuit de statistische bedrijfseenheden die samen ondernemingengroepen vormen.³⁰ Ondernemingengroepen bestaande uit meerdere bedrijfseenheden kunnen hun S&O-activiteiten verspreid uitvoeren over meerdere regio's en in meerdere sectoren. Als we de kennisvoorraden direct op het niveau van de ondernemingengroep zouden bepalen, hebben enkel bedrijven in de nabijheid van de hoofdlocatie of binnen dezelfde sector als de ondernemingengroep mogelijk profijt van externe kennisspillovers. Dit zou leiden tot een overschatting van de externe kennisvoorraden nabij de ondernemingengroep en een onderschatting van de externe kennisvoorraden nabij de afzonderlijke bedrijfseenheden. Bovendien is het aannemelijker dat kennisuitwisseling in de praktijk eerder plaatsvindt tussen bedrijfseenheden onderling dan op het niveau van de gehele ondernemingengroepen (zie, bijv. Lychagin et al., 2016).

Allereerst zijn de gewogen interne kennisvoorraden van andere bedrijfseenheden in de nabijheid van de bedrijfseenheid gebruikt om de externe kennisvoorraden op het niveau van bedrijfseenheden samen te stellen. Vervolgens zijn deze externe kennisvoorraden geaggregeerd naar de ondernemingengroep. Bij het aggregeren van bedrijfseenheden naar ondernemingengroep wegen we de bedrijfseenheden naar het aandeel van de werkzame personen in de totale ondernemingengroep. Een grotere bedrijfseenheid telt zo zwaarder mee in de externe kennisvoorraad van de onderneming. Ook worden hiermee dubbeltellingen voorkomen. Een onderneming met twee locaties in dezelfde gemeente kan zo niet tweemaal profiteren van de externe kennisvoorraad in deze gemeente.

30 Voor de technologische gerelateerdheid kunnen de externe kennisvoorraden enkel worden berekend op het niveau van ondernemingengroepen. Dit komt door de aard van de patentdata, die enkele op ondernemingsgroepsniveau beschikbaar zijn.

4 Descriptieve statistieken

Dit hoofdstuk toont voor onze onderzoekspopulatie de beschrijvende statistieken over de reguliere productiefactoren, arbeid en kapitaal, en de interne kennisvoorraad. Eerst beschrijven we de totalen en gewogen gemiddelden (over bedrijven) voor het laatste jaar (2022) in onze dataset.³¹ Dit geeft een beeld van de samenstelling van de onderzoekspopulatie. Vervolgens laten we de groei van de variabelen over tijd zien, weergegeven in de vorm van indexcijfers (met 2014 = 100). Daarna tonen we de distributie en samenstelling van de verschillende externe kennisvoorraden.

4.1 Relevante variabelen in de onderzoekspopulatie

In 2022 bestaat de dataset uit 238.501 unieke ondernemingen (Tabel 2). Gezamenlijk creëren zij ruim 491 miljard euro aan toegevoegde waarde, wat neerkomt op gemiddeld ongeveer 2,1 miljoen euro per bedrijf. Het totaal aantal werkzame personen bij deze ondernemingen in 2022 is ruim 4,5 miljoen. De arbeidsproductiviteit komt daarmee uit op 108.500 euro per werkzaam persoon in 2022. De kapitaalafschrijvingen (als proxy voor kapitaal) bedragen gemiddeld 244 duizend euro per bedrijf. De S&O-investeringen per bedrijf zijn gemiddeld 21 duizend euro op jaarbasis, wat laag is doordat een flink aantal bedrijven simpelweg niet in S&O investeert. De berekende interne S&O-kennisvoorraad komt hierdoor uit op gemiddeld 141 duizend euro per bedrijf.

Tabel 2: Totalen en gemiddelden van relevante variabelen, 2022

	2022
Aantal ondernemingen	238.501
Toegevoegde waarde, mln. euro	491.410
per bedrijf, x1000 euro	2.060
Werkzame personen	4.530.746
per bedrijf	19,0
Arbeidsproductiviteit, x1000 euro	108,5
Afschrijvingen, mln. euro	58.225
per bedrijf, x1000 euro	244,1
S&O-investeringen, mln. euro	5.017
per bedrijf, x1000 euro	21,0
S&O-kennisvoorraad, mln. euro	33.565
per bedrijf, x1000 euro	140.734

Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Alle financiële waarden zijn gecorrigeerd voor inflatie en uitgedrukt in contante waarde tegen prijzen van 2022. Arbeidsproductiviteit is berekend per werkzaam persoon.

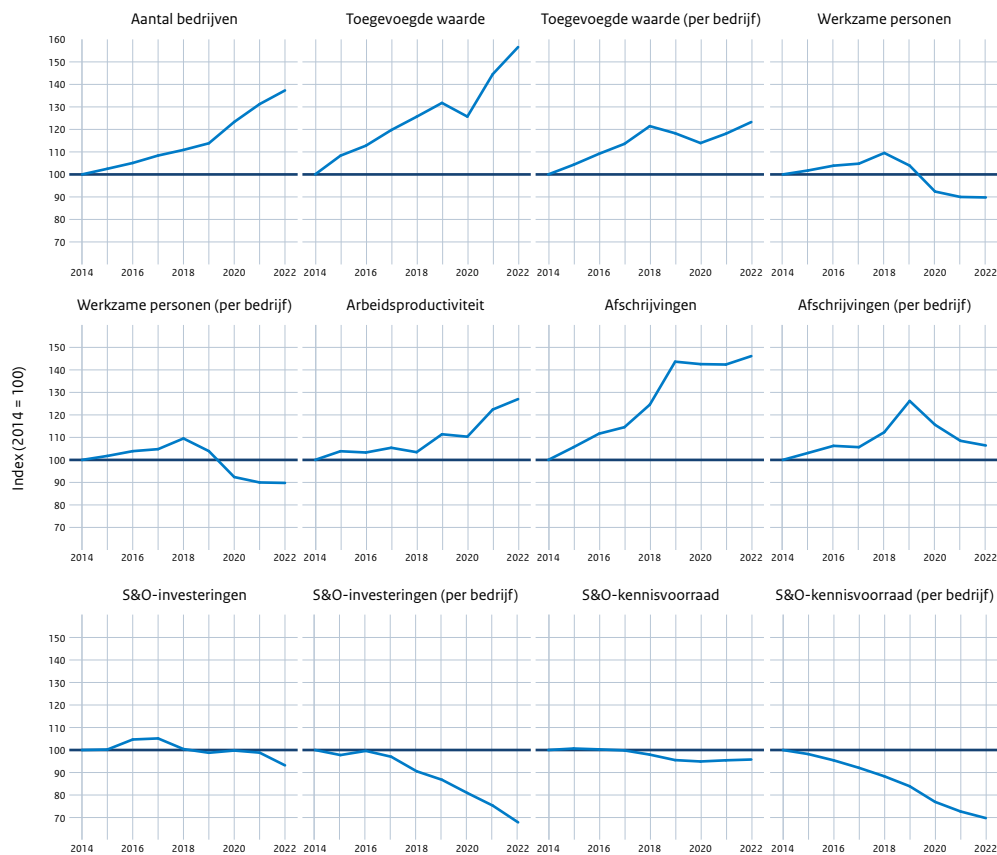
Figuur 2 toont de ontwikkeling van deze variabelen door de tijd, weergegeven in de vorm van indexcijfers (met 2014 = 100).³² Het aantal bedrijven laat een consistente groei zien. Ook de toegevoegde waarde – zowel in totaal als gemiddeld per bedrijf – vertoont een groei, met een duidelijke, maar tijdelijke, coronadip in 2021. Een vergelijkbaar patroon, zij het in mindere mate, is zichtbaar bij het totaal aantal werkzame personen. Het gemiddeld aantal werkzame personen per bedrijf stijgt aanvankelijk tot en met 2018, maar daalt daarna aanzienlijk en ligt in 2022 onder het niveau van 2014. De kapitaalafschrijvingen per bedrijf volgen een vergelijkbare ontwikkeling; na een periode van groei volgt een daling, hoewel het niveau in 2022 nog steeds hoger ligt dan in 2014.

31 Bijlage B geeft dezelfde beschrijvende statistieken voor het eerste jaar van de observatieperiode (2014). De waarden in deze tabellen kunnen afwijken van de landelijke totalen en gemiddelden, door de dataselecties beschreven in paragraaf 3.1.

32 Ontwikkelingen door de tijd kunnen deels het gevolg zijn van compositie-effecten.

De totale S&O-investeringen en -kennisvoorraad zijn min of meer stabiel over de gehele periode, en komen in 2022 iets onder het niveau van 2014 uit. Doordat het aantal bedrijven in de onderzoeksperiode aanzienlijk toeneemt, nemen de gemiddelde investeringen en kennisvoorraad per bedrijf aanzienlijk af.

Figuur 2: Ontwikkeling in relevante variabelen over de periode 2014 – 2022

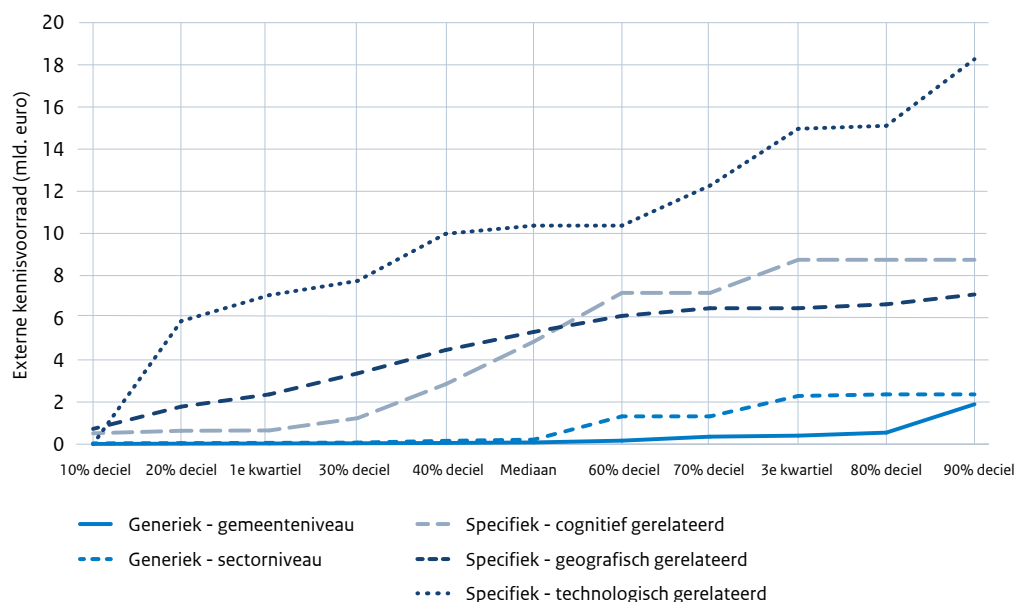


Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Ontwikkelingen weergegeven in de vorm van indexcijfers (2014 = 100).

Externe kennisvoorraden

De externe kennisvoorraden zijn in dit onderzoek op verschillende manieren samengesteld. We maken onderscheid tussen enerzijds twee methodes (generiek en specifiek) en anderzijds verschillende transmissiekanalen (zie ook paragraaf 3.2). In Figuur 3 staan de decielen en de mediaan voor de vijf verschillende externe kennisvoorraden weergegeven. De lijnen voor de twee externe kennisvoorraden gebaseerd op de generieke methode liggen lager dan de lijnen op basis van de specifieke methode. Dit komt omdat bij de eerstgenoemde methode enkel naar externe kennisvoorraden in de eigen sector/regio is gekeken, terwijl de tweede methode ook rekening houdt met externe kennis in andere (gerelateerde) sectoren en regio's. Hierdoor zijn er meer bedrijven die van elkaar kunnen profiteren, waardoor de externe kennisvoorraad hoger ligt. Verder valt op dat er meer variatie is bij de technologische kennisvoorraden. Dit wordt veroorzaakt door grote verschillen in de mate van patenteren tussen sectoren. In sommige sectoren wordt niet gepatenteerd waardoor zij niet blootgesteld worden aan technologische kennisvoorraden (het eerste deciel heeft een kennisvoorraad van nul). In andere sectoren wordt het juist erg veel gebruikt waardoor ze ook blootgesteld worden aan kennis in veel andere sectoren.

Figuur 3: Verdeling van de verschillende externe kennisvoorraden, 2022



Bron: CBS Microdata (2025), bewerking BAT-lab

Figuur 4 toont de tien sectoren en tien gemeenten met de hoogste externe kennisvoorraad, die op basis van de *generieke methode* zijn samengesteld. Per sector of gemeente is de (ongewogen) som van alle interne kennisvoorraden berekend. In lijn met de algemene verdeling van R&D-investeringen en kennis binnen bedrijven, waarbij enkele bedrijven doorgaans het leeuwendeel van de totale investeringen voor hun rekening nemen, is ook de verspreiding van externe kennis sterk geconcentreerd in een beperkt aantal sectoren en/of gemeenten. Zoals blijkt uit het bovenste paneel, staat de sector 'Speur- en ontwikkelingswerk' (SBI 72) op de eerste plek met een externe kennisvoorraad van bijna 7,5 miljard euro, terwijl de sector 'Vervaardiging van farmaceutische grondstoffen en producten' (SBI 21) met een externe kennisvoorraad van 850 miljoen euro op de tiende plek staat. Daarna volgen ongeveer 80 andere sectoren met een nog kleinere kennisvoorraad. Het onderste paneel laat zien dat de externe kennisvoorraad op gemeenteniveau met name geconcentreerd is in de gemeenten (rondom) Eindhoven,³³ en de gemeenten Amsterdam en Utrecht.

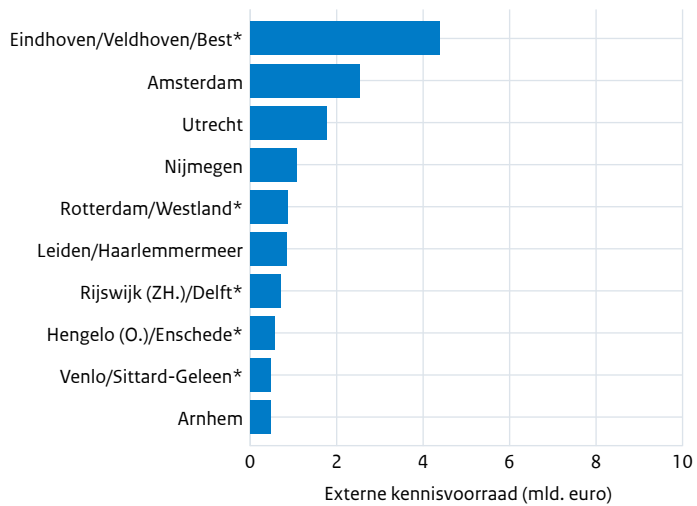
Figuur 4: Externe kennisvoorraad uitgesplitst naar sector (boven) en gemeente (onder) in het jaar 2022, generieke methode

Externe kennisvoorraad op sectorniveau (SBI 2-digit)



33 Het gaat hier om een samenvoeging van drie gemeenten.

Externe kennisvoorraad op gemeenteniveau



Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Enkel de tien sectoren/gemeenten met de grootste externe kennisvoorraad zijn getoond; enkele sectoren/gemeenten zijn samengevoegd ter voorkoming van onthulling door dominantie (d.w.z. sectoren/gemeenten waarbij zonder samenvoeging >50% van de externe kennisvoorraad afkomstig is van één bedrijf). Deze samenvoegingen zijn te herkennen door een *.

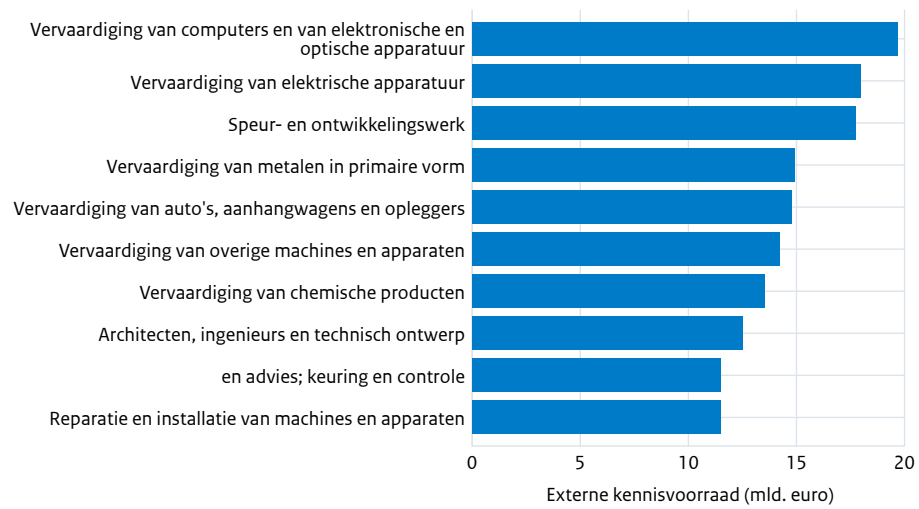
In Figuur 5 worden de externe kennisvoorraden berekend op basis van de technologische, geografische en cognitieve gerelateerdheid getoond. Deze voorraden tonen in grote lijnen hetzelfde beeld als die gebaseerd op de generieke methode. Een uitzondering vormt de geografische gerelateerdheid: daarin verschijnen vrijwel alleen gemeenten rond Eindhoven/Veldhoven/Best, terwijl andere grote steden zoals Amsterdam, Utrecht en Rotterdam niet langer in de top staan. Ook liggen de externe kennisvoorraden een stuk hoger, en zijn de verschillen tussen sectoren en gemeenten veel kleiner dan in de generieke methode. Het eerste komt doordat in de specifieke methode meer bedrijven kunnen profiteren van de kennis van andere bedrijven. Het tweede komt doordat sectoren en gemeenten in de nabijheid van een kennisintensieve sector of gemeente daar nu ook van meeprofiteren, waardoor er clusters ontstaan rond gebieden met een grote externe kennisvoorraad.

Figuur 5: Totale kennisvoorraad uitgesplitst naar technologische (boven), cognitieve (midden) en geografische gerelateerdheid (onder) in het jaar 2022, specifieke methode

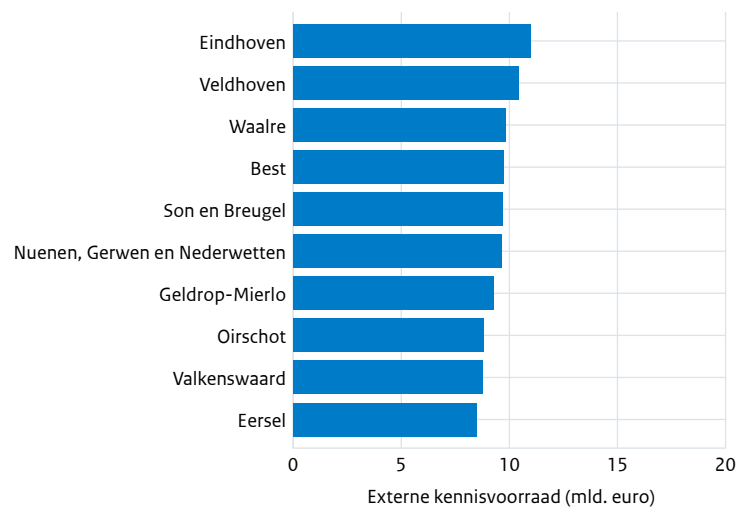
Externe kennisvoorraad naar technologische gerelateerdheid (SBI 2-digit)



Externe kennisvoorraad naar cognitieve gerelateerdheid (SBI 2-digit)



Externe kennisvoorraad naar geografische gerelateerdheid (gemeenten)



Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Enkel de tien sectoren/gemeenten met de grootste externe kennisvoorraad zijn getoond. In tegenstelling tot de externe kennisvoorraden op basis van de generiek methode getoond in Figuur 4, zijn er hier geen sectoren/gemeenten samengevoegd.

5 Econometrische methodologie

Dit hoofdstuk beschrijft de modellen die zijn gebruikt in de econometrische analyses. Middels een regressieanalyse schatten we het verband tussen de toegevoegde waarde van een bedrijf als output-variabele en de productiefactoren arbeid en kapitaal, aangevuld met de interne en externe kennisvoorraden als inputvariabelen. Voor de externe kennisvoorraad hanteren we zowel een generieke als een specifieke methode, zoals uitgelegd in hoofdstuk 3, zodat we de uitkomsten daarvan onderling kunnen vergelijken.

Uit de regressieresultaten leiden we vervolgens af hoeveel extra economische waarde (in de vorm van bbp) een extra euro aan R&D oplevert, zowel voor het bedrijf zelf (privaat rendement) als voor de economie als geheel (publiek rendement). Het private rendement volgt direct uit de regressieresultaten. Voor het spillovereffect vermenigvuldigen we het resultaat van het externe kennisvoorraad met het totaal aantal bedrijven dat hier gebruik van kan maken en een correctiefactor voor de mate van gerelateerdheid. Het publieke rendement is de som van het private rendement en de spillovers. Tot slot voeren we robuustheidsanalyses uit om de gevoeligheid van de resultaten te toetsen. We kijken onder andere naar alternatieve specificaties, datasetselecties, aannames over de opbouw van kennisvoorraden, en de mate waarin bedrijven aan elkaar gerelateerd zijn.

De rest van dit hoofdstuk gaat in op de achterliggende econometrische specificatie van het gebruikte model.

Basismodelspecificatie

Ons econometrisch model gaat uit van een *Cobb-Douglas* productiefunctie met de toegevoegde waarde van een onderneming als outputvariabele en de traditionele productiefactoren (arbeid en kapitaal) en R&D-gerelateerde kennisvoorraden (intern en extern) als inputvariabelen. Hiermee volgen we de traditie van het werk op het gebied van R&D en productiviteit, zoals gepioneerd door Griliches (zie 1998, voor een uitgebreid overzicht).³⁴

In wiskundige termen ziet deze productiefunctie er als volgt uit:

$$Y_{i,t} = A_{i,t} L_{i,t}^{\alpha} C_{i,t}^{\beta} [K_{i,t}^I]^{\gamma} [K_{i,t}^E]^{\theta},$$

waarbij $Y_{i,t}$ de output (toegevoegde waarde) van bedrijf i in jaar t is; $A_{i,t}$ de residuele productiviteits-term; $L_{i,t}$ is arbeid (werkzame personen); $C_{i,t}$ is kapitaal (afschrijvingen); en $K_{i,t}^I$ en $K_{i,t}^E$ respectievelijk interne en externe kennisvoorraad.

Na een log-linearisatie en het toevoegen van bedrijfs- en jaar fixed effects:

$$\log(Y_{i,t}) = \alpha \log(L_{i,t}) + \beta \log(C_{i,t}) + \gamma \log(K_{i,t}^I) + \theta \log(K_{i,t}^E) + \eta_{s(i)} + \varphi_{k(i)} + \delta_t + \varepsilon_{i,t},$$

wat een eenvoudig te schatten lineair regressiemodel is.³⁵ Hierbij hebben we impliciet aangenomen dat $\log(A_{i,t}) = \eta_{s(i)} + \varphi_{k(i)} + \delta_t + \varepsilon_{i,t}$,³⁶ oftewel de residuele productiviteit is op te delen in een bedrijfstak, COROP-regio en een jaarspecifiek effect en een willekeurige storingsterm.

34 Er zijn ook flexibeler specificaties van de productiefunctie mogelijk, zoals *translog*- en *augmented* productiefuncties. Zie bijvoorbeeld Van Elk et al. (2019) voor een toepassing van deze verschillende specificaties in de context van het rendement op (publieke) R&D, evenals een bespreking van de voor- en nadelen van elke benadering.

35 In onze schattingen hanteren we de invers hyperbolische sinustransformatie (*asinh*), wat een benadering is van de gangbare logaritmische transformatie, alleen dan ook gedefinieerd voor negatieve waarnemingen en nullen (Bellemare & Wichman, 2020). Door de hoge aantallen nullen in (onze proxies) voor arbeid, kapitaal en kennis, leidt de logtransformatie van de inputvariabelen tot het verlies van te veel waarnemingen.

36 $s(i)$ betekent “de bedrijfstak s van bedrijf i ” en $k(i)$ de “COROP-regio k van bedrijf i ”.

We hebben tevens geëxperimenteerd met specificaties die bedrijfsspecifieke effecten opnemen. Zoals bekend uit de literatuur (zie onder meer Akerberg et al., 2007; Van Beveren, 2012), leiden dergelijke modellen vaak tot een onderschatting van de kapitaalcoëfficiënt en tot identificatieproblemen, mede doordat de benodigde variatie wordt beperkt tot enkel de tijdsdimensie. In onze toepassing resulteert dit in onvoldoende variatie om de spillovereffecten op robuuste wijze te schatten.³⁷ De gehanteerde modelvariant zonder bedrijfseffecten levert daarentegen economisch plausibele en statistisch stabiele schattingen op, en heeft daarom onze voorkeur.

Afleiden van private en publieke rendement van R&D-investeringen

Aangezien zowel de output als de kennisvoorraden logaritmisches getransformeerd zijn, kunnen de coëfficiënten γ en θ worden geïnterpreteerd als constante elasticiteiten van respectievelijk interne en externe kennis. Met andere woorden: een toename van de interne (of externe) kennisvoorraad met 1 procent leidt tot een stijging van de output met γ procent (of θ procent). Het marginale rendement in een gegeven jaar (bijv.,) $t = 0$ van een euro extra interne (of externe) kennisvoorraad volgt rechtstreeks uit deze elasticiteit:

$$\rho_{t=0} = \frac{\partial Y_{t=0}}{\partial K_{t=0}^I} = \gamma \frac{\bar{Y}_{t=0}}{\bar{K}_{t=0}^I}, \text{ (marginaal rendement op interne kennisvoorraad)}$$

$$\tau_{t=0} = \frac{\partial Y_{t=0}}{\partial K_{t=0}^E} = \theta \frac{\bar{Y}_{t=0}}{\bar{K}_{t=0}^E}, \text{ (marginaal rendement op externe kennisvoorraad)}$$

waarbij $\bar{Y}_{t=0}$, $\bar{K}_{t=0}^I$ en $\bar{K}_{t=0}^E$ de gemiddelde toegevoegde waarde, interne en externe kennisvoorraad in jaar $t = 0$ zijn.

Het marginale rendement op de interne kennisvoorraad, ρ , geeft een directe schatting van het *private rendement* van een investering van één euro in R&D voor het investerende bedrijf zelf. Het marginale rendement op de externe kennisvoorraad, τ , laat zien wat één extra euro externe kennisvoorraad een individueel bedrijf oplevert aan toegevoegde waarde via spillovers. Een belangrijk kenmerk van externe kennis is echter dat *meerdere* bedrijven tegelijkertijd van de externe kennisvoorraad profiteren.

Om de *totale spillovereffecten* van één extra euro R&D-investering te schatten, vermenigvuldigen we daarom het marginale rendement op de externe kennisvoorraad (τ) met het aantal bedrijven dat (potentieel) spillovereffecten geniet ($N - 1$) en de gemiddelde gewichten in een gegeven jaar (wederom, $t = 0$):

$$\tau_{t=0}^{\text{tot}} = \tau_{t=0} \cdot (N - 1) \cdot \bar{\omega}_{ij,t=0}$$

Merk op dat dit voor de generieke kennisvoorraden op sector en gemeenteniveau, equivalent is aan vermenigvuldiging met het gewogen gemiddeld aantal bedrijven per sector of gemeente. Voor de specifieke kennisvoorraden naar technologisch, geografische en cognitieve gerelateerdheid zorgt deze vermenigvuldiging ervoor dat rekening wordt gehouden met de mogelijkheid dat bedrijven slechts gedeeltelijk kunnen profiteren van kennis die door andere bedrijven is gegenereerd. Bijvoorbeeld, twee bedrijven met een gerelateerdheid van $\omega_{ij} = 0,5$ genieten slechts de helft van het marginale rendement op elkaars kennisvoorraad.³⁸

Het *publieke rendement* is vervolgens gelijk aan de som van het private rendement en de totale spillovereffecten: $\kappa_{t=0} = \rho_{t=0} + \tau_{t=0}^{\text{tot}}$.

Langetermijneffecten en R&D-multipliers

Het publieke en private rendement zoals hierboven afgeleid geeft het rendement weer in het jaar dat de R&D-investering plaatsvindt, oftewel het korte termijn rendement. Als we willen weten hoeveel een eenmalige investering van één euro in R&D op de langere termijn aan private en publieke rendementen oplevert, dienen we ook alle toekomstige baten mee te nemen. Hierbij houden we er rekening mee dat kennis veroudert, middels het afschrijvingspercentage dat we ook hanteren voor het samenstellen van de kennisvoorraden, en dat de investeringen over de tijd dus telkens minder opleveren, zowel privaat als publiek. Ook rekenen we alle toekomstige baten terug naar het huidige jaar (verdisconteren), zodat we uiteindelijk uitkomen op de bruto contante baten. Deze bruto contante baten worden vaak de

37 Een specificatie met specifieke effecten op een lager niveau dan bedrijfstak/COROP-regio (bijv. SBI2-digit sector en gemeente) leidt tot dezelfde problemen met een te lage temporele variatie, omdat de externe kennisvoorraden op dit niveau zijn opgebouwd en variëren.

38 We nemen op deze wijze enkel de spillovereffecten binnen onze steekproef mee. Voor zover er nog bedrijven buiten onze steekproef (bijv. bedrijven buiten de NFO, of buitenlandse bedrijven) spillovereffecten genieten, vertegenwoordigen onze schattingen van het publieke rendement een onderschatting van het daadwerkelijke publieke rendement.

‘R&D-multiplier’ genoemd en geven weer hoeveel economische baten voortkomen uit een initiële R&D-investering.

Aangenomen dat het aantal bedrijven, de gerelateerdheidsmaatstaven tussen deze bedrijven en de andere productiefactoren over de tijd niet in omvang wijzigen, is het publieke rendement in jaar t voor een eenmalige investering in jaar $t = 0$ gelijk aan:

$$\kappa_t = \kappa_{t=0}(1-\delta)^t, \forall t$$

waarbij $\delta = 0,15$ het eerder gedefinieerde standaardafschrijvingspercentage is.

Als we deze vergelijking invoegen in de standaardvergelijking voor de bruto constante waarde en dat herschrijven, dan resulteert:

$$GPV = \sum_{t=0}^{\infty} \frac{\kappa_{t=0}(1-\delta)^t}{(1+r)^t} = \kappa_{t=0} \sum_{t=0}^{\infty} \frac{(1-\delta)^t}{(1+r)^t}$$

waarbij $r = 0,25$ de jaarlijkse discontovoet weergeeft. Merk op dat $\sum_{t=0}^{\infty} \frac{(1-\delta)^t}{(1+r)^t} = \frac{(1+r)}{(\delta+r)}$, zodat:

$$GPV = \kappa_{t=0} \frac{(1+r)}{(\delta+r)}$$

waarmee we de bruto contante publieke baten van een extra euro in R&D dus kunnen berekenen als een functie van het publieke rendement in het jaar waarin de investering plaatsvindt, het afschrijvingspercentage en de discontovoet. Middels bovenstaande vergelijking kunnen ook de bruto contante private baten van een extra euro in R&D worden berekend (door $\kappa_{t=0}$ te vervangen door $\rho_{t=0}$). Het verschil tussen beide geeft een schatting van de cumulatieve spillovereffecten op de lange termijn.³⁹

39 De cumulatieve spillovereffecten op de lange termijn kunnen ook berekend worden door $\kappa_{t=0}$ te vervangen door τ^{tot} in de vergelijking voor de bruto contante baten.

6 Resultaten

Dit hoofdstuk beschrijft de resultaten van de econometrische modellen. Eerst beschrijven we de resultaten van de modellen met de externe kennisvoorraden zoals berekend volgens de generieke methode en dan de modellen met de externe kennisvoorraden zoals berekend volgens de specifieke methode. Vervolgens presenteren we een aantal gevoeligheidsanalyses die de robuustheid van de resultaten onderschrijven. Tot slot berekenen en interpreteren we de langetermijneffecten en R&D-multipliers op basis van onze schattingsresultaten.

6.1 Modellen met generieke externe kennisvoorraden

Tabel 3 presenteert onze schattingsresultaten voor de modellen met externe kennisvoorraden gebaseerd op de generieke methode. Model 1 bevat enkel de interne productiefactoren (arbeid, kapitaal en interne kennis) en geen externe kennisvoorraad. Model 2 en Model 3 nemen aanvullend een externe kennisvoorraad mee op respectievelijk sectorniveau (SBI 2-digit) en gemeenteniveau. In Model 4 zijn beide externe kennisvoorraden tezamen meegenomen.

In alle modellen ligt de geschatte coëfficiënt voor arbeid rond de 0,72, voor kapitaal rond de 0,26 en voor interne kennis rond de 0,01. Deze drie factoren samen tellen nagenoeg op tot 1, wat duidt op constante schaalopbrengsten en daarmee consistent is met de gangbare economische theorie over productiefuncties onder competitieve marktstructuren. Het private rendement op kennisinvesteringen op basis van de elasticiteiten in het model ligt tussen de 14 en 17 eurocent per euro R&D, wat betekent dat één extra euro aan interne kennis in een jaar leidt tot gemiddeld 14 tot 17 cent aan extra toegevoegde waarde. Deze uitkomsten liggen in lijn met schattingen uit de bestaande literatuur (Hall et al., 2010; Ugur et al., 2016).

De coëfficiënten voor externe kennis zijn in alle modellen positief en statistisch significant, wat duidt op positieve spillovereffecten. De coëfficiënten voor externe kennis zijn over het algemeen hoger dan de coëfficiënten voor interne kennis, wat te herleiden is naar het feit dat het elasticiteiten betreft en de externe kennisvoorraad (per definitie) veel groter is dan de interne kennisvoorraad. Zoals onze schattingen van het publieke rendement op externe kennis laten zien, genereert één euro extra investering in R&D gemiddeld 15 eurocent aan extra rendement via sectorale spillovers en gemiddeld 5 eurocent aan gemeentelijke spillovers.⁴⁰

Opvallend is dat de sectorale en gemeentelijke spillovers niet afnemen in het model waarin ze tegelijkertijd zijn meegenomen (Model 4), ten opzichte van de modellen waarin ze afzonderlijk zijn geschat (Model 2 en 3). Dit duidt erop dat deze kennisspillovers via twee verschillende transmissiekanalen tot stand komen, waar weinig overlap tussen zit. Een mogelijke verklaring is dat gemeentelijke spillovers vooral worden gedreven door fysieke nabijheid, terwijl sectorale spillovers eerder voortkomen uit overeenkomsten in werkzaamheden en technologieën. Deze twee kanalen lijken daarmee grotendeels onafhankelijk van elkaar te functioneren.

Onze schattingen bevestigen dat het publieke rendement op R&D-investeringen hoger ligt dan het private rendement. Op basis van de uitkomsten uit Model 4 bedraagt het publieke rendement circa 2,4 keer het private rendement.⁴¹ Deze verhouding sluit nauw aan bij de bevindingen van Bloom et al. (2013), waarin een factor van 2,6 werd gevonden tussen publiek en privaat rendement.

40 Merk op dat deze publieke rendementen verdeeld worden over een groot aantal bedrijven – voor één enkel bedrijf is het gemiddelde effect van een euro hogere externe kennisvoorraad vele malen lager (ongeveer 0,005 eurocent voor externe kennis op sectoraal niveau en 0,007 eurocent voor externe kennis op gemeentelijk niveau).

41 Dit is de verhouding tussen publiek en privaat rendement: 0,34/0,14.

Tabel 3: Schattingsresultaten modellen met generieke externe kennisvoorraden

	Afhankelijke variabele: log(toegevoegde waarde)			
	Model 1	Model 2	Model 3	Model 4
Arbeid	0,7162*** (0,0034)	0,7172*** (0,0034)	0,7151*** (0,0034)	0,7160*** (0,0033)
Kapitaal	0,2569*** (0,0023)	0,2575*** (0,0024)	0,2575*** (0,0023)	0,2581*** (0,0023)
Interne kennis	0,0113*** (0,0004)	0,0098*** (0,0004)	0,0111*** (0,0004)	0,0096*** (0,0004)
Externe kennis op sectorniveau (SBI 2-digit)		0,0226*** (0,0039)		0,0231*** (0,0039)
Externe kennis op gemeenteniveau			0,0166*** (0,0014)	0,0170*** (0,0014)
Adj. R ²	0,66761	0,66789	0,66781	0,66811
Aantal waarnemingen	1.793.441	1.793.441	1.793.441	1.793.441
<i>Geschat publiek en privaat rendement</i>				
Privaat rendement	0,17	0,14	0,16	0,14
Spillovers via sectorale externe kennis		0,15		0,15
Spillovers via gemeentelij- ke externe kennis			0,05	0,05
Totaal publiek rendement	0,17	0,29	0,21	0,34

Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Alle inputvariabelen zijn getransformeerd met de invers hyperbolische sinustransformatie (asinh); Bedrijfstak, COROP-regio en jaarspecifieke effecten opgenomen in alle modellen; standaardfouten gegeven tussen haakjes (geclusterd op bedrijfstak-COROP-jaar niveau); *** $p < 0,001$ ** $p < 0,01$ * $p < 0,05$. $p < 0,1$.

6.2 Modellen met specifieke externe kennisvoorraden

Tabel 4 presenteert onze schattingsresultaten voor de modellen met externe kennisvoorraden gebaseerd op de specifieke methode. Model 5 tot en met Model 7 bevatten de externe kennisvoorraden naar technologische, cognitieve en geografische gerelateerdheid afzonderlijk, terwijl Model 8 alle drie de externe kennisvoorraden gelijktijdig meeneemt. De coëfficiënten voor arbeid (~0,72), kapitaal (~0,26) en interne kennis (~0,01) komen overeen met de schattingen in Tabel 3, wederom consistent met constante schaalopbrengsten.

Ook hier zijn de coëfficiënten voor alle externe kennisvoorraden positief en statistisch significant. Via technologische gerelateerdheid wordt er voor elke euro aan R&D gemiddeld 5 tot 6 eurocent aan spillovers gerealiseerd. Dat dit relatief laag is, kan mede komen doordat technologische gerelateerdheid is geoperationaliseerd met patentendata en niet alle sectoren en bedrijven patenteren.⁴² Het is mogelijk dat er ook nog andere sectoren zijn die *niet* patenteren, maar wel profiteren van de technologische kennis die met R&D gegenereerd wordt. Deze spillovers worden niet opgepakt door onze technologische gerelateerdheidsmaatstaf.

Spillovers via cognitieve gerelateerdheid leveren naar schatting gemiddeld ongeveer 35 tot 38 eurocent aan publiek rendement op voor elke euro die in R&D geïnvesteerd wordt. Voor spillovers via geografische gerelateerdheid is dat gemiddeld zelfs 57 eurocent. Net als bij de generieke methode zien we dat de spillovereffecten niet afnemen wanneer de verschillende specifieke externe kennisvoorraden tegelijkertijd meegenomen worden in het model (Model 8) ten opzichte van de afzonderlijke modellen (Model 5 tot en met 7). Dit laat zien dat de verschillende typen van gerelateerdheid additioneel zijn aan

42 In veel van de literatuur vormt dit een minder groot probleem, doordat de analyses zich vaak beperken tot bedrijven die *wel* patenteren (zie, bijv., Bloom et al., 2013).

elkaar en onderstreept het belang om verschillende transmissiekanalen voor spillovers mee te nemen. Interessant is bovendien dat de spillovereffecten geschat middels de specifieke methode veel hoger zijn dan de spillovereffecten, die we hierboven hebben geschat op basis van de generieke methode. Onze interpretatie van deze verschillen behandelen we in de laatste sectie van dit hoofdstuk.

Tabel 4: Schattingsresultaten modellen met specifieke externe kennisvoorraden

	Afhankelijke variabele: log(toegevoegde waarde)			
	Model 5	Model 6	Model 7	Model 8
Arbeid	0,7178*** (0,0034)	0,7179*** (0,0033)	0,7160*** (0,0034)	0,7189*** (0,0033)
Kapitaal	0,2561*** (0,0023)	0,2575*** (0,0024)	0,2570*** (0,0023)	0,2568*** (0,0024)
Interne kennis	0,0107*** (0,0004)	0,0097*** (0,0004)	0,0112*** (0,0004)	0,0092*** (0,0004)
Externe kennis naar technologische gerelateerdheid	0,0058*** (0,0004)			0,0049*** (0,0004)
Externe kennis naar cognitieve gerelateerdheid		0,0402*** (0,0067)		0,0372*** (0,0068)
Externe kennis naar geografische gerelateerdheid			0,0576*** (0,0052)	0,0577*** (0,0052)
Adj, R ²	0,66787	0,66807	0,66767	0,66832
Aantal waarnemingen	1.793.441	1.793.441	1.793.441	1.793.441
Geschat publiek en privaat rendement				
Privaat rendement	0,16	0,14	0,16	0,14
Spillovers via technologische gerelateerdheid	0,06			0,05
Spillovers via cognitieve gerelateerdheid		0,38		0,35
Spillovers via geografische gerelateerdheid			0,57	0,57
Totaal publiek rendement	0,22	0,52	0,73	1,11

Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Alle inputvariabelen zijn getransformeerd met de invers hyperbolische sinustransformatie (asinh); Bedrijfstak, COROP-regio en jaarspecifieke effecten opgenomen in alle modellen; standaardfouten gegeven tussen haakjes (geclusterd op bedrijfstak-COROP-jaar niveau); *** p<0,001 ** p<0,01 * p<0,05 . p<0,1.

6.3 Gevoeligheidsanalyses

Tabel 5 bevat de schattingsresultaten van de uitgevoerde gevoeligheidsanalyses. De verschillende gevoeligheidsanalyses zijn toegepast op alle modellen. Om het overzichtelijk te houden worden uitsluitend de resultaten weergegeven voor Model 4, waarin alle generieke kennisvoorraden zijn opgenomen, en Model 8, waarin alle specifieke kennisvoorraden zijn meegenomen. De gevoeligheidsanalyses voor de overige modellen leveren geen noemenswaardige aanvullende inzichten op. De algehele conclusie uit deze sensitiviteitsanalyses is dat de geschatte coëfficiënten zeer robuust zijn voor verschillende aannames in de opbouw van de dataset, dataselecties en modelspecificaties.

Toestaan van één verhuizing

In onze basisschattingen hebben we de sector en gemeente van ondernemingsgroepen gelijkgetrokken, zodat elke ondernemingsgroep de gehele periode dezelfde sector en gemeente heeft (op basis van de sector/gemeente waarin ze het grootste deel van de periode actief zijn). Dit voorkomt dat bedrijven

abrupte schokken ondervinden in de externe kennisvoorraad (of dat verhuizingen abrupte schokken veroorzaken in de kennisvoorraden van andere bedrijven). In de eerste sensitiviteitsanalyse (Model 4a en 8a) staan we wél toe dat bedrijven (maximaal één keer) veranderen van sector en/of gemeente.⁴³ De geschatte coëfficiënten komen sterk overeen met de coëfficiënten uit de basisschattingen. Zodoende kunnen we stellen dat het gelijktrekken van de sector en gemeente geen doorslaggevende invloed heeft op onze bevindingen.

Externe kennisvoorraad vanuit ondernemingengroepsniveau

In de basisschattingen is de externe kennisvoorraad opgebouwd op basis van de vestigingslocaties en sectoren van alle bedrijfseenheden, waardoor ondernemingen profiteren van de kennis in de nabijheid van al hun bedrijfseenheden. In de tweede sensitiviteitsanalyse (Model 4b en 8b) bouwen de externe kennisvoorraad op vanuit het ondernemingengroep niveau. Een bedrijf profiteert dan alleen maar van de externe kennis in de nabijheid van de hoofdvestiging en hoofdsector. Hoewel dit een sterk versimpelde weergave is van onze basisschattingen, heeft ook dit geen noemenswaardige invloed op de geschatte coëfficiënten.

Gebalanceerd panel

De derde sensitiviteitsanalyse (Model 4c en 8c) schat het model op een gebalanceerd panel van bedrijven die de gehele onderzoeksperiode in de dataset zitten. Hoewel het aantal observaties hierdoor fors afneemt, met ongeveer een derde, blijven ook hier de geschatte coëfficiënten sterk in lijn met de basisresultaten. Wel neemt de verklarende kracht van het model (Adj. R^2) substantieel toe, omdat de groep met bedrijven constant wordt gehouden over de tijd (het model hoeft geen rekening te houden met de entry en exit van bedrijven).

Enkel industriële sector

De vierde sensitiviteitsanalyse (Model 4d en 8d) beperkt de data verder tot enkel bedrijven actief in de bedrijfstak Industrie. De richting en orde van grootte van de geschatte coëfficiënten veranderen niet, maar er treden wel enkele kwantitatieve verschuivingen op. Het effect van arbeid en, met name, kapitaal neemt toe, wat consistent is met de relatief hogere kapitaalintensiteit van industriële bedrijven ten opzichte van andere sectoren. De elasticiteit van interne kennis is juist lager, mogelijk door de grotere interne kennisvoorraad. Voor de generieke externe kennisvoorraad is er ook een duidelijke verschuiving richting een hogere elasticiteit voor externe kennis op sectorniveau, en een lagere voor externe kennis op gemeentelijk niveau. Voor de specifieke externe kennisvoorraad zijn er minder grotere verschillen met de basisresultaten, wat er mogelijk op wijst dat deze methode voor het samenstellen van externe kennisvoorraad robuuster is voor populatieselecties. Dit model heeft veruit de hoogste verklarende kracht (Adj. R^2), wat in lijn is met de homogene(re) selectie van bedrijven waarop het geschat wordt.

Eén jaar vertraagde inputvariabelen

In de vijfde sensitiviteitsanalyse (Model 4e en 8e) worden alle inputvariabelen opgenomen met één jaar vertraging. Dit corrigeert deels voor eventuele endogeniteit, omdat externe schokken in de outputvariabele geen invloed kunnen hebben op de inputvariabelen van het voorgaande jaar.⁴⁴ Het is geruststellend dat dit niet zorgt voor substantieel andere coëfficiënten.

43 We kiezen hier voor maximaal één verhuizing van sector/gemeente, omdat bij meerdere veranderingen binnen de steekproefperiode vermoedelijk sprake is van 'administratieve verhuizingen'.

44 Dit betreft geen volledige correctie voor mogelijk endogeniteitsproblemen in productiefunctieschattingen. Daarvoor kunnen modelspecificaties in de lijn van Olley & Pakes (1996) en/of Petrin & Levinsohn (2003) ingezet worden. Deze modellen kennen echter weer hun eigen aannames en beperkingen, zie Akerberg & Caves (2004); Katayama (2009); Wooldridge (2009).

Tabel 5: Schattingsresultaten gevoeligheidsanalyses

	Incl. één verhuizing		Sector/gemeente op concernniveau		Gebalanceerd panel		Enkel industriële sector		Eén jaar vertraagde inputvariabelen	
	Model 4a	Model 8a	Model 4b	Model 8b	Model 4c	Model 8c	Model 4d	Model 8d	Model 4e	Model 8e
Arbeid	0,7181*** (0,0033)	0,7191*** (0,0033)	0,7160*** (0,0033)	0,7186*** (0,0033)	0,7030*** (0,0028)	0,7068*** (0,0028)	0,7312*** (0,0043)	0,7330*** (0,0042)	0,7054*** (0,0034)	0,7099*** (0,0034)
Kapitaal	0,2574*** (0,0024)	0,2568*** (0,0023)	0,2579*** (0,0023)	0,2567*** (0,0024)	0,2389*** (0,0023)	0,2373*** (0,0024)	0,2684*** (0,0032)	0,2660*** (0,0032)	0,2346*** (0,0025)	0,2334*** (0,0026)
Interne kennis	0,0097*** (0,0004)	0,0093*** (0,0004)	0,0094*** (0,0004)	0,0096*** (0,0004)	0,0100*** (0,0003)	0,0097*** (0,0003)	0,0061*** (0,0004)	0,0066*** (0,0004)	0,0127*** (0,0004)	0,0125*** (0,0004)
Externe kennis op sectorniveau (SBI 2-digit)	0,0228*** (0,0038)		0,0262*** (0,0040)		0,0244*** (0,0040)		0,0442*** (0,0016)		0,0287*** (0,0041)	
Externe kennis op gemeenteniveau	0,0178*** (0,0014)		0,0173*** (0,0015)		0,0210*** (0,0016)		0,0097*** (0,0017)		0,0202*** (0,0016)	
Externe kennis naar technologische gerelateerdheid		0,0050*** (0,0004)		0,0052*** (0,0004)		0,0054*** (0,0003)		0,0056*** (0,0004)		0,0057*** (0,0004)
Externe kennis naar cognitieve gerelateerdheid		0,0366*** (0,0067)		0,0260*** (0,0066)		0,0368*** (0,0067)		0,0703*** (0,0028)		0,0451*** (0,0072)
Externe kennis naar geografische gerelateerdheid		0,0579*** (0,0052)		0,0521*** (0,0050)		0,0675*** (0,0054)		0,0508*** (0,0070)		0,0620*** (0,0058)
Adj. R ²	0,66813	0,66811	0,66814	0,66815	0,72311	0,72327	0,81201	0,81298	0,64855	0,64993
Aantal waarnemingen	1.793.441	1.798.356	1.798.356	1.798.356	1.085.672	1.085.672	149.333	149.333	1.428.664	1.412.745

Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: Alle inputvariabelen zijn getransformeerd met de invers hyperbolische sinustransformatie (asinh); Bedrijfstak, COROP-regio en jaarspecifieke effecten opgenomen in alle modellen; standaardfouten gegeven tussen haakjes (geclusterd op bedrijfstak-COROP-jaar niveau); *** p<0,001 ** p<0,01 * p<0,05 . p<0,1.

6.4 Langetermijneffecten en R&D-multipliers

Op basis van onze schattingsresultaten berekenen we de langetermijneffecten van investeringen in R&D ('R&D multipliers'). We maken hiervoor gebruik van de vergelijking voor de bruto contante waarde van een R&D-investering zoals afgeleid aan het einde van hoofdstuk 5. We presenteren een bandbreedte waarbij het model met de twee generieke externe kennisvoorraden (model 4, uit Tabel 3) als ondergrens fungeert en het model met de drie specifieke externe kennisvoorraden als bovengrens (model 8, uit Tabel 4).

Voor het model met de twee generieke externe kennisvoorraden, komt het totale publieke rendement in het jaar van de initiële R&D-investering ($K_{t=0}$) uit op 36 eurocent (zie Tabel 3). Wanneer we rekening houden met het standaardafschrijvingspercentage van 15 procent op jaarbasis ($\delta = 0.15$) en een jaarlijkse discontovoet van 2,5 procent ($r = 0.025$), en deze getallen invoegen in de vergelijking voor bruto contante waarde, dan resulteert een R&D-multiplier van gemiddelde 2,00 euro aan extra toegevoegde waarde voor elke euro aan R&D-investeringen. Zie de eerste kolom van Tabel 6 voor een schematische weergave van de berekening.

Voor het model met de drie specifieke externe kennisvoorraden ligt het totale publieke rendement in het jaar van de R&D-investering op 1,11 euro (zie Tabel 4). Doorgerekend naar de langetermijn publieke baten, komt dit neer op een gemiddelde R&D-multipliers van 6,50 euro. Zie de tweede kolom van Tabel 6 voor een schematische weergave van de berekening.

Tabel 6: Berekening van langetermijneffecten en R&D-multipliers, ondergrens en bovengrens

		Ondergrens (model 4 in Tabel 3)	Bovengrens (model 8 in Tabel 4)
Econometrische schattingen			
Totaal publiek rendement in jaar t = 0 (euro)	$K_{t=0} = \rho_{t=0} + \tau_{t=0}^{tot}$	0,36	1,11
Private rendement in jaar t = 0 (euro)	$\rho_{t=0}$	0,14	0,14
Spillovereffecten in jaar t = 0 (euro)	$\tau_{t=0}^{tot}$	0,22	0,97
Aannames			
Afschrijvingspercentage (%)	δ		0,15
Discontovoet (%)	r		0,025
Doorrekening			
Bruto contante waarde <i>publiek</i> rendement (euro)	$K_{t=0} \frac{(1+r)}{(\delta+r)}$	2,00	6,50
waarvan <i>privaat</i> rendement (euro)	$\rho_{t=0} \frac{(1+r)}{(\delta+r)}$	0,82	0,82
waarvan <i>spillovers</i> (euro)	$\tau_{t=0}^{tot} \frac{(1+r)}{(\delta+r)}$	1,18	5,68

Bron: CBS Microdata (2025), bewerking BAT-lab. Noot: de ondergrensberekening is gebaseerd op model 4 in Tabel 3 en de bovengrensberekening op model 8 in Tabel 4.

Deze onder- en bovengrens berekening resulteert in een brede bandbreedte voor de R&D-multiplier.⁴⁵ In de literatuur zijn ook sterk uiteenlopende R&D-multipliers gevonden. Donselaar en Koopmans (2015) komen op basis van een meta-analyse uit op een 'best guess'-berekening voor Nederland van 4,50 euro voor elke euro aan privaat geïnvesteerde R&D. Een recentere meta-analyse van Witteman en Bijlsma (2024) komt uit op 2,40 euro aan totale economische baten voor elke euro geïnvesteerd in R&D. Deze schatting ligt dicht bij de 2,60 euro berekend door Erken en Groenewegen (2019) voor elke euro investering in private R&D. Onze schattingen op basis van de generieke methode liggen iets onder deze R&D-multipliers uit deze literatuur, terwijl onze schattingen op basis van de specifieke methode er juist boven liggen.

6.5 Interpretatie en discussie

Onze resultaten laten allereerst zien dat R&D-kennisspillovers in Nederland economisch substantieel zijn. Op basis van de, meer conservatieve, generieke modellen vertegenwoordigen de kennisspillovers ruim zestig procent van het totale publieke rendement.⁴⁶ In de specifieke modellen loopt de omvang van kennisspillovers in het hoogste geval op tot bijna negentig procent van het totale publieke rendement. Deze aanzienlijke spillovers rechtvaardigen overheidsingrijpen in de R&D-stimulerings sfeer. Bovendien komen onze schattingen van het langetermijn bruto contante private rendement onder de één uit, oftewel elke euro die in R&D wordt geïnvesteerd levert naar verwachting minder dan een euro op. Dit benadrukt dat overheidsingrijpen op de marge nodig is om bedrijven aan te zetten in R&D te investeren, aangezien de private investeringsafweging op basis van dit verwachte private rendement negatief uit zou vallen.

Verder zien we dat de geschatte spillovereffecten en publieke baten van R&D op basis van de specifieke methode voor het vaststellen van externe kennisvoorraden aanzienlijk hoger uitvallen dan op basis van de generieke methode. Door deze grote verschillen komen we ook uit op een relatief brede bandbreedte voor onze schatting van het publiek rendement (tussen de 2,00 en 6,50 euro op de langere termijn).

Dit komt waarschijnlijk deels omdat we bij de specifiekere methode een breder beeld hebben van welke bedrijven kunnen profiteren van elkaars kennis, waarbij ook gekeken wordt naar effecten buiten de eigen sector/gemeente.⁴⁷ Mogelijk is de generieke methode gevoelig voor onderschattingen doordat wordt aangenomen dat bedrijven buiten de eigen sectoren en/of gemeenten sowieso niet profiteren, wat het lage(re) publieke rendement dat we hier vinden verklaart.

Daarentegen is het ook mogelijk dat de specifieke methode leidt tot overschattingen door endogeniteitsproblemen, zoals het oppakken van generieke regio- en/of agglomeratie-effecten (bijv. regionale economische schokken die zowel de toegevoegde waarde als de kennisinvesteringen van bedrijven beïnvloeden). Hoewel we corrigeren voor bedrijfstak en jaareffecten en verschillende sensitiviteitsanalyses hebben uitgevoerd,⁴⁸ passen we geen formele endogeniteitscorrecties (à la Olley & Pakes, 1996, en/of Petrin & Levinsohn, 2003) toe.⁴⁹ Met name de geografische spillovers leiden tot onwaarschijnlijk hogere R&D-multipliers op de langere termijn, wat het aannemelijk maakt dat er hier factoren meespelen waarvoor we nog niet geheel gecorrigeerd hebben in onze huidige modelspecificaties.

45 Het grote verschil tussen de onder- en bovenkant van deze schattingen wordt voor een belangrijk deel bepaald door de spillovers via geografische gerelateerdheid in de model met specifieke externe kennisvoorraden. Van de 6,50 euro aan de bovengrens, komt ongeveer de helft voort uit geografische spillovers (en de andere helft uit private baten en technologische en cognitieve spillovers).

46 Te berekenen als *spillovers* gedeeld door *publiek rendement* in Tabel 6.

47 Ook geven de modellen waarin meerdere kanalen voor spillovers zijn meegenomen unaniem aan dat de verschillende kanalen die we hier bekijken (gemeentelijk, sectoraal, technologisch, etc.) additioneel zijn aan elkaar. Modelschattingen waarin enkel één van deze kanalen wordt meegenomen geven dus per definitie een onderschatting. Daarnaast missen we in elk model de buitenlandse spillovers.

48 De sensitiviteitsanalyses (sectie 6.2) bevat een modelvariant waarin we proberen te corrigeren voor endogeniteit door de met één jaar vertraagde productiefactoren en kennisvoorraden in het model op te nemen.

49 Dit vormt wel een belangrijk aanknopingspunt voor vervolgonderzoek.

7 Conclusies

Met dit onderzoek zetten we een eerste stap om het publieke rendement van R&D inclusief kennispillovers in Nederland in kaart te brengen, op basis van microdata over de gehele markteconomie in Nederland. De aanwezigheid van kennispillovers wordt vaak als rechtvaardiging voor overheidsingrijpen genoemd, maar dit wordt zelden empirisch getest. Bestaande literatuur biedt al veel bewijs voor het bestaan van spillovers, maar kent twee belangrijke tekortkomingen: (1) er ontbreekt recent onderzoek naar de Nederlandse context en (2) de focus ligt vaak op (een steekproef van) specifieke technologische sectoren. Ons onderzoek richt zich daarentegen op de volledige Nederlandse business economy in de periode 2014 – 2022.

Daarnaast onderzoeken we via welke transmissiekanalen kennispillovers plaatsvinden. Enerzijds gebruiken we een generieke methode met versimpelende aannames: kennis verspreidt zich uitsluitend binnen de eigen sector of gemeente. Anderzijds passen we een specifiekere methode toe, waarin ook andere sectoren en gemeenten van de nieuwe kennis kunnen profiteren, afhankelijk van de mate van gerelateerdheid tussen sectoren en gemeenten. Deze gerelateerdheid is op drie manieren bepaald: tussen sectoren op basis van technologische (overlap in patentgebieden) en cognitieve gerelateerdheid (op basis van arbeidsstromen) en tussen gemeenten op basis van geografische gerelateerdheid (op basis van de afstand tussen de vestigingslocaties van bedrijven).

Onze resultaten sluiten aan bij de bestaande literatuur en laten zien dat ook in Nederland geldt dat de publieke baten ruim groter zijn dan de private baten. Bij zowel de generieke als de specifieke methode vinden we dat er aanzienlijke kennispillovers plaatsvinden tussen bedrijven, die in verhouding vaak zelfs groter zijn dan het private rendement. Deze spillovers rechtvaardigen overheidsingrijpen in de R&D-stimulerings sfeer. Zo blijkt uit onze schattingen dat het private rendement aan de marge gemiddeld onder één euro per geïnvesteerde euro ligt, wat betekent dat bedrijven zonder overheidsondersteuning onvoldoende prikkels hebben om extra R&D-investeringen te doen.

Daarnaast zien we dat de schattingen van de publieke baten in de modellen met specifieke kennisvoorraden groter zijn dan in de modellen met generieke kennisvoorraden. Dit komt onder andere doordat rekening wordt gehouden met kennisverspreiding buiten de eigen sector en gemeente. Ook tonen de resultaten aan dat de verschillende transmissiekanalen elkaar aanvullen in plaats van elkaar vervangen. Spillovers vinden dus via meerdere kanalen plaats. Voor het verkrijgen van een totaalbeeld van de spillovereffecten en het publieke rendement van R&D is het belangrijk om deze verschillende kanalen mee te nemen.

Voor het afleiden van de lange termijneffecten hebben we de generieke modellen als ondergrens (2,00 euro) en de specifieke modellen als bovengrens gebruikt (6,50 euro). Deze laat zien dat onze generieke schattingen iets onder de R&D-multipliers liggen uit de bestaande literatuur, terwijl onze specifieke schattingen er juist boven liggen. De geografische spillovereffecten zijn bijvoorbeeld veel groter in de specifieke modellen ten opzichte van de generieke modellen. Hier kunnen meerdere verklaringen voor zijn: bijvoorbeeld dat de gemeentegrenzen mogelijk te beperkt zijn in het generieke model, of dat de manier waarop gerelateerdheid is berekend nog verder geoptimaliseerd zou moeten worden (bijv. door beter te corrigeren voor endogeniteit). De technologische spillovers in de specifieke modellen zijn veel lager dan in bestaand onderzoek. Dit komt doordat we kijken naar de gehele business economy waarin een groot deel van de bedrijven simpelweg niet patenteert, terwijl bestaande literatuur vaak focust op een steekproef van bedrijven die allemaal patenteren.

Suggesties vervolgonderzoek

Ons onderzoek dient gezien te worden als een eerste stap naar het schatten van kennisspillovers van R&D in Nederland op basis van microdata. Om te weten bij wie en waar deze spillovers precies plaatsvinden is doorontwikkeling van het model nodig. Idealiter resulteert dit in een model waarmee de (orde van grootte van) maatschappelijke baten van verschillende vormen van R&D-beleid kunnen worden gekwantificeerd en doorgerekend voor een groot deel van de Nederlandse economie.

Hiervoor achten we verschillende uitbreidingen van het model nodig. Nu schatten we een gemiddeld effect voor alle verschillende type bedrijven, terwijl er in beleid ook nadrukkelijk aandacht is voor bijvoorbeeld het stimuleren van innovatie bij het mkb, in specifieke sectoren of bepaalde regio's. Het model houdt op dit moment nog geen rekening met de mogelijkheid dat spillovereffecten variëren over sectoren, regio's, bedrijfsgrootteklassen en andere bedrijfskenmerken. Ook is er op dit moment geen rekening gehouden met dat bedrijven die zelf aan R&D doen mogelijk meer kunnen profiteren van kennisspillovers (de zogenaamde "*two faces of R&D*"). Daarnaast gebruiken we in dit onderzoek S&O als proxy voor R&D, in het volle besef dat dit een onvolledige maatstaf is. Wanneer er meer duidelijkheid is over de oorzaak van de toenemende wig tussen R&D en S&O kan hiervoor gecorrigeerd worden. Tot slot kan in vervolgonderzoek gebruik worden gemaakt van formele endogeniteitscorrecties, om tot nauwkeurigere schattingen te komen dan de huidige aanpak, waarin is gewerkt met een simpele vertraging van de inputvariabelen. Deze toevoegingen helpen om uiteindelijk de maatschappelijke baten van verschillende vormen van R&D-beleid door te rekenen en effectiever vorm te geven.

- Abolhassani, M. (2024). Productivity spillovers of superior firms through worker mobility. *De Economist*, 172: 1-23.
- Akerberg, D. A., & Caves, K. (2004). *Structural Identification of Production Functions-An Application to the Timing of Input Choice*. mimeo.
- Akerberg, D., Benkard, C. L., Berry, S., & Pakes, A. (2007). Econometric tools for analyzing market outcomes. *Handbook of econometrics*, 6, 4171-4276.
- Agrawal, A., Rosell, C., & Simcoe, T. (2020). Tax credits and small firm R&D spending. *American Economic Journal: Economic Policy*, 12(2), 1-21.
- Aldieri, L., & Cincera, M. (2009). Geographic and technological R&D spillovers within the triad: micro evidence from US patents. *Journal of Technology Transfer*, 34: 196-211.
- Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *The American economic review*, 86(3), 630-640.
- Becker, B. (2015). Public R&D policies and private R&D investment: A survey of the empirical evidence. *Journal of economic surveys*, 29(5), 917-942.
- Bijlsma, M., Brouwer, E., De Jong, G., Busschots, T., den Hertog, P., Bongers, F., Smeitink, A., Massop, M. (2024). *Evaluatie van de Innovatiebox 2010-2019. SEO-rapport 2023-125*. Amsterdam: SEO & Dialogic.
- Bloom, N., Schankerman, M., & Van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4), 1347-1393.
- Bloom, N., Van Reenen, J., & Williams, H. (2019). A toolkit of policies to promote innovation. *Journal of economic perspectives*, 33(3), 163-184.
- Brouwer, E., De Jong, G., Content, J., Hertog, P. den, Brennenraedts, R., Smeitink, A., Verhagen, P., Urselmann, E., Bresser, B. (2025). *Evaluatie Wet Bevordering Speur- en Ontwikkelingswerk (WBSO) 2018-2022. SEO-rapport 2025-07*. Amsterdam: SEO
- Chesbrough, H. (2003) *Open Innovation*. Cambridge, MA: Harvard University Press.
- Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: the two faces of R & D. *The economic journal*, 99(397), 569-596.
- Dechezleprêtre, A., Einiö, E., Martin, R., Nguyen, K. T., & Van Reenen, J. (2023). Do tax incentives increase firm innovation? An RD design for R&D, patents, and spillovers. *American Economic Journal: Economic Policy*, 15(4), 486-521.
- De Jong, G., Lensink, A., Brouwer, E. (2025). *Toegepast onderzoek voor en met bedrijven. De impact van toegepast onderzoek organisaties (TO's) op het bedrijfsleven*. SEO-rapport 2025-40. Amsterdam: SEO
- Erken, H., Groenewegen, J. (2019). 50 miljard euro investeren in onderwijs en innovatie verdubbelt economische groei", *Me Judice*, 3 oktober 2019.
- Grant, R.M. (1996). Toward a Knowledge-Based Theory of the Firm. *Strategic Management Journal*, 17, 109-122.
- Griffith, R., Redding, S., & Reenen, J. V. (2004). Mapping the two faces of R&D: Productivity growth in a panel of OECD industries. *Review of economics and statistics*, 86(4), 883-895.
- Hall, B. H., Mairesse, J., & Mohnen, P. (2010). Measuring the returns to R&D. In *Handbook of the Economics of Innovation* (Vol. 2, pp. 1033-1082). North-Holland.
- Hernandez, E., Kee, J.K., & Shaver, J.M. (2025). Toward an improved causal test of network effects: Does alliance network position enhance firm innovation? *Strategic Management Journal*, 46(4), 863-897.
- Jacobs, B., Nahuis, R., & Tang, P. J. (2002). Sectoral productivity growth and R&D spillovers in the Netherlands. *De Economist*, 150(2), 181-210.

- Jaffe, A. B. (1986). Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value. *American Economic Review*, 76, 984-1001.
- Katayama, H., Lu, S., & Tybout, J. R. (2009). Firm-level productivity studies: illusions and a solution. *International Journal of Industrial Organization*, 27(3), 403-413.
- König, M. D., Liu, X., & Zenou, Y. (2019). R&D networks: Theory, empirics, and policy implications. *Review of Economics and Statistics*, 101(3), 476-491.
- Koopmans, C. & Donselaar, P. (2015). Een meta-analyse van het effect van R&D op productiviteit. *Economische Statistische Berichten*, 100(4717).
- Lucking, B., Bloom, N., & Van Reenen, J. (2019). Have R&D spillovers declined in the 21st century?. *Fiscal Studies*, 40(4), 561-590.
- Lychagin, S., Pinkse, J., Slade, M. E., & Reenen, J. V. (2016). Spillovers in space: does geography matter?. *The Journal of Industrial Economics*, 64(2), 295-335.
- Mazzucato, M. (2018). Mission-oriented innovation policies: challenges and opportunities. *Industrial and corporate change*, 27(5), 803-815.
- Møen, J., & Sandvig Thorsen, H. (2017). Publication bias in the returns to R&D literature. *Journal of the Knowledge Economy* 8(3), 987-1013.
- Neffke, F. M., Otto, A., & Weyh, A. (2017). Inter-industry labor flows. *Journal of Economic Behavior & Organization*, 142, 275-292.
- Olley, G. S., & Pakes, A. (1996). The Dynamics of Productivity in the Telecommunications Equipment Industry. *Econometrica*, 1263-1297.
- Levinsohn, J., & Petrin, A. (2003). Estimating production functions using inputs to control for unobservables. *The review of economic studies*, 70(2), 317-341.
- Soete, L. L., & Ter Weel, B. J. (1999). Innovation, knowledge creation and technology policy: the case of the Netherlands. *De Economist*, 147(3), 293-310.
- Stoyanov, A., & Zubanov, N. (2014). The distribution of the gains from spillovers through worker mobility between workers and firms. *European Economic Review*, 70, 17-35.
- Rao, N. (2016). Do tax credits stimulate R&D spending? The effect of the R&D tax credit in its first decade. *Journal of Public Economics*, 140, 1-12.
- Ugur, M., Trushin, E., Solomon, E., & Guidi, F. (2016). R&D and productivity in OECD firms and industries: A hierarchical meta-regression analysis. *Research Policy*, 45(10), 2069-2086.
- Ugur, M., Churchill, S. A., & Luong, H. M. (2020). What do we know about R&D spillovers and productivity? Meta-analysis evidence on heterogeneity and statistical power. *Research Policy*, 49(1), 103866.
- Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, 26(1), 98-128.
- Wooldridge, J. M. (2009). On estimating firm-level production functions using proxy variables to control for unobservables. *Economics letters*, 104(3), 112-114.

Bijlage A: Achtergrondinformatie samenstelling dataset

In deze bijlage wordt meer achtergrond gegeven over de dataset die voor dit onderzoek is samengesteld.

Correcties SBI- en gemeentecodes

Niet voor alle ondernemingengroepen is sector- of vestigingslocatie beschikbaar. Informatie over de gemeente komt uit het ABR. Wanneer deze informatie mist, is dit aangevuld aan de hand van de postcode (indien deze wel beschikbaar was) of op basis van de gemeentecode uit een toekomstig jaar (indien beschikbaar). Hierdoor daalde het percentage observaties met een missende gemeente in het ABR van 0,21% naar 0,13%. Wanneer de SBI-code van een ondernemingengroep miste, omdat deze ondernemingengroep niet voorkwam in de NFO, is de SBI-code uit het ABR van de grootste onderliggende bedrijfseenheid genomen.

Daarnaast kunnen ondernemingen over tijd van sector en/of vestigingslocatie veranderen (4,6 procent wisselt ten minste één keer van sector en 11,6 procent wisselt ten minste één keer van gemeente). Naast daadwerkelijke verhuizingen of veranderingen in de werkzaamheden van het bedrijf, kunnen deze veranderingen ook administratief van aard zijn of het gevolg van andere databewerkingen. Aangezien sector en gemeente centraal staan in het bepalen van de kennisvoorraden, kunnen veranderingen hierin de kennisvoorraden waar bedrijven toegang tot hebben sterk beïnvloeden. We zijn met name geïnteresseerd in de relatie tussen externe kennisvoorraden en bedrijfsprestaties, en niet in de invloed van veranderingen in sector of locatie. We hebben daarom de meest voorkomende waarde voor de gemeente en de sector geselecteerd.

S&O-uitgaven als proxy voor R&D

We gebruiken de uitgaven aan S&O onder de WBSO als proxy voor R&D-investeringen, in het volle besef dat dit een onvolledige maatstaf is. Niet alle R&D-investeringen vallen onder de definitie van S&O zoals die in de WBSO wordt gehanteerd en niet alle bedrijven die in S&O investeren, vragen daarover WBSO aan (zie de recente WBSO-evaluatie van Brouwer et al. 2025, voor een uitgebreide uitleg hierover). Over de verschillen tussen S&O en R&D – en de oorzaken daarachter – is nog teveel onduidelijk om hier op een zinnige manier voor te corrigeren. De S&O-uitgaven zijn daarbij integraal beschikbaar — dat wil zeggen: voor alle ondernemingen en jaren — in tegenstelling tot de R&D-investeringen die door het CBS slechts steekproefsgewijs worden gemeten. Een integrale waarneming van S&O-uitgaven is essentieel voor de opbouw van kennisvoorraden in onze methode. Informatie over uitgaven is enkel voor inhoudingsplichtigen beschikbaar. Zelfstandigen blijven daardoor buiten beschouwing, wat gegeven onze afbakening naar rechtspersonen niet tot een groot verlies aan informatie leidt. In de definitie van S&O-uitgaven nemen we zowel de loon- als niet-loonuitgaven mee. De uitgaven zijn gecorrigeerd voor inflatie op basis van macro-indices en uitgedrukt in contante waarde tegen prijzen van 2022.

Achtergrondinformatie over het bepalen van arbeidsstromen

De cognitieve gerelateerdheid tussen bedrijven wordt bepaald op basis van arbeidsstromen tussen sectoren. De arbeidsstromen zijn gebaseerd op wisselingen in banen van werknemers en op basis van wisselingen in de activiteiten van zelfstandigen.

Eerst hebben we op basis van de SPOLISHOOFDBAAN⁵⁰ in kaart gebracht welke personen op t_0 (december 2017) en t_1 (december 2019) werkzaam waren bij welke onderneming. We kozen voor twee peilmomenten voor de start van de coronacrisis, omdat deze periode voor veel afwijkende arbeidsstromen zorgde. In de SPOLIS⁵¹ komen ook een aantal fictieve bedrijven voor. Deze zijn buiten beschou-

50 Zie [documentatierapport van het CBS over SPOLISHOOFDBAAN](#).

51 Zie [documentatierapport van het CBS over SPOLIS](#)

wing gelaten. Via de SPOLIS en de BETAB⁵² is het bedrijf en de bijbehorende SBI-code geïdentificeerd. Er is een selectie gemaakt op minimaal 40 en maximaal 200 gewerkte uren per maand, zodat bijbanen buiten beschouwing blijven.

Voor zelfstandigen is via de INPBEID⁵³ bekend wat hun belangrijkste bedrijf was in een bepaald jaar. Directeur-groootaandeelhouders (DGA's) zijn verwijderd, want deze komen als werknemers zijnde namelijk ook voor in de SPOLIS. Via het BDK⁵⁴ koppelen we de SBI-code en de oprichtings- en opheffingsdatum aan. Enkel bedrijven die op peilmoment t_0 (december 2017) en t_1 (december 2019) bestonden zijn geselecteerd.

Als een persoon zowel als werknemer als zelfstandige aan de slag was op het peilmoment is op basis van de voornaamste inkomensbron uit de SECMBUS⁵⁵ een selectie gemaakt. Als de voornaamste inkomensbron 'zelfstandige' was, dan worden de gegevens van de zelfstandige geselecteerd. Als dit niet zo was, dan worden de gegevens van de werknemer gebruikt.

Ieder persoon op t_0 heeft nu een bedrijf en een SBI-code. Hieraan is de situatie op t_1 gekoppeld. Een deel van de personen is niet meer werkzaam op t_1 (12%). Een groot deel is nog steeds actief bij hetzelfde bedrijf (68,3%). Beide situaties worden buiten beschouwing gelaten in ons verdere onderzoek zodat we enkel records meenemen waar personen van bedrijf zijn gewisseld (en die daadwerkelijk arbeidsstromen weergeven) (19,6%).

Achtergrondinformatie over het gebruik van patenten

Technologische verwantschap geeft aan of bedrijven gebruikmaken van dezelfde (nieuwe) technologieën. De standaardmethode in de literatuur om technologische gerelateerdheid te berekenen is op basis van patentklassen in patentaanvragen. Een patent reflecteert de ontwikkeling van een nieuwe technologie.

Het CBS heeft gegevens beschikbaar over patentaanvragen door Nederlandse bedrijven bij OCNL en de EPO.⁵⁶ In de periode 2013-2021 gaat dit om zo'n 41.200 unieke patentfamilies, oftewel groepen van patenten die betrekking hebben op dezelfde uitvinding maar in verschillende landen kunnen zijn ingediend. De patenten zijn door het CBS gekoppeld aan een ondernemingengroep. Hierdoor is het voor deze maatstaf niet mogelijk om eerst de kennisvoorraden voor bedrijfseenheden te berekenen en vervolgens te aggregeren. Deze maatstaf wijkt hierin af van de andere maatstaven. Patenten die door het CBS niet gekoppeld konden worden aan specifieke bedrijven zijn verwijderd. Ook bedrijven met een onbekende SBI-code zijn verwijderd. Een patent kan bij zowel EPO als OCNL ingediend zijn. Als ze bij beide instanties zijn ingediend, is gekozen voor de EPO-aanvraag, aangezien deze een groter bereik heeft.

Elk patentaanvraag is door het octrooibureau toegewezen aan één of meerdere patentklassen. Deze klasse geeft een technologisch gebied weer. De patenten zijn door het CBS op basis van deze patentklassen ingedeeld in 35 klassen op basis van de technologie-concordantietabel van de World Intellectual Property Organisation (WIPO). De microdata van het CBS bevat helaas niet de originele patentklassen, waardoor onze maatstaf minder nauwkeurig is dan bestaand onderzoek. De CBS-microdata bevat ook een indeling op basis van de IPC-classificatie. Deze bestaat uit 25 klassen en is minder gedetailleerd dan de WIPO-indeling. Wanneer een patent betrekking heeft op meerdere technologiegebieden is door het CBS een gewicht per gebied bepaald (gebaseerd op hoe vaak een code voorkomt in een patentaanvraag). Het totale gewicht per patentaanvraag is 1. In sommige SBI-codes wordt er erg weinig gepatenteerd. We nemen enkel SBI-codes mee waarin in de gehele onderzoeksperiode meer dan 30 patenten zijn aangevraagd. Alleen voor de dan resterende 50 unieke SBI-codes kunnen technologische gerelateerdheidsscores worden berekend.

52 Zie [documentatierapport van het CBS over BETAB](#)

53 Zie [documentatierapport van het CBS over INPBEID](#)

54 Zie [documentatierapport van het CBS over BDK](#)

55 Zie [documentatierapport van het CBS over SECMBUS](#)

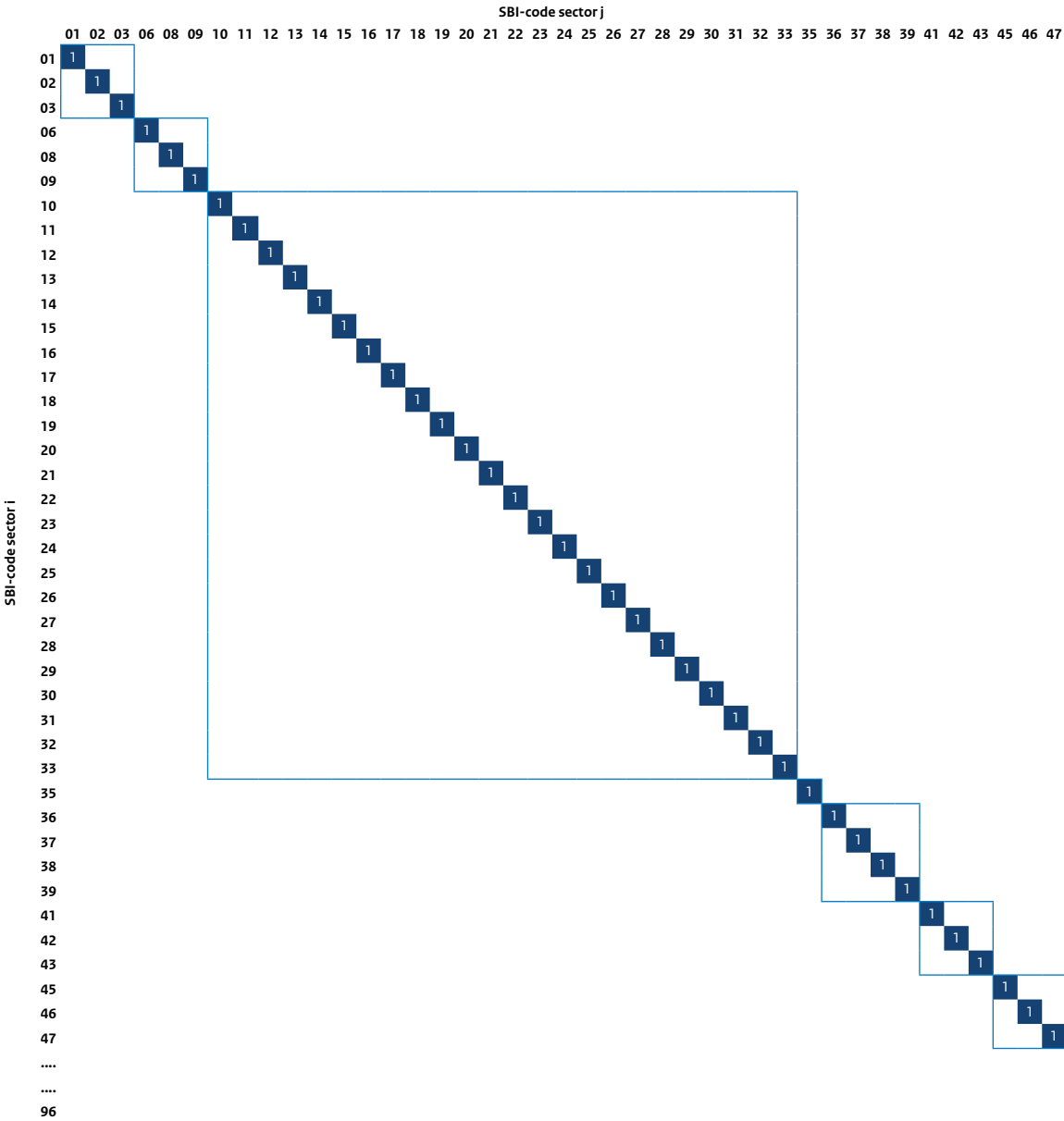
56 Zie [documentatierapport van het CBS over PAT](#)

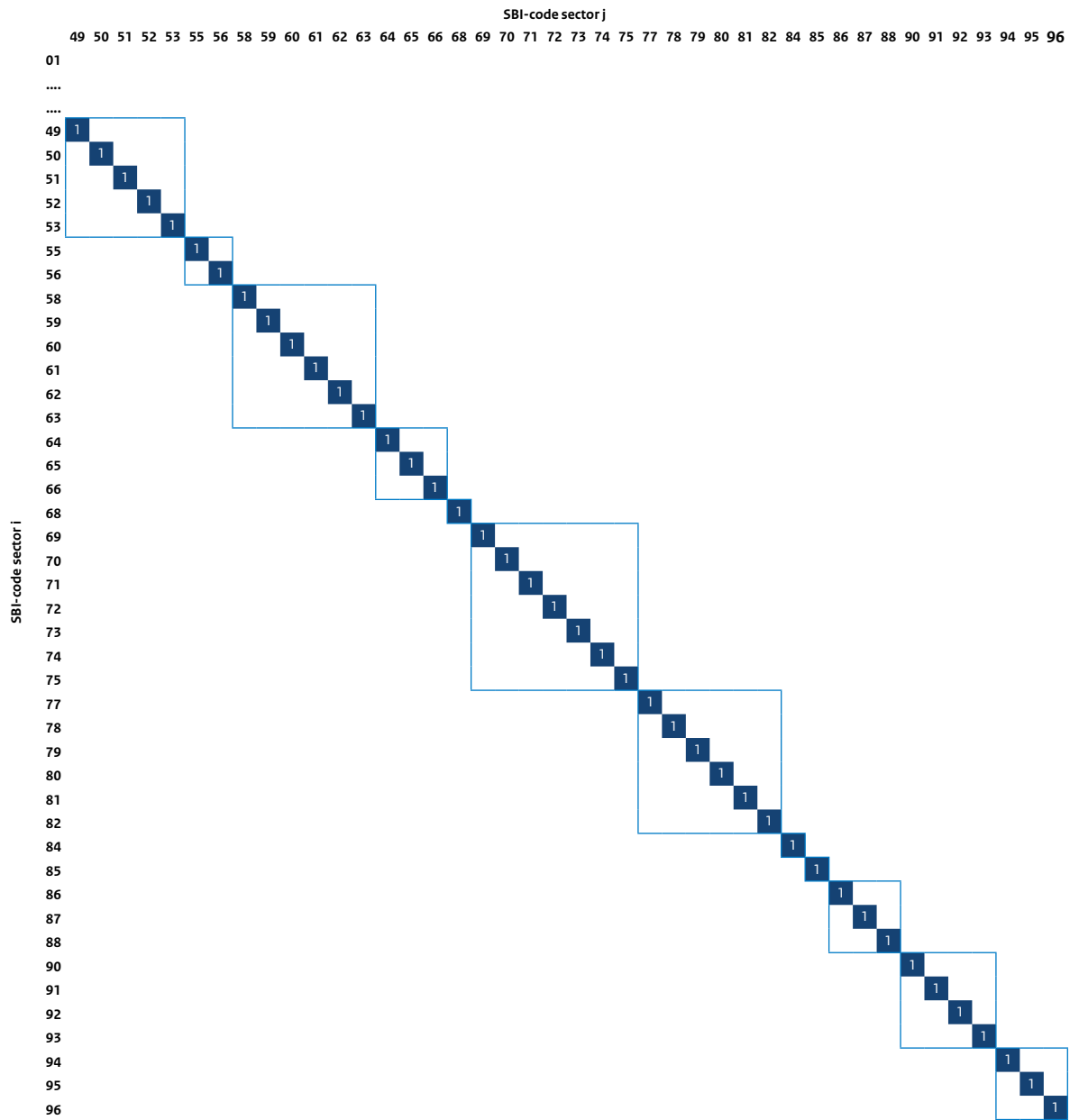
Bijlage B: Additionele figuren en tabellen

Bij de *generieke* methode wordt aangenomen dat alle bedrijven binnen dezelfde sector aan elkaar gerelateerd en bedrijven buiten de eigen sector niet gerelateerd zijn. Hierdoor ontstaat een zogenaamde identiteitsmatrix waarbij de diagonaal gelijk is aan 1 en alle andere cellen gelijk aan 0 (zie Tabel B.1). De lichtblauwe lijnen geven de grenzen van een bedrijfstak aan.

Dezelfde redenatie geldt voor het berekenen van de gewichten tussen gemeenten: bedrijven in dezelfde regio zijn volledig aan elkaar gerelateerd (gewicht is gelijk aan 1) en bedrijven buiten de eigen regio zijn niet gerelateerd. Deze matrix is vanwege de omvang niet weergegeven.

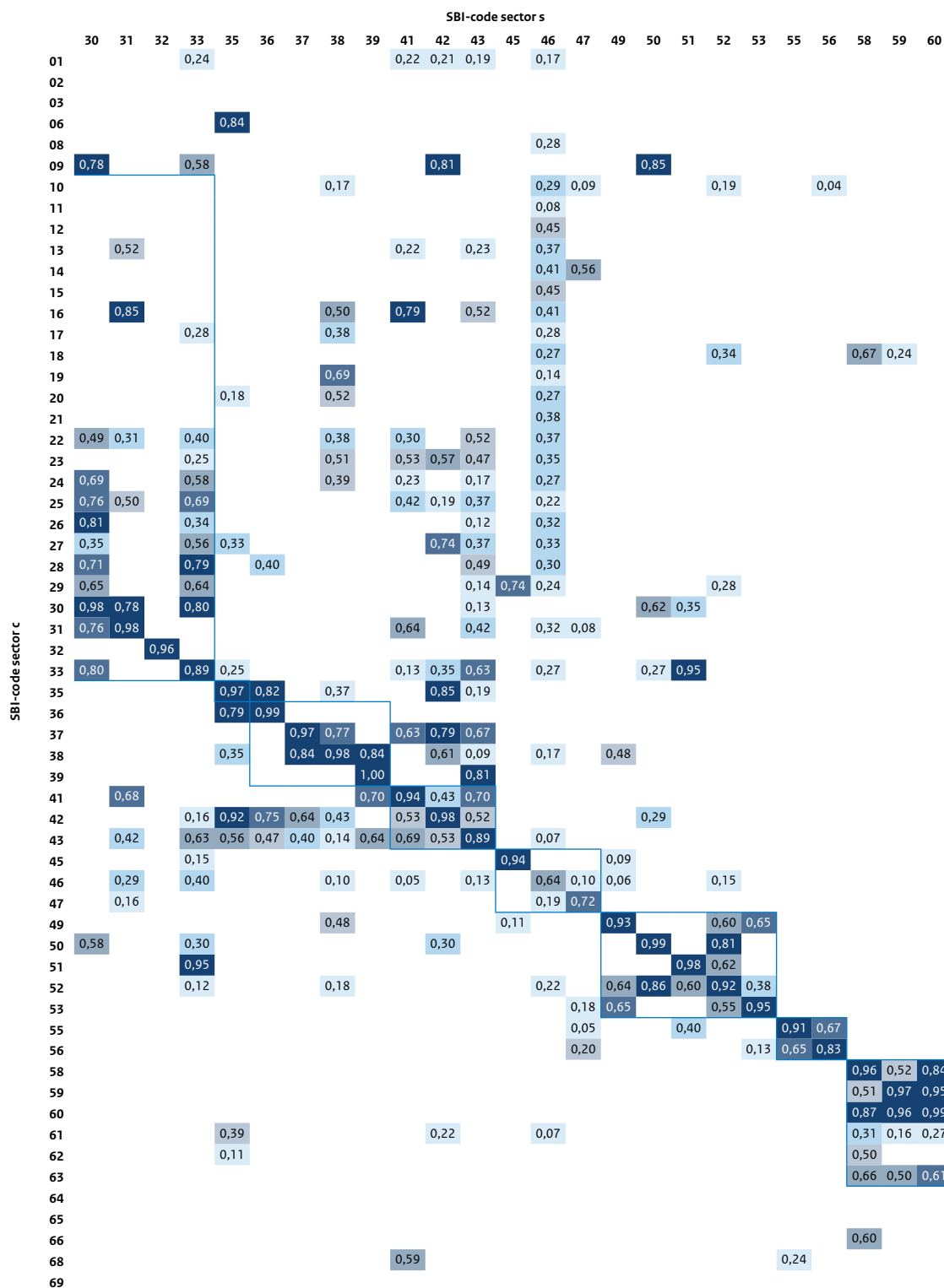
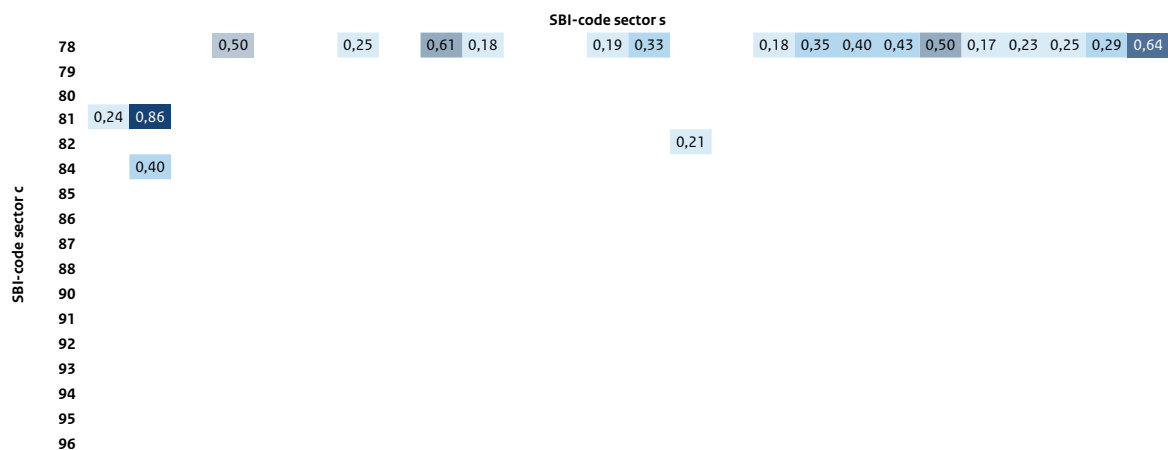
Tabel B.1: Generieke gewichten op sector-niveau (SBI 2-digit).

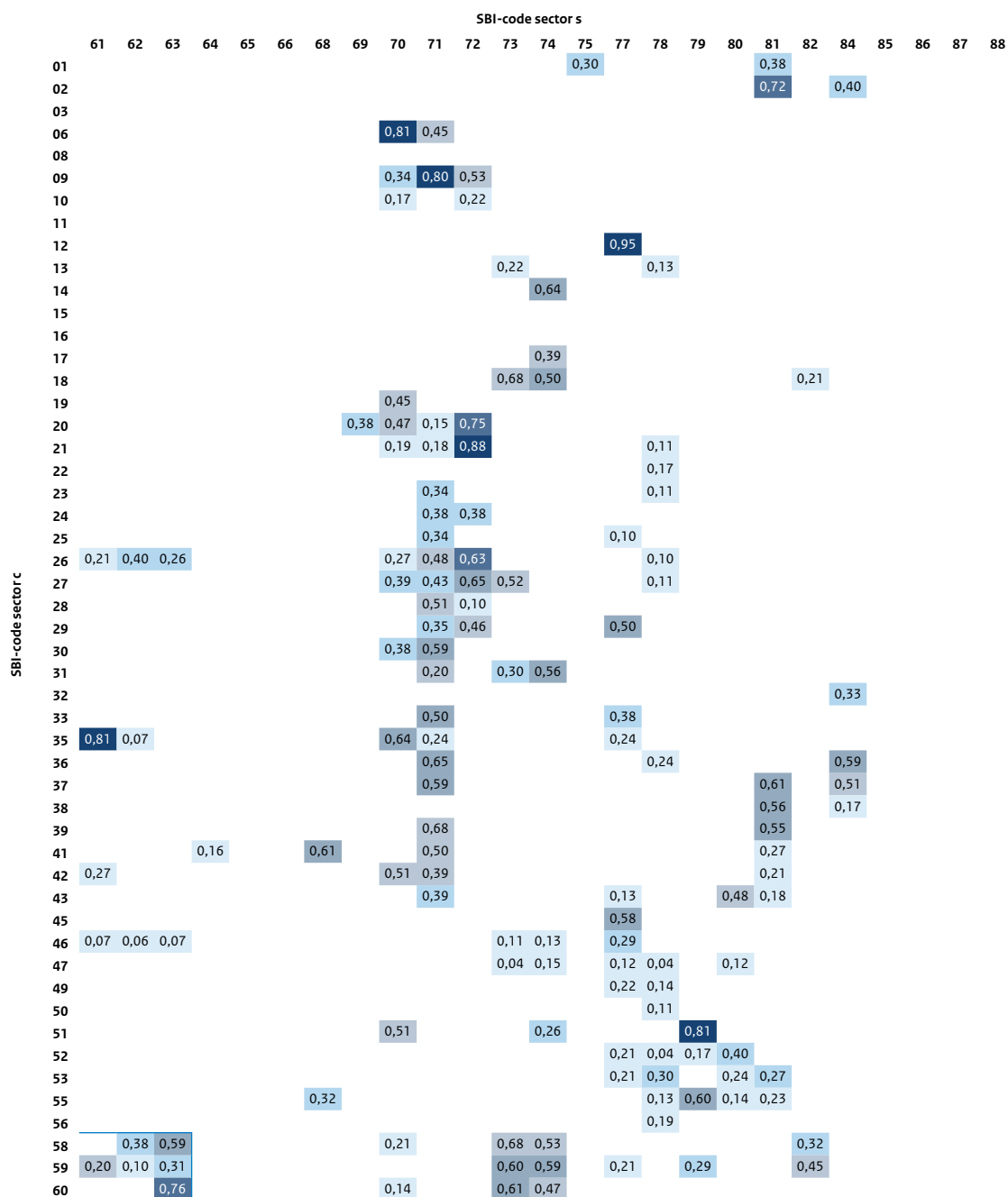
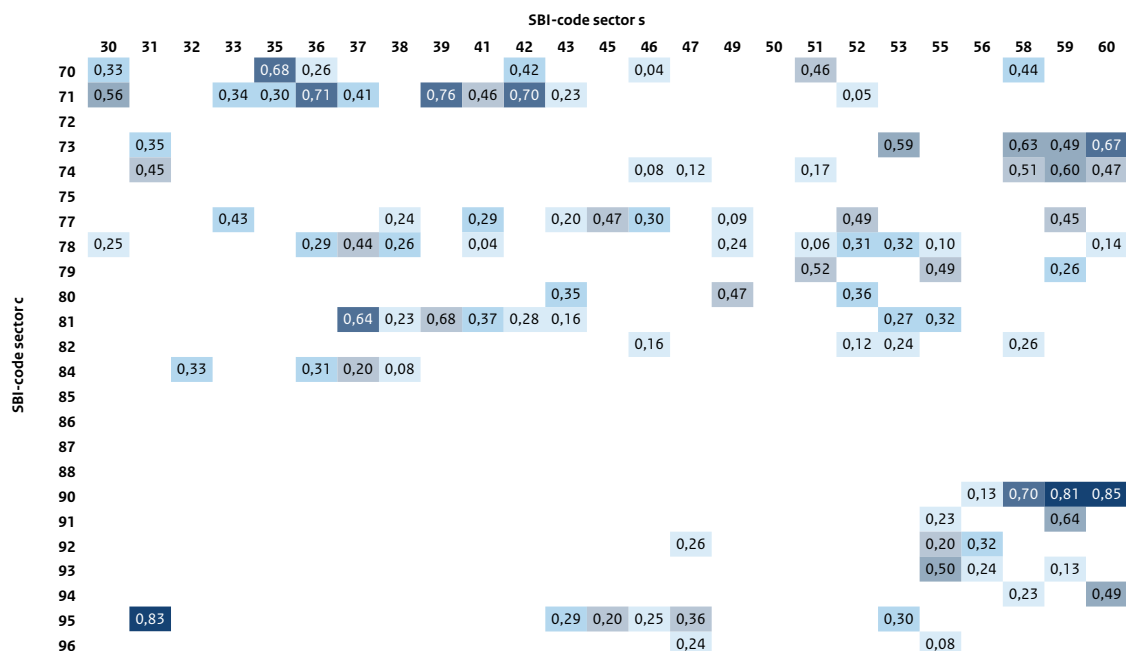


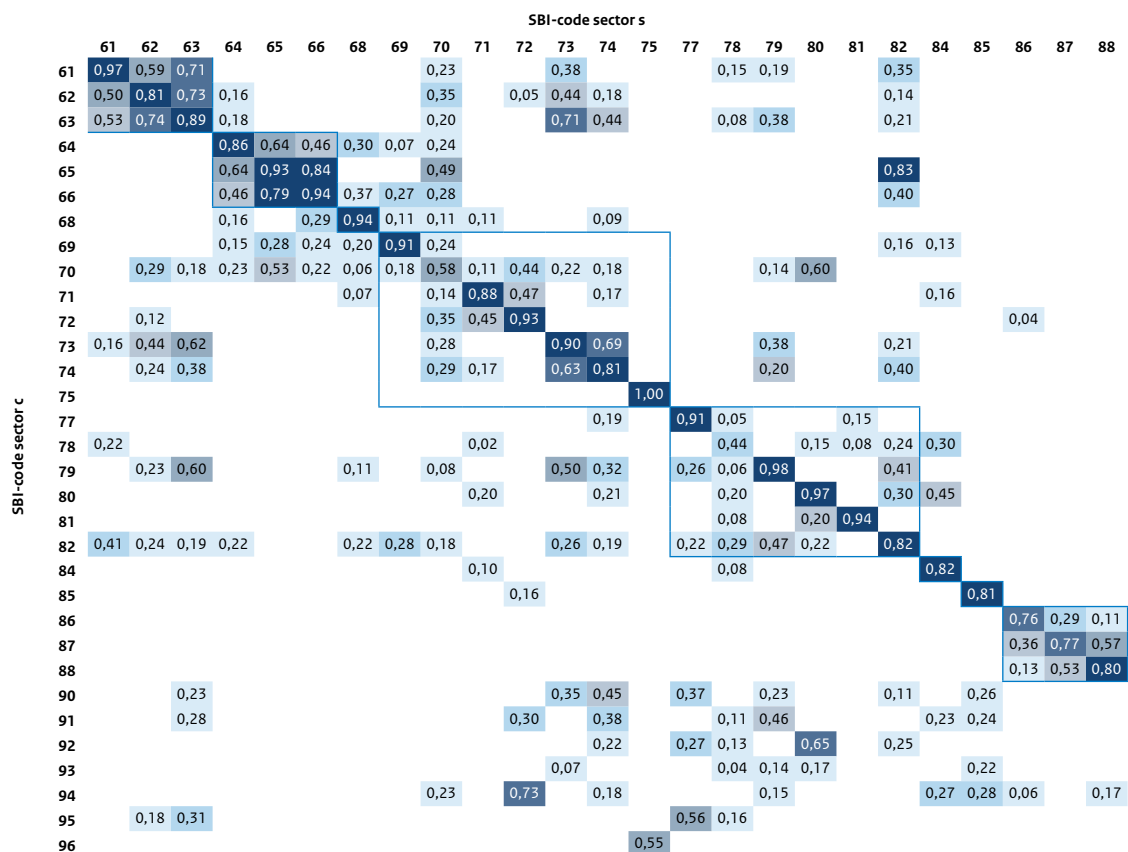


Tabel B2: Specifieke gewichten op basis van cognitieve gerelateerdheid

42







	SBI-code sector s						
	90	91	92	93	94	95	96
01							
02		0,95			0,79		
03							
06							
08							
09							
10							
11							
12							
13							
14		0,90					
15							
16							
17							
18							
19							
20							
21							
22							
23							
24							
25							
26							
27							
28							
29							
30							
31						0,58	
32							
33							
35					0,12		
36							
37							
38							
39							
41							
42							
43						0,19	
45							
46						0,22	
47			0,20	0,10		0,45	0,16
49							
50							
51							
52							
53						0,46	
55	0,10	0,26	0,53	0,37			0,28
56	0,16		0,41	0,35			0,04
58	0,70	0,39			0,46		
59	0,77	0,36	0,36	0,17	0,17		
60	0,87				0,60		
61			0,36				
62						0,17	
63	0,37				0,13		
64							
65							
66							
68							
69							
70			0,21		0,26		
71							
72					0,47		
73	0,36		0,20		0,11		
74	0,42	0,48	0,23	0,09	0,27	0,25	
75							0,31
77	0,41			0,13			
78		0,13					
79	0,23	0,27	0,41	0,30	0,14		
80		0,59					
81							
82	0,34				0,16	0,31	
84		0,36		0,06	0,14		
85	0,26	0,23			0,37		
86					0,16		
87							
88					0,23		
90	0,93	0,82		0,47	0,53		
91	0,74	0,96		0,36	0,70		
92			0,99	0,37			
93	0,35		0,49	0,91	0,21		0,20
94	0,45	0,71		0,34	0,85		
95						0,98	
96			0,35	0,15		0,27	0,95

Tabel B3: Totalen en gemiddelden van relevante variabelen voor eerste jaar van de observatieperiode (2014)

	2014
Aantal ondernemingen	173.687
Toegevoegde waarde, mln. euro	313.837
<i>per bedrijf, x1000 euro</i>	1.806
Werkzame personen	3.675.431
<i>per bedrijf</i>	21,2
Arbeidsproductiviteit, x1000 euro	85,4
Afschrijvingen, mln. euro	39.846
<i>per bedrijf, x1000 euro</i>	229,4
S&O-investeringen, mln. euro	5.385
<i>per bedrijf, x1000 euro</i>	31,0
S&O-kennisvoorraad, mln. euro	35.047
<i>per bedrijf, x1000 euro</i>	201.781

Bron: CBS Microdata (2025). Noot: Alle financiële waarden zijn gecorrigeerd voor inflatie en uitgedrukt in contante waarde tegen prijzen van 2022. Arbeidsproductiviteit is berekend per werkzaam persoon.

Dit is een uitgave van
Ministerie van Economische Zaken

Maart 2026