

De verklaarde variantie verklaard: Een vergelijking van sociologische onderzoeksartikelen in de periode 1975-1998

Matthijs Kalmijn en Gerbert Kraaykamp

Summary

Explaining the R-square: A comparison of sociological research articles 1975-1998

The R^2 is often seen as one indicator of how well researchers are able to explain the phenomena they study. The R^2 is usually not high but there are large differences among studies. In this study, we try to explain these differences by examining the influence of four factors on the value of the R^2 : the field researchers are working in, the type of models and data they use, and the characteristics of the researchers themselves. To examine these influences empirically, we drew a sample of 285 regression models in 119 research articles, published in major journals in the Netherlands and the United States between 1975 and 1998. Multilevel analyses indicate that there are large differences among fields of study: stratification research and labor market studies have the highest R^2 , studies on norms and values have the lowest R^2 . The type of samples researchers work with appear to have very little impact on the R^2 , but the type of models they estimate do matter, although not as much as is often believed. Finally, we do not find that the R^2 has increased over the past decades, nor do we find that American researchers are doing better than Dutch researchers. Most surprisingly, we find that full professors tend to have a lower R^2 on average than other researchers. Our own R^2 is 50 percent, indicating that we are able to explain about half of the differences in the R^2 we observe among studies.

1. Inleiding

Hoe groot is de verklaringskracht van kwantitatief sociologisch onderzoek? Als men een onderzoeksartikel doorbladert, valt het oog vaak als eerste op de R^2 in de onderste rij van de tabellen. In het algemeen wordt de 'proportie verklaarde variantie' opgevat als een aanwijzing voor hoe goed een socioloog het verschijnsel dat hij of zij bestudeert inhoudelijk kan verklaren. Hoe hoger de R^2 , hoe beter de verklaring, zo luidt dan het oordeel. En vaak is dat oordeel kritisch. Waarden rond de twintig procent zijn gebruikelijk, en dat wordt over het algemeen als laag

beschouwd, misschien niet zozeer door sociologen, maar wel door collegawetenschappers uit de meer exacte hoek.

Of deze kritiek terecht is kan men betwisten. Om te beoordelen of twintig procent laag is, moet men nagaan hoe de verklaarde variantie is opgebouwd. De maximaal te bereiken waarde van honderd procent is ruwweg te splitsen in drie delen: een deel dat verklaard wordt, een deel dat men nog kan verklaren, en een deel dat onverklaarbaar is. Het laatste, onverklaarbare deel komt tot stand doordat menselijke gedragingen en sociale structuren een 'random' component kennen. In het leven van mensen spelen toeval, pech en geluk een rol, en veel sociale structuren ontstaan in een lange en grillige reeks opeenvolgende gebeurtenissen. Omdat niet te achterhalen is hoe groot deze random component is, weten sociologen ook niet of de verklaringen die ze geven toereikend zijn.

Niet alleen over de random component bestaat onduidelijkheid, ook over het deel dat men nog kan verklaren is te discussiëren. Verklaart men twintig procent en is de random component vijftig procent, dan nog is de vraag of we de resterende dertig procent wel willen verklaren. Immers, een deel van deze rest zal zijn terug te voeren op psychologische of biologische factoren, en hierin zijn sociologen minder geïnteresseerd. In de sociologie gaat het immers om de verklaring van sociale fenomenen en zijn theorieën over individueel gedrag vooral een hulpmiddel. Binnen het onverklaarde deel is er dus ook een gedeelte dat we niet willen verklaren. Om te beoordelen of sociologen 'het goed doen', moet men dus weten hoe groot de random component is in het te verklaren verschijnsel, en hoe groot de niet-sociale component is, en dat weten we niet.

We kunnen de verklaringkracht van sociologen dus niet goed vergelijken met die van onze critici uit de exacte hoek, maar we kunnen sociologen wel onderling vergelijken. Als we dit doen, valt op dat er tussen studies veel variatie is in de hoogte van de R^2 . Waar komen deze verschillen vandaan, dat is de onderzoeksvraag in ons artikel. Om deze vraag te beantwoorden, kijken we naar de invloed van vier soorten factoren op de hoogte van de verklaarde variantie in sociologisch onderzoek: kenmerken van de onderzoekers, kenmerken van de gegevens, kenmerken van de gehanteerde modellen en kenmerken van het bestudeerde onderzoeksterrein. Met bovengenoemde vraag richten we ons op het deel van de variantie dat we wel willen, maar vaak nog niet kunnen verklaren. Het belang daarvan is dat als we weten welke factoren de hoogte van de R^2 beïnvloeden, we ook zouden kunnen aangeven op welke wijze verklaringen in sociologisch onderzoek zijn te verbeteren. Onze meta-analyse geeft tevens een beeld van de sociologische onderzoekspraktijk in Nederland in de afgelopen decennia: welke soorten modellen en variabelen worden gebruikt, welke onderwerpen worden bestudeerd, wat voor soort gegevens hanteert men, en hoe verhoudt de Nederlandse praktijk zich tot hoogstaand internationaal onderzoek?

Hoewel de R^2 een belangrijke maat is in onderzoek, zegt zij niet alles over de kracht van verklaringen. De R^2 meet in hoeverre verschillen tussen individuen of andersoortige eenheden in het te verklaren verschijnsel kunnen worden toegeschreven aan individuele verschillen in andere, achterliggende kenmerken. De R^2 kan dus ook hoog zijn zonder dat de statistische verklaring theoretisch interessant is. Onderzoekers die modellen maken om toekomstig gedrag te

voorspellen, bijvoorbeeld, zullen altijd belang hebben bij een hoge R^2 , maar niet per se bij een goede inhoudelijke verklaring. Dergelijke mechanische wijzen van verklaren zullen in het wetenschappelijke onderzoek dat wij bestuderen weinig voorkomen. Ook is verklaren in de zin van variantie verklaren niet de enige vorm van verklaren die men in de sociologie tegenkomt. Een andere vorm is het uiteenleggen van verbanden of effecten in indirecte en directe effecten via stelsels van regressievergelijkingen. Hier richt men zich op het statistisch verklaren van invloeden in plaats van op het statistisch verklaren van individuele verschillen. Weer een andere manier van verklaren ligt in de theoretische modelbouw. In deze benadering probeert men via formele modellen en simulaties inzicht te krijgen in de achtergronden van individuele verschillen of verbanden tussen variabelen. In deze bijdrage beperken we ons tot de eerste vorm: het statistisch verklaren van verschillen tussen individuen of andere eenheden.

2. Opzet van het onderzoek

We vergelijken R^2 -en uit een groot aantal gepubliceerde studies en relateren deze via een multivariate regressieanalyse aan een reeks kenmerken van deze studies. Hiervoor is in de eerste plaats een selectie gemaakt van drie gerenommeerde Nederlandse sociologische tijdschriften die geregeld kwantitatief onderzoek publiceren, te weten *Mens & Maatschappij*, *Sociale Wetenschappen* en de *Sociologische Gids*. Om het Nederlands sociologisch onderzoek goed te kunnen beoordelen, maken we een vergelijking met internationaal toponderzoek. Daartoe zijn drie gerenommeerde Amerikaanse tijdschriften geselecteerd: *American Journal of Sociology*, *American Sociological Review*, en *Social Forces*. Om ook een historische context te bieden, hebben we tijdschriften bestudeerd over de periode 1975-1998.

Uit deze tijdschriften is een steekproef van artikelen getrokken. In Nederland is uit elk tijdschrift vanaf 1975 elke twee jaar een jaargang bekeken, en zijn uit elke jaargang de eerste drie artikelen geselecteerd waarin een R^2 voorkwam (13 jaargangen, waaronder ook de meest recente jaargang, 1998). Omdat in Amerikaanse tijdschriften frequenter kwantitatief onderzoek wordt gepubliceerd, is in de Verenigde Staten om de vier jaar een jaargang bekeken (zeven jaargangen, inclusief 1998). De volledige referenties naar de door ons geselecteerde artikelen staan vermeld in de bijlage. We hebben ons beperkt tot onderzoeksartikelen met regressiemodellen voor als continu beschouwde afhankelijke variabelen omdat alleen dan onderlinge vergelijking goed mogelijk is. Bij modellen voor discrete afhankelijke variabelen worden immers andere – en onderling vaak verschillende – maten gebruikt om de verklaringskracht te bepalen. Discrete afhankelijke variabelen vallen dus buiten het bestek van deze studie.

Omdat in een artikel vaak meerdere modellen voorkomen, is een selectie gemaakt van (maximaal) drie modellen per artikel. Als er voor een bepaalde afhankelijke variabele meerdere modellen zijn gepresenteerd, is het meest volledige model gekozen, dat wil zeggen, het model met het grootste aantal onafhankelijke variabelen. Als een gegeven model voor subgroepen is geschat, bijvoorbeeld voor mannen en vrouwen afzonderlijk, zijn beide modellen in de steekproef opgenomen. Het komt voor dat er zowel meerdere afhankelijke variabelen zijn

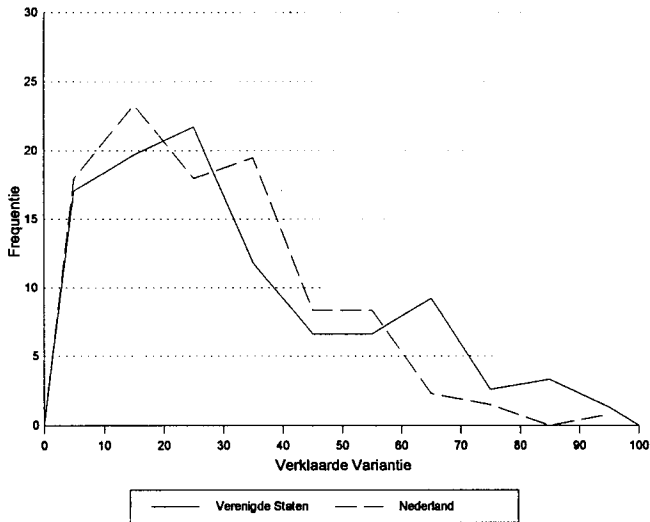
en elk model ook voor subgroepen is geschat. In dat geval is ervoor gekozen om binnen een artikel variatie te krijgen in de afhankelijke variabelen. Voor elke afhankelijke variabele is dan steeds het (meest volledige) model voor de eerste subgroep gebruikt. In totaal zijn er uit Nederlandse tijdschriften 56 artikelen met 133 R^2 -en geselecteerd. Onze Amerikaanse steekproef bevat 63 artikelen en 152 R^2 -en. De Amerikaanse modellen zijn gemiddeld wat 'ouder' dan de Nederlandse modellen (zie Tabel 1), hetgeen erop wijst dat regressie-analyse hier later voet aan de grond kreeg.

3. Onderzochte variabelen en beschrijving

We analyseren steeds de 'ongecorrigeerde' R^2 . Waar nodig is deze uit de 'gecorrigeerde' R^2 en het aantal eenheden en variabelen terugerekend (in 19 procent van de modellen). We nemen overigens aan dat als er in de tabel ' R^2 ' of proportie 'verklaarde variantie' staat, dit ook feitelijk een ongecorrigeerde R^2 is. Dit is een plausibele aanname: onderzoekers hebben er immers weinig belang bij om een – meestal lagere – gecorrigeerde R^2 als een ongecorrigeerde R^2 te presenteren.

De gemiddelde R^2 in Nederland is 27 procent, in de Verenigde Staten significant hoger, 32 procent. Kijken we in Figuur 1 naar de frequentieverdeling van de R^2 -en dan zien we duidelijke verschillen: de verdeling van Nederland ligt links van die van de Verenigde Staten. Verder zien we dat de verdeling in Nederland unimodaal en links-scheef is, met het grootste deel van de R^2 -en tussen de tien en veertig procent. Voor de Verenigde Staten vinden we nog een tweede piek tussen de zestig en zeventig procent; een piek die, zoals we later zullen zien, vooral bestaat uit modellen met aggregaatgegevens. De hoogste R^2 is 94 procent, behaald in een aggregaatanalyse van beroepssegregatie tussen mannen en vrouwen in arbeidsmarktregio's. De laagste R^2 is 2 procent, behaald in een analyse van opvattingen over de rechten van daklozen in een opinieonderzoek.

We denken dat er ruwweg vier groepen van factoren zijn die invloed hebben op de verklarende kracht in sociologisch onderzoek. Ten eerste zal de kwaliteit van de data een rol spelen: wat voor soort steekproef heeft men, welk type data bestudeert men? In de tweede plaats is een onderzoeker met goede data in staat zijn verklaring te verbeteren door meer of minder uitgekende modellen te gebruiken. Ten derde zijn er kwaliteitsverschillen tussen auteurs: de ene socioloog is een betere theoreticus dan de ander, en zal daardoor tot een betere verklaring komen. Tot slot zou het onderwerp een rol kunnen spelen. Sommige verschijnselen hebben een relatief grote 'random' component of een sterk psychologische achtergrond, en zijn daardoor minder goed vanuit een sociologisch perspectief te verklaren. Kortom, de hoogte van de R^2 hangt af van de data die men verzamelt, de modellen die men schat, de mensen die het bedenken, en de vragen die zij stellen. Hieronder bespreken we de operationalisering van deze kenmerken en de verdeling van die kenmerken zoals naar voren komt in een vergelijking tussen Nederlands en Amerikaans onderzoek (Tabel 1).



Figuur 1: Verdeling van R^2 -en in categorieën van tien procent voor Nederland ($N=133$) en de Verenigde Staten ($N=152$)

3.1. Onderzoeksonderwerpen

Om de onderwerpen van de studies te coderen, is steeds gekeken naar de afhankelijke variabele. Omdat het aantal modellen beperkt is, hebben we de afhankelijke variabelen ingedeeld in een klein aantal categorieën. De onderwerpen zijn van tevoren gedefinieerd en de afhankelijke variabelen zijn tijdens het coderen door ieder van ons afzonderlijk in één van de categorieën ingedeeld. In geval van twijfel zijn de afhankelijke variabelen genoteerd en later gezamenlijk aan een categorie toegewezen.

De categorie *stratificatie- en arbeidsmarktstudies* bevat onderzoek naar onder andere status- en beroepsverwerving, inkomensverschillen en uitkomsten in het onderwijs (hierna aangeduid als 'stratificatie'). *Sociaal-psychologische studies* gaan over subjectieve kenmerken van personen, zoals intenties en ambities, wensen en verwachtingen, normen en opvattingen, en persoonlijke gevoelens. Het doet er daarbij niet toe of opvattingen bijvoorbeeld betrekking hebben op de arbeidsmarkt of ongelijkheid: meningen over de rechtvaardigheid van inkomensverschillen horen binnen dit thema en niet binnen het thema stratificatie. *Politiek-historische studies* richten zich op onderwerpen, als modernisering, opkomst bij verkiezingen en democratisering (hierna aangeduid als 'politiek'). *Demografische en criminologische studies* gaan over huwelijks- en gezinspatronen, gezondheidsverschillen, en criminaliteit (hierna aangeduid als 'demografie'). Studies naar bedrijven, overheidsinstellingen, netwerken, en urbanisatieprocessen zijn tot slot ingedeeld in de categorie *organisatie- en stadsstudies* (hierna aangeduid als 'organisatie').

In Tabel 1 laat de verdeling van modellen naar onderwerp zien dat stratificatie- en arbeidsmarktstudies en sociaal-psychologische studies dominant zijn; samen maken zij een kleine 70

Tabel 1: Percentages en gemiddelden voor kenmerken van sociologische tijdschriftartikelen in Nederland en de Verenigde Staten in de periode 1975-1998.

| | Nederland | Verenigde Staten | Toets op verschil ^a |
|--|-----------|------------------|--------------------------------|
| Ongecorrigeerde R-kwadraat (in procenten) | 27,3 | 32,0 | 0,06- |
| Onderzoeksonderwerpen (in procenten) | | | |
| Stratificatie- en arbeidsmarktstudies | 41 | 30 | 0,05* |
| Sociaal-psychologische studies | 38 | 28 | 0,07- |
| Politiek-historische studies | 5 | 6 | 0,81 |
| Demografische en criminologische studies | 9 | 22 | 0,00** |
| Organisatie- en stadsstudies | 7 | 14 | 0,04* |
| Kenmerken van onderzoekers ^b | | | |
| Gemiddeld jaar van verschijnen (1975-1998) | 1991 | 1987 | 0,00** |
| Niet-hoogleraren alleen (%) | 75 | 43 | 0,00** |
| Hoogleraren en niet-hoogleraren gezamenlijk (%) | 20 | 35 | 0,06- |
| Hoogleraren alleen (%) | 5 | 14 | 0,11 |
| Toegepast onderzoek (niet-universitaire auteurs) (%) | 20 | 8 | 0,06- |
| Kenmerken gegevens | | | |
| Gemiddeld aantal eenheden (39-22000) | 1811 | 1844 | 0,94 |
| Gemiddeld aantal eenheden (studies op individueel niveau) | 2048 | 2419 | 0,47 |
| Lokale steekproef (%) | 13 | 35 | 0,00** |
| Sociaal-demografische subset (%) | 54 | 63 | 0,12 |
| Aggregaatgegevens (%) | 12 | 32 | 0,00** |
| Kenmerken van modellen | | | |
| Gemiddeld aantal X-variabelen (1-40) | 8 | 11 | 0,00** |
| Gemiddeld aantal X-variabelen (studies individueel niveau) | 9 | 12 | 0,00** |
| Sociaal-psychologische X-variabele(n) (%) | 30 | 34 | 0,53 |
| Niet-lineaire X-variabele(n) (%) | 27 | 41 | 0,02* |
| Meervoudige meting van X-variabele(n) (%) | 38 | 47 | 0,13 |
| Eerdere meting van Y als X-variabele (%) | 5 | 11 | 0,10- |
| Aantal modellen | 133 | 152 | |
| Aantal artikelen | 56 | 63 | |

^a *t*-toets voor continue variabelen, χ^2 -toets voor discrete variabelen, ** = $p < 0,01$, * = $p < 0,05$, ~ = $p < 0,10$.

^b Berekend met artikelen als eenheden.

Bron: DSO7598.

procent van de modellen uit. In Nederland is de overheersende positie van deze thema's groter, hetgeen wijst op een geringere diversiteit in het Nederlandse kwantitatieve onderzoek. Daarnaast valt op dat demografische studies in de Verenigde Staten relatief meer voorkomen dan in Nederland. Bij deze cijfers dienen we in ogenschouw te nemen dat onze steekproef betrekking heeft op modellen met continue variabelen. Sommige onderzoeksonderwerpen hebben relatief veel discrete kenmerken en zullen in onze studie zijn ondervertegenwoordigd.

3.2. Kenmerken van gegevens

Om kenmerken van de gegevens te meten, is gebruik gemaakt van drie variabelen.

Lokale steekproef. We maken een onderscheid tussen landelijke en lokale steekproeven met als verwachting dat de kwaliteit van lokale steekproeven over het algemeen wat minder is dan die van landelijke steekproeven. In Nederland gebruikt 13 procent een lokale steekproef, significant minder vaak dan in de Verenigde Staten (35 procent). Een verklaring hiervoor is dat het daar duurder is om een landelijke steekproef te organiseren.

Sociaal-demografische subset. Hierbij gaat het om selecties op basis van kenmerken zoals leeftijd, geslacht en etniciteit. We verwachten dat er in dergelijke selecties minder variatie is in de onafhankelijke variabelen waardoor de R^2 lager zal zijn. Daar komt bij dat in een subset ook de variabele ontbreekt waarmee de subset is gemaakt, wat ook de verklaarde variantie verlaagt. Sociologisch onderzoek maakt frequent gebruik van dergelijke subsets, in ongeveer 60 procent van de gevallen.

Aggregaatgegevens. We onderscheiden gegevens over personen of huishoudens enerzijds, en gegevens over aggregaateenheden, zoals steden, landen, regio's, tijdstippen of organisaties anderzijds. Omdat men in een aggregaatopzet allerlei individuele verschillen uitmiddelt, is het gemakkelijker om met sociologische variabelen een goede verklaring te realiseren. Aggregaatanalyses zijn in Nederland minder populair dan in de Verenigde Staten (12 versus 32 procent). Wellicht is de Amerikaanse sociologie meer geïnteresseerd in macroproblemen dan de Nederlandse sociologie.

3.3. Kenmerken van modellen

Om in kaart te brengen hoe de modellen eruit zien, maken we gebruik van vijf variabelen.

Aantal onafhankelijke variabelen per eenheid. Ongetwijfeld de belangrijkste factor in de (ongecorrigeerde) R^2 is het aantal variabelen dat men in een model opneemt. Omdat een gegeven aantal onafhankelijke variabelen meer verklaringskracht geeft bij een klein aantal eenheden, hebben we het aantal variabelen gedeeld door het aantal eenheden. Het gemiddeld aantal onafhankelijke variabelen in Nederland is acht, kleiner dan in de Verenigde Staten (11). Het gemiddeld aantal eenheden is in Nederland niet kleiner, 1811 versus 1844. Als we alleen naar individuele analyses kijken, hebben de Amerikanen wel een voorsprong (400 eenheden), maar dit verschil is niet significant.

Sociaal-psychologische variabelen. Hierbij gaat het om subjectieve kenmerken van personen, zoals intenties en ambities, wensen en verwachtingen, normen en opvattingen, en persoonlijke gevoelens. Omdat dergelijke kenmerken vaak dicht tegen het te verklaren verschijnsel aanliggen, verwachten we dat hiermee de R^2 hoger wordt. Dit type variabele wordt relatief vaak toegepast, in ongeveer 30 procent van de modellen.

Niet-lineaire variabelen. Voor de onafhankelijke variabelen in de modellen is vastgesteld of er een wiskundige bewerking op is toegepast, zoals bijvoorbeeld een logaritmische functie voor bedrijfsgrootte, een kwadratische term voor leeftijd, of een opdeling van opleidingsniveau in dummyvariabelen. Er wordt vaak gesuggereerd dat sociale verschijnselen een niet-lineaire vorm aannemen, zodat de R^2 na dergelijke transformaties hoger zal zijn. Transformaties komen veel voor, maar duidelijk minder in Nederland dan in de Verenigde Staten (27 versus 41 procent).

Meervoudige metingen. Hierbij is nagegaan of onafhankelijke kenmerken zijn gemeten met meerdere indicatoren. Voorbeelden zijn het meten van modernisering via een reeks van landkenmerken of het combineren van attitude-items in een schaal voor waardenoriëntaties. De algemeen geldende wijsheid is dat hiermee random meetfouten worden geminimaliseerd waardoor de systematische component relatief groter wordt en er meer kan worden verklaard. Ook hier zien we dat Nederlandse onderzoekers iets achterlopen. Terwijl in de helft van de Amerikaanse modellen meervoudige metingen worden toegepast, blijft dit in Nederland beperkt tot 38 procent.

Eerdere metingen van de afhankelijke variabelen. Min of meer ter controle is nagegaan of er in de modellen onafhankelijke variabelen zijn opgenomen die metingen zijn van de afhankelijke variabele op een eerder tijdstip (het bruto nationaal product van landen in 1980 als verklaring van het bruto nationaal product in 1990, of het prestige van iemands eerste beroep als voorspeller van het prestige van zijn huidige beroep). Dit komt in Nederland nauwelijks voor (5 procent), in Amerika vaker (11 procent).

3.4. Kenmerken van onderzoekers

Omdat het niet goed mogelijk is om voor alle jaren en beide landen de kwaliteit van wetenschappers op een vergelijkbare wijze te meten, is gebruik gemaakt van vier eenvoudig vast te stellen kenmerken die op zijn best een samenhang zouden moeten vertonen met de kwaliteit van de auteurs.

Land van publicatie. Gecodeerd is of het artikel verscheen in een Amerikaans of in een Nederlands tijdschrift. De selectie in Amerikaanse tijdschriften is strenger en door het grotere aantal sociologen in de Verenigde Staten zal de kwaliteit van de Amerikaanse top hoger zijn.

Jaar van verschijnen. Dit kenmerk is opgenomen in de verwachting dat wetenschappers op elkaars werk voortbouwen en dat cumulatie van kennis leidt tot een hogere R^2 .

Senioriteit van auteurs. Een hoogleraar wordt door de bank genomen meer kwaliteit toegedacht dan anderen. Men zou daarom van hen hogere percentages verklaarde variantie verwachten. Om dit te meten maken we een onderscheid in drie groepen: artikelen geschreven

door één of meer hoogleraren, artikelen geschreven door hoogleraren in samenwerking met niet-hoogleraren, en artikelen geschreven door één of meer niet-hoogleraren. In de Verenigde Staten zijn alleen 'full professors' als hoogleraar geteld. Voor artikelen in *Social Forces* en voor een deel van de artikelen in de *American Journal of Sociology* kon de status van de auteur niet uit het tijdschrift zelf worden gehaald en is gebruik gemaakt van de *Guide to Graduate Departments*, waaruit de status van de meeste auteurs kon worden afgeleid. In Nederland zijn bij 25 procent van de artikelen hoogleraren betrokken, in Amerika meer, 49 procent. Daarnaast valt op dat als hoogleraren publiceren, ze dat in Nederland vaker met niet-hoogleraren doen dan in de Verenigde Staten (in Nederland 5 keer zo vaak, in de Verenigde Staten 2,5 keer zo vaak).

Toegepast onderzoek. Voor zover mogelijk hebben we tot slot gemeten of de onderzoekers een wetenschappelijke of een meer toegepaste benadering hebben. Dit is vastgesteld door na te gaan of één of meer van de auteurs werkzaam was bij een niet-universitair onderzoeksinstituut. Hier gaat het minder om een samenhang met kwaliteit, en meer om het feit dat toegepast onderzoek dikwijls gericht is op beschrijvingsvragen en minder op verklaringsvragen. Zowel in Nederland als de Verenigde Staten heeft het toegepaste onderzoek niet veel airplay in de gerenommeerde tijdschriften (21 en 10 procent).

4. Uitkomsten uit de regressieanalyses

Om na te gaan welke invloed bovengenoemde factoren hebben op verschillen in de R^2 , voeren we een multivariate regressie uit. Omdat er per artikel meerdere R^2 -en zijn, is er geen sprake van onafhankelijke waarnemingen. De 'error'-term van een model in een artikel zal gecorreleerd zijn met de 'error'-term van een ander model uit hetzelfde artikel. Hiervoor corrigeren we door coëfficiënten te schatten via 'random effects' regressiemodellen (in STATA). Dit is een vorm van multilevelanalyse waarbij de afhankelijkheid tussen modellen binnen artikelen wordt meegenomen en zodoende het probleem van een te groot aantal vrijheidsgraden in gewone regressie wordt opgelost.

We schatten het model voor Nederlandse en Amerikaanse modellen samen en nemen een variabele op die aangeeft in welk land het artikel werd gepubliceerd. Het totale aantal modellen is 285, gepubliceerd in 119 artikelen. Er worden twee modellen geschat: het eerste model bevat alleen variabelen over het onderwerp van studie, het tweede model voegt de andere kenmerken toe (Tabel 2). Door deze modellen te vergelijken gaan we na hoe eventuele verschillen in de R^2 tussen onderzoeksonderwerpen totstandkomen.

Het eerste model in Tabel 2 laat zien dat er substantiële verschillen zijn in de R^2 , afhankelijk van het onderwerp dat men analyseert. De hoogste verklaarde variantie vindt men in organisatiestudies, met een gemiddelde van ruim 40 procent. Sociaal-psychologische studies hebben de minste verklaringskracht, gemiddeld bijna 20 procent. Studies over stratificatie, demografie en politiek nemen een middenpositie in.

Als wordt gecontroleerd voor de overige kenmerken van studies en onderzoekers, verandert de volgorde aanzienlijk. Na controle blijken organisatiestudies niet meer de koploper te zijn,

zij doen het dan zelfs iets minder goed dan stratificatiestudies. De in het eerste model geconstateerde voorsprong van organisatiestudies op stratificatiestudies is te wijten aan het feit dat zij relatief vaker gebruik maken van aggregaateenheden. Verder blijken politieke en demografische studies nu een significant kleinere verklaringskracht te hebben dan stratificatiestudies.

Als we kijken naar de tweede set kenmerken, kenmerken van onderzoekers, dan blijkt de kleine voorsprong van Amerikaanse studies er na controle niet meer te zijn. Dit komt niet overeen met wat men zou verwachten. Over de tijd blijkt er ook geen sprake te zijn van een toename. In tegenstelling tot wat we op grond van het idee van wetenschappelijke cumulatie zouden verwachten, bereikt meer recent onderzoek geen significant hogere R^2 . Wel zien we een negatief effect voor toegepast onderzoek. Artikelen waarbij niet-universitaire auteurs betrokken zijn, hebben gemiddeld een 5 procent lagere R^2 .

Een onverwacht resultaat is dat artikelen waarbij hoogleraren zijn betrokken niet een hogere R^2 hebben, eerder een lagere. De twee dummyvariabelen voor de hoogleraar hebben beide een negatief effect, hoewel slechts één daarvan significant is. Als we deze twee aspecten samen nemen, blijkt het totale hoogleraars-effect negatief te zijn en bijna significant ($b = -4,3$, $p = 0,06$). Omdat het onderscheid tussen hoogleraren en niet-hoogleraren in Nederland scherper is dan in de Verenigde Staten, zijn we tevens nagegaan of er een interactie is tussen hoogleraarsvariabele en land. Hieruit blijkt dat het effect in Nederland sterker negatief is, maar het interactie-effect is niet significant ($p = 0,15$).

Deze resultaten spreken onze hypothese over meer kwaliteit en ervaring bij hoogleraren tegen. Een mogelijke interpretatie ligt in het bekende Mattheuseffect: misschien worden artikelen van hoogleraren minder kritisch geëvalueerd en daardoor met een minder optimale verklaring geaccepteerd. Opvallend is dat als hoogleraren samen met junioronderzoekers publiceren de R^2 weer lager is dan de R^2 van publicerende hoogleraren samen. Waarom de samenwerking tussen junior- en senioronderzoekers negatiever uitpakt weten we niet. De verklaring ligt niet in een mogelijk negatief effect van samenwerking zelf, want het verschil blijft bestaan als we solo-auteurs uit onze analyses verwijderen.

Bij de effecten van het type gegevens is verreweg het belangrijkste effect het verschil tussen individuele en aggregaatgegevens. Als gebruik gemaakt wordt van aggregaatgegevens is de R^2 gemiddeld 26 procentpunten hoger. Onze interpretatie hiervoor is dat de tweede component in de verklaarde variantie – die van psychologische en individuele verschillen – in aggregaatgegevens voor een groot deel is weggemiddeld. Het deel dat men kan verklaren met sociale factoren is daardoor groter geworden en het is dus gemakkelijker om hier een hoge R^2 te krijgen. Dat er in de Verenigde Staten vaker van macroanalyses gebruik wordt gemaakt, is een van de redenen dat de voorsprong van Amerikaanse op Nederlandse onderzoekers na controle verdwijnt. De overige variabelen in dit rijtje bieden een beperkte ondersteuning voor de gedachte dat het soort data wat uitmaakt. In overeenstemming met onze verwachting lijken verklaringen wat minder goed te lukken in lokale steekproeven en sociaal-demografische subsamples, maar deze effecten zijn klein (1 procent) en kennen grote standaardfouten.

Het meest voor de hand liggende effect bij het soort model dat men gebruikt is dat van het aantal onafhankelijke variabelen, gerelateerd aan het aantal eenheden. Het effect bedraagt 0,64,

Tabel 2: Random effects regressie-analyse van percentage verklaarde variantie op onderzoeksonderwerpen, kenmerken van onderzoekers, kenmerken van gegevens en kenmerken van modellen: sociologische tijdschrift-artikelen in Nederland en de Verenigde Staten in de periode 1975-1998.

| | Model A | Model B |
|--|---------|---------|
| Constante | 35,4** | 29,2** |
| Onderzoeksonderwerpen | | |
| Stratificatie- en arbeidsmarktstudies (referentie) | - | - |
| Sociaal-psychologische studies | -17,0** | -14,6** |
| Politiek-historische studies | 4,0 | -11,4* |
| Demografische en criminologische studies | 0,1 | -9,8** |
| Organisatie- en stadsstudies | 5,9 | -5,8- |
| Kenmerken van onderzoekers | | |
| Nederlands tijdschrift | | 1,0 |
| Jaar van verschijnen (1998=0, decennia) | | 0,5 |
| Niet-hoogleraren alleen (referentie) | | - |
| Hoogleraren en niet-hoogleraren gezamenlijk | | -5,1* |
| Hoogleraren alleen | | -1,3 |
| Toegepast onderzoek | | -4,9- |
| Kenmerken van gegevens | | |
| Lokale steekproef | | -1,0 |
| Sociaal-demografische subset | | -1,0 |
| Aggregaatgegevens | | 26,0** |
| Kenmerken van modellen | | |
| Ratio X-variabelen / eenheden | | 0,64** |
| Sociaal-psychologische X-variabele(n) | | 6,0* |
| Niet-lineaire X-variabele(n) | | 3,8- |
| Meervoudige meting van X-variabele(n) | | -1,1 |
| Eerdere meting van Y als X-variabele | | 18,3** |
| Modelgegevens | | |
| Aantal modellen | 285 | 285 |
| Aantal artikelen | 119 | 119 |
| R-kwadraat overall (%) | 20,4 | 49,6 |
| R-kwadraat tussen artikelen (%) | 27,2 | 60,3 |
| R-kwadraat tussen modellen (%) | 2,4 | 15,2 |

** = $p < 0,01$, * = $p < 0,05$, - = $p < 0,10$ (eenzijdig getoetst).

Bron: DSO7598.

wat erop wijst dat men zo'n 6 procent meer verklaart met elke tien nieuwe onafhankelijke variabelen, constant gehouden op het aantal eenheden. Evenmin opvallend is dat modellen die eerdere metingen van de afhankelijke variabele opnemen het beter doen, gemiddeld zo'n 18 procentpunten meer verklaarde variantie. Met de opname van een dergelijk effect hebben de meeste auteurs overigens niet de intentie om de verklaring theoretisch te verbeteren, maar meer om effecten van andere variabelen zuiverder te schatten.

Van de andere drie modelkenmerken, heeft het gebruiken van sociaal-psychologische kenmerken het duidelijkst een effect: dit leidt gemiddeld tot 6 procentpunten meer verklaarde variantie. De reden daarvoor is, naar wij menen, dat dergelijke variabelen inhoudelijk erg lijken op de afhankelijke variabelen. Men verklaart examencijfers bijvoorbeeld uit prestatie-motivatie, of gebruikt een meting van zelfvertrouwen als predictor voor psychische gezondheidsproblemen. Tot slot heeft het transformeren van onafhankelijke variabelen een kleine positieve invloed op de hoogte van de R^2 . Meervoudige metingen van de onafhankelijke variabelen blijken echter niet tot een significant hogere R^2 te leiden. Hetzelfde is het geval als we kijken naar meervoudige metingen voor de afhankelijke variabele; ook deze modellen kennen geen significant hogere R^2 .

5. Conclusie

Voor we de bevindingen samenvatten, noemen we een aantal beperkingen van onze aanpak. In de eerste plaats speelt het probleem van selectie. Dat we nauwelijks effecten vinden van bijvoorbeeld, meervoudige meetinstrumenten of niet-lineaire variabelen, zou daarmee verband kunnen houden. Als men in een bepaald model enkelvoudige meetinstrumenten vervangt door meervoudige instrumenten, gaat de R^2 dikwijls wel wat omhoog, zo is althans de algemene ervaring van onderzoekers. In onze opzet vergelijken we echter modellen met en zonder meervoudige meetinstrumenten die ook in allerlei andere opzichten verschillen, en niet al deze verschillen zijn door ons uitgeschakeld. Wellicht worden methodes als meervoudige meetinstrumenten pas toegepast in gevallen waar de verklaring om andere redenen moeilijker is. In dat geval is er sprake van selectie bias en zullen we de verwachte effecten in een cross-sectioneel design als het onze niet terugvinden. Naast het probleem van selectie, zijn er beperkingen in het soort onderwerp dat we bestuderen. Zo hebben we geen discrete afhankelijke variabelen onderzocht en hebben we geen gespecialiseerde vaktijdschriften beschouwd. Ook moest de kwaliteit van onderzoekers in deze bijdrage indirect worden gemeten. Op deze punten kunnen eventuele replicaties of vervolgstudies verbetering aanbrengen.

Samenvattend kunnen we vaststellen dat de gemiddelde R^2 in Nederlands onderzoek 27 procent is, met een omvangrijke spreiding rond dat gemiddelde. In totaal verklaren we met ons 'random effects' regressiemodel 50 procent van de variantie in de verklaarde variantie. We kunnen meer variantie tussen artikelen verklaren dan variantie tussen modellen (60 en 15 procent), hetgeen ook niet verwonderlijk is, omdat veel van onze onafhankelijke variabelen binnen artikelen niet verschillen.

De belangrijkste verschillen in de R^2 komen voort uit het onderwerp dat men bestudeert. Stratificatie- en organisatiestudies hebben gemiddeld een hogere R^2 dan sociaal-psychologische, demografische en politieke studies. Voor de bevinding dat stratificatie- en arbeidsmarkt-onderzoek aan kop staat, is ook wel een verklaring te geven: in dit type onderzoek zit meestal een correlatie tussen de statusposities van ouders en kinderen of een correlatie tussen opleiding en inkomen, en dat zijn nu eenmaal sterke verbanden. Of dit ook in theoretische zin een betere verklaring is, is natuurlijk de vraag.

Wellicht interessanter is dat studies naar normen en waarden, toch een belangrijk onderwerp binnen de sociologie, de laagste R^2 hebben. Blijkbaar is het moeilijker om te verklaren (of te meten) wat er in het hoofd van mensen omgaat dan het is om te verklaren hoe mensen zich gedragen. Met de geconstateerde verschillen tussen onderwerpen kan overigens weinig worden gezegd over hoe verklaringen verbeterd kunnen worden. Wat het wel aantoonst is dat men de meetlat niet overal even hoog kan leggen: men dient een verklaring te beoordelen in het licht van het onderzoeksgebied waarin men werkt.

Daarnaast zijn verschillen in de R^2 afhankelijk van de vraag of men met aggregaatgegevens werkt, of men eerdere metingen van de afhankelijke variabele gebruikt, en of men gebruik maakt van sociaal-psychologische voorspellers. Op deze drie punten is echter ook niet direct vooruitgang te boeken. De aanbeveling om individuele gegevens te aggregeren of 'lagged' variabelen op te nemen, ligt immers niet voor de hand. Daarnaast kan men zich afvragen of het aantrekkelijk is om psychologische kenmerken in de vergelijking te betrekken, vaak liggen deze kenmerken zo dicht tegen de afhankelijke variabele aan dat de uitkomst empirisch wel krachtiger is, maar theoretisch niet.

Van de andere kenmerken die men als onderzoeker zou kunnen manipuleren, zoals de steekproef of de metingen, blijken er in onze analyses geen sterke effecten te zijn op de verklaarde variantie. Eenvoudig gezegd, luidt de conclusie dan dat het weinig uitmaakt voor de verklaringskracht wat voor gegevens men heeft of wat voor modellen men precies schat. Als men naar vooruitgang in verklaringskracht zoekt, dan zal men deze meer in de theorie moeten zoeken dan in de data.

Databron

Kalmijn, M. & G. Kraaykamp. 1999. *Databank Sociologische Onderzoeksartikelen 1975-1998* [DSO7598] [databestand]. Universiteit Utrecht/Katholieke Universiteit Nijmegen.

Bijlage

Hieronder staan de artikelen die in het DSO7598 zijn opgenomen, gesorteerd op tijdschrift en aangeduid met volumenummer, beginpagina en eindpagina.

American Sociological Review:

40:55-70; 40:174-200; 40:215-228; 44:235-252; 44:253-261; 44:262-279; 48:46-59;
48:60-77; 48:78-90; 52:28-43; 52:83-95; 52:96-112; 56:60-71; 56:72-84; 56:117-123;
60:126-140; 60:299-308; 60:449-468; 63:27-38; 63:250-263; 63:536-553.

American Journal of Sociology:

81:109-128; 81:139-146; 81:324-342; 85:109-134; 85:156-166; 85:167-177; 89:350-372;
89:410-419; 89:612-650; 93:141-165; 93:348-382; 93:383-405; 97:96-113;
97:347-375; 97:760-778; 101:578-610; 101:950-992; 101:1306-1332; 103:112-143;
103:429-460; 103:1673-1712.

Social Forces:

53:381-398; 53:553-562; 53:581-594; 57:887-914; 57:915-930; 57:952-959; 61:755-773;
61:774-796; 61:797-824; 65:695-718; 65:746-766; 65:816-833; 69:693-713;
69:715-732; 69:733-761; 73:811-839; 73:841-872; 73:895-915; 76:883-904;
76:905-935; 76:1087-1115.

Mens & Maatschappij:

50:368-382; 56:42-62; 58:5-27; 58:151-166; 58:325-359; 60:119-141; 60:274-281;
60:325-344; 62:153-175; 62:382-400; 64:5-21; 64:42-65; 64:235-252; 66:5-24;
66:180-199; 66:257-276; 68:133-152; 68:153-174; 68:234-256; 70:120-151;
70:203-242; 70:289-303; 72:48-67; 72:132-149; 72:149-165; 73:4-26; 73:27-46;
73:108-129.

Sociale Wetenschappen:

26:29-52; 26:173-179; 30:179-206; 30:207-217; 30:279-300; 32:101-119; 32:239-252;
32:253-262; 34:1-21; 36:23-43; 36:25-51; 38:18-27; 40:1-22; 40:36-61; 40:106-126;
41:54-78; 41:70-89.

Sociologische Gids:

26:265-282; 34:322-336; 36:398-412; 38:144-161; 38:162-173; 40:126-139; 40:320-332;
40:482-501; 42:232-252; 42:426-444; 44:453-462.

Noten

1. Matthijs Kalmijn is als universitair hoofddocent verbonden aan de vakgroep sociologie van de Universiteit Utrecht. Gerbert Kraaykamp is als universitair docent verbonden aan de vakgroep sociologie van de Katholieke Universiteit Nijmegen. Corres-

pondentie richten aan Matthijs Kalmijn, Vakgroep Sociologie, Universiteit Utrecht, Heidelberglaan 1, 3584 CS Utrecht, e-mail: M.Kalmijn@fss.uu.nl.