

Meet risicobereidheid met een kwantitatieve methode

Wanneer een consument een financieel product wil aanschaffen, moet de aanbieder een inschatting maken van de risicobereidheid van de consument. Daarmee wordt bijvoorbeeld gewaarborgd dat iemand niet te veel risico's neemt. Er zijn echter verschillende methodes om risicobereidheid in kaart te brengen. We onderzochten er een aantal in de pensioencontext.

IN HET KORT

- Op dit moment worden vaak kwalitatieve methodes gebruikt om risicobereidheid te inventariseren in de pensioencontext.
- Er bestaan verschillende methodes die kwantitatieve criteria gebruiken.
- Deze kwantitatieve methodes leiden tot een verbeterde inventarisering van risicobereidheid.

GUST VAN DER MEEREN

Voormalig stagiair bij de Autoriteit Financiële Markten (AFM)

HANNIE DE CLOE-VOS

Onderzoeker bij de AFM

ALEXANDRA VAN GEEN

Onderzoeker bij de AFM

Van der Meeren (2017) bevat een uitgebreidere versie van het experiment en de literatuurstudie

Pensioenuitvoerders, banken, verzekeraars en beleggingsadviseurs brengen de risicobereidheid van consumenten in kaart om zo passende producten en diensten te kunnen leveren. Die eis staat ook in de Wet financieel toezicht en de Pensioenwet. Financiële ondernemingen doen dat doorgaans door aan de consument vragen te stellen over de risicobereidheid, de financiële positie, de doelstelling en kennis en ervaring. Die informatie wordt gebruikt om een risicoprofiel van de consument op te stellen. De risicobereidheid bepaalt mede het uiteindelijke beleggingsproduct voor de consument, en het is dus belangrijk dat dit accuraat gemeten wordt.

Het is lastig om de risicobereidheid te bepalen, omdat risicobereidheid niet direct te observeren is. De financiële sector gebruikt veelal niet-wetenschappelijk gevalideerde methodes om de risicobereidheid van consumenten te achterhalen. Een voorbeeld hiervan is een vraag die informeert naar slaapproblemen bij negatieve koersdalingen. En dat terwijl er veel wetenschappelijk onderzoek is gedaan naar de risicobereidheidsinventarisatie (Charness et al., 2013; Holt en Laury, 2014), wat een breed scala aan gevalideerde methodes heeft opgeleverd.

Om tot een betere inschatting van de risicobereidheid in de pensioencontext te komen, hebben we gang-

bare wetenschappelijke methodes om risicobereidheid te meten geëvalueerd aan de hand van vijf criteria. Dit heeft tot een selectie geleid van twee gevalideerde methodes die resulteren in een objectieve maat van risicobereidheid. Bij toetsing in het AFM-consumentenpanel blijken de twee kwantitatieve methodes minstens even goed, zo niet beter, te presteren dan de huidige, gangbare methodes.

Selectie van methodes

In de wetenschap wordt er een grote verscheidenheid aan gevalideerde methodes gebruikt om de risicobereidheid te inventariseren. Ten eerste worden vragenlijsten gebruikt om met kwalitatieve vragen de risicobereidheid in kaart te brengen. Ten tweede zijn er ook *dynamische methodes* waarbij mensen tijdens een spelsituatie kunnen kiezen tussen stoppen met een zekere uitkomst of doorspelen met een onzekere uitkomst. Voorbeelden zijn het knappendeballon-scenario (*Balloon Analogue Risk Task*, BART) of het ontploffende-bom-scenario (*Bomb Risk Elicitation Task*, BRET) (Lejuez et al., 2002; Crosetto en Filippin, 2013). Ten slotte laten *kwantitatieve methodes* mensen kiezen tussen verschillende vooruitzichten met een wisselende hoeveelheid aan risico. Daarbij zijn er verschillende varianten en moeten respondenten vaak een aantal keren kiezen tussen twee opties die onder elkaar worden gepresenteerd (Dohmen et al., 2005; 2011; Holt en Laury, 2002) of juist na elkaar waarbij de opties kunnen variëren al naar gelang eerder gemaakte keuzes (Barsky et al., 1997).

Niet al deze methodes zijn toepasbaar bij het inventariseren van de risicobereidheid ten aanzien van financieel risico van consumenten. Daarom hebben we eerst een kader ontwikkeld om de geschiktheid van een methode te beoordelen, het zogenoemde FRAME-kader. Dit omvat vijf criteria waar een succesvolle methode om risicobereidheid te meten in onze ogen ten minste aan moet voldoen: *Feasible, Rationalising, Appropriate, Measurable* en *Errable*.

Feasible (uitvoerbaar)

Een eerste belangrijke vereiste van een methode is dat deze van consumenten een lage belasting vergt. De inventarisatie moet daarom snel en gemakkelijk kunnen worden afgenomen. Dit is zeker het geval als het om risicobereidheid bij pensioen gaat, omdat mensen dit onderwerp vaak als lastig en vervelend ervaren.

Rationalising (rationaliserend)

Een tweede eis is dat de methode irrationele effecten zo veel mogelijk beperkt. Uit de gedragswetenschappelijke literatuur weten we dat mensen moeite hebben met bijvoorbeeld de inschatting en weging van kleine kansen, verliesaversie en *framing*. Wanneer we de ware risicobereidheid van personen willen achterhalen, is het daarom noodzakelijk om dergelijke effecten tijdens de inventarisatie van de risicobereidheid te voorkomen of te verminderen.

In de pensioencontext kan dit onder andere door de uitkomsten te tonen als totaalinkomen (Alserda et al., 2016), bijvoorbeeld door de consument voor te leggen “u krijgt in totaal 2000 euro pensioen”, in plaats van “u krijgt 100 euro meer pensioen”. Ook het gebruiken van gematigde in plaats van extreme kansen – tussen dertig en zeventig procent – beperkt het effect van het irrationeel wegen van kansen (Holt en Laury, 2014). En verliesaversie kan gereduceerd worden door zorgvuldige framing. Meer inzichten in hoe dergelijke effecten het gedrag van beleggers beïnvloeden, heeft de AFM beschreven in een rapport over zelfstandige beleggers (AFM, 2015).

Appropriate (passend)

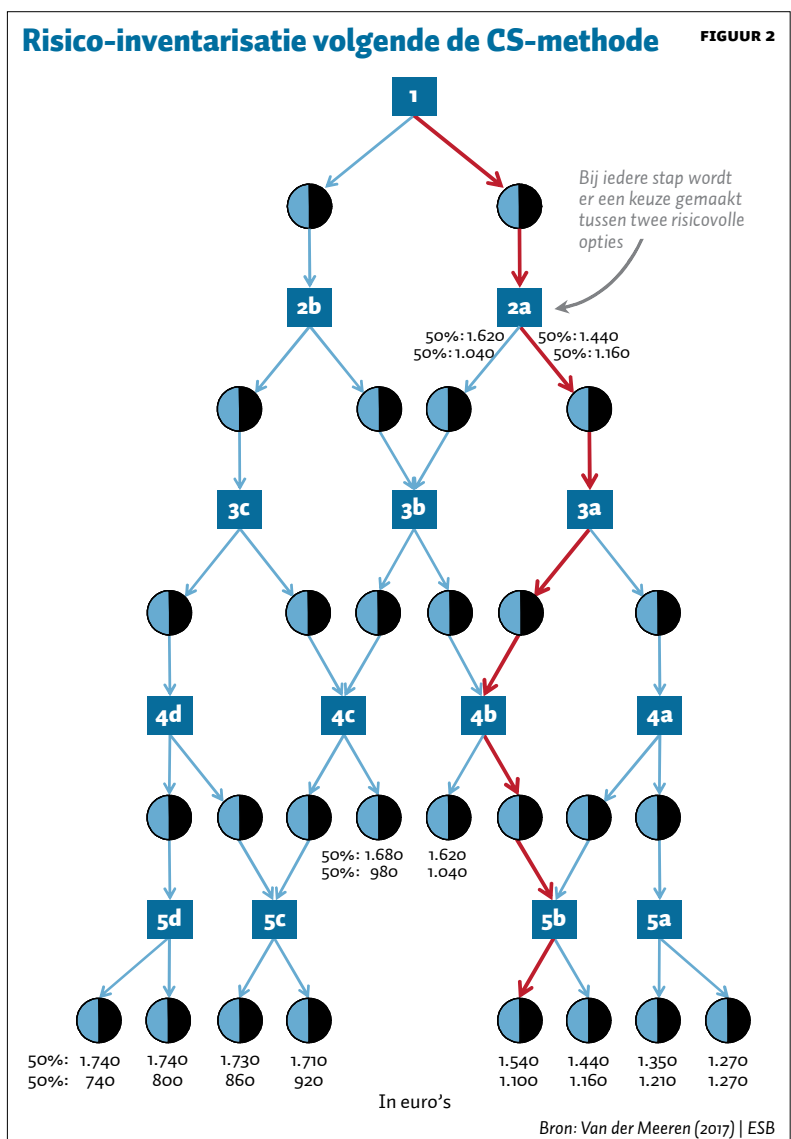
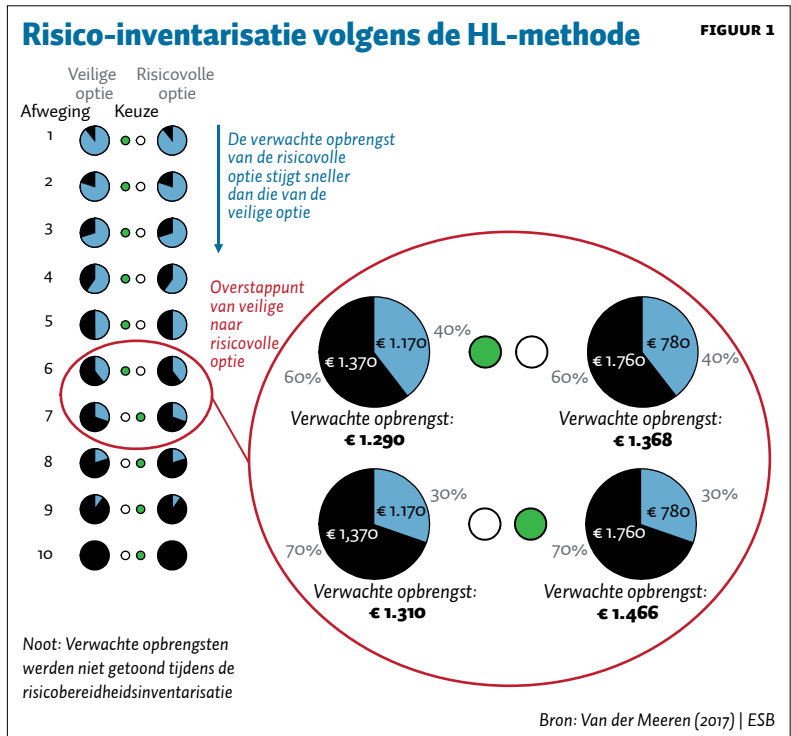
Risicobereidheid is contextafhankelijk en daarom maakt een goede methode gebruik van passende scenario's gegeven de context (Holt en Laury, 2014; Weber et al., 2002). Iemand kan bijvoorbeeld risicoavers zijn wanneer het gaat om financiële zaken, en tegelijkertijd risicozoekend in een sociale of gezondheidscontext. Daarom is het belangrijk dat de risicobereidheidsinventarisatie daaraan wordt aangepast, met hierbij horende bedragen en kansen.

Measurable (meetbaar)

De methode levert een kwantitatieve, objectief interpreteerbare score. Bij kwalitatieve methodes – die bijvoorbeeld informeren naar participatie bij gokspellen – is het wel mogelijk om een *relatieve* positie van personen ten opzichte van elkaar vast te stellen, de een gokt vaker dan de ander, maar is het niet mogelijk om een goede vertaalslag naar beleggingsrisico's te maken. Voor een persoon met de hoogste risicobereidheid binnen een bepaalde groep is namelijk niet noodzakelijkerwijs het meest risicovolle beleggingsproduct passend.

Door gebruik te maken van kwantitatieve methodes is er een directe vertaalslag mogelijk van gemeten risicobereidheid naar beleggingskeuzes, omdat het om een *absolute* maatstaf van risicobereidheid gaat die kan worden gemodelleerd. Het keuzemodel dat hiervoor vaak wordt gebruikt, *Expected Utility Theory* (EUT), ligt tevens ten grondslag aan de afweging tussen rendement en risico in moderne portfoliotheorieën bij beleggingen (Markowitz, 1952).

Uiteindelijk moet de methode een kwantitatieve en bruikbare maat van risicobereidheid opleveren: een score die door zowel de consument als de financiële onderneming eenvoudig te interpreteren is. Dat kan door de consument hypothetische scenario's voor te leggen waarbij er een keuze gemaakt moet worden tussen concrete bedragen en kansen, bijvoorbeeld 2.000 euro met zekerheid, of vijftig procent kans op 4.500 euro en vijftig procent kans op 0 euro.



Methodes voor risicobereidheid-correlaties

TABEL 1

	Holt-Laury (HL)	Choice Sequence (CS)	Eén vraag (EG)
Gem. categorie ± SD	3,60 ± 2,5	3,76 ± 2,6	3,39 ± 2,9
Gem. CRRA ± SD	1,95 ± 3,3	1,79 ± 3,2	2,24 ± 3,1
Aandeel risicoavers	83%	78%	86%
Aantal dat op alle antwoorden consistent antwoordt	321	294	426
Aandeel dat minstens één inconsistent antwoord gaf	25%	31%	-
Correlatie			
met EG-methode	0,61 ***	0,67 ***	-
met CS-methode	0,67 ***	-	0,67 ***
met SOEP-vragenlijst	0,30 ***	0,39 ***	0,27 ***
met DOSPERT-vragenlijst	0,10*	0,11*	0,08*

*/**/** Significant op respectievelijk tien-, vijf- en eenprocentniveau

Bron: Van der Meeren (2017) | ESB

Errable (feilbaar)

Ten slotte staat een goede methode ook inconsistente antwoorden toe. Daardoor worden mensen herkend die de methode niet begrijpen. Het is bekend dat mensen zaken zoals risico en kansen moeilijk kunnen begrijpen, maar een effectieve inventarisatiemethode van risicobereidheid houdt hier rekening mee. De methode haalt de mensen die de vragen niet begrijpen eruit, zodat ze op een andere manier kunnen worden geholpen.

Drie methodes

Op basis van bovenstaande criteria zijn er twee methodes vastgesteld die voldeden aan alle criteria: de HL-methode (Holt en Laury, 2002) en de CS-methode (*Choice Sequence*).

De HL-methode, weergegeven in figuur 1, is een lijst van binaire keuzes tussen twee opties die variëren in risico en verwachte opbrengst. De verwachte opbrengst van de meest risicovolle optie wordt geleidelijk aantrekkelijker. De verwachting is dat een persoon na een aantal stappen overgaat van de 'veilige' optie links naar een meer risicovolle optie rechts. Dat overstappunt, in de figuur met een rode cirkel weergegeven, bepaalt de risicobereidheid. In de figuur is de respondent vanaf de zevende keuze overgestapt op het meer risicovolle alternatief. Het feit dat zij in de zesde keuze nog niet was overgestapt, ondanks dat ook daar het risicovollere alternatief al een hogere verwachte opbrengst had, geeft aan dat zij risicoavers is. Net zoals bij de volgende methodes, kan vervolgens, op basis van de keuze en de bijhorende kansen en uitkomsten, via een EUT-model de risicobereidheidscoëfficiënt van de respondent(e) worden berekend.

De CS-methode, weergegeven in figuur 2, legt aan respondenten een reeks van keuzes tussen twee risicovolle opties voor. De gepresenteerde optie is afhankelijk van de eerder gemaakte keuzes van de respondent, waardoor de risicobereidheid steeds preciezer in kaart gebracht wordt. De figuur toont de verschillende mogelijke 'routes' naar eenzelfde uiteindelijke keuze, zodat foutjes niet catastrofaal zijn en ook inconsistenties kunnen worden vastgesteld.

Voor het experiment is ook nog een derde gevalideerde kwantitatieve methode meegenomen. Deze EG-methode, naar Eckel en Grossman (2002), bestaat uit één vraag, waarbij mensen uit tien verschillende opties kunnen kiezen. Omdat de EG-methode maar één vraag is, voldoet de methode niet aan het feilbaarheids criterium: mensen die de vraag niet snappen, zijn niet te achterhalen op basis van hun antwoord.

Risico-inventarisaties vergeleken

De methodes zijn getest in het AFM Consumentenpanel onder 426 respondenten in het voorjaar van 2017. We hebben de scenario's in een pensioencontext geplaatst en de gepresenteerde risico's gebaseerd op de inkomenscategorie van de respondent. Ook is er een oefenvraag toegevoegd, en zijn de HL-, CS- en EG-methode licht aangepast om ze onderling te kunnen vergelijken en om ook inconsistente antwoorden toe te laten (zodat aan de 'errable'-component van het FRAME-kader wordt voldaan). De risicobereidheidscategorieën hebben in alle methodes gelijke CRRA-grenzen (*Constant Relative Risk Aversion*). Alle drie de methodes zijn aan elke respondent in gerandomiseerde volgorde voorgelegd, zodat de resultaten vergeleken kunnen worden.

Omdat we geen objectieve maatstaf van individuele risicobereidheid hebben, beoordelen we elke methode op basis van hoe goed ze de andere twee methodes benadert (net zoals bijvoorbeeld in Deck et al., 2010; Reynaud en Couture, 2012; Corsetto en Flippin, 2016). Hieronder ligt de aanname dat alle methodes in zekere mate de onderliggende ware risicobereidheid meten. Daarnaast vergelijken we de methodes ook met twee gevalideerde kwalitatieve methodes: de risicobereidheidsvragen uit de SOEP-questionnaire (*German Socio-Economic Panel*) (Dohmen et al., 2005) en de psychometrische DOSPERT-vragenlijst (*Domain Specific Risk-Taking*) in het financiële domein (Blais en Weber, 2006).

Kwantitatieve methodes vergeleken

De resultaten van de vergelijking tussen de methodes zijn weergegeven in tabel 1. Daarbij wordt de respondent bij elke methode in een categorie geplaatst van sterk risicoavers (1) tot sterk risicozoekend (10), met in het midden (5–6) risiconutraal. De respondenten zijn volgens alle methoden redelijk risicoavers: het gemiddelde is 3,60 bij de HL-methode, 3,76 bij de CS-methode en 3,39 bij de EG-methode.

Om de verschillende methoden met elkaar te vergelijken is bovendien gebruik gemaakt van de corresponderende CRRA-risicobereidheidscoëfficiënt. Bij een negatieve CRRA-score is een respondent risicozoekend, en bij een positieve score risicoavers. Ook met deze maatstaf zijn de respondenten gemiddeld genomen risicoavers met CRRA-scores tussen de 1,79 en 2,24. Deze maatstaf is robuust voor verschillende methodes en onderzoeken. Voor onze vergelijking van methodes is de CRRA echter minder van belang, omdat we alle drie de methodes zo hebben ontworpen dat ze allemaal tien categorieën hebben met precies dezelfde CRRA-bandbreedtes.

Alle methodes geven een andere risicobereidheid aan respondenten en deze verschillen zijn statistisch significant. De HL-methode verschilt echter het minst van de andere

twee methodes, en presteert dus het beste. De CS-methode volgt, en de EG-methode presenteert verder het slechtst.

Als we kijken naar de correlatie tussen de drie methodes, dan zien we dat deze voor elke combinatie significant is. De verschillende kwantitatieve testmethodes komen dus grotendeels overeen wat betreft de risicobereidheid van de respondent. Bovendien zijn de gevonden correlaties (0,61–0,67) relatief hoog in vergelijking met de correlatiehoogte in andere wetenschappelijke literatuur. Deze drie kwantitatieve methodes lijken dus duidelijk een en dezelfde onderliggende risicobereidheid te meten.

Om de correlaties te kunnen rangschikken, zodat duidelijk wordt welke methode het beste presteert, nemen we respondenten die inconsistente antwoorden hebben gegeven niet mee in de vergelijking, waardoor we 224 respondenten overhouden. Uit de analyse blijken er geen statistisch significante verschillen te zijn in de correlaties. Alle methodes presteren hier dus even goed.

Vergelijking met kwalitatieve methodes

Vervolgens bepalen we de kwaliteit van de drie kwantitatieve methodes door te kijken of er een correlatie bestaat tussen de kwantitatieve methodes en de twee kwalitatieve questionnaires. De correlaties tussen de SOEP-questionnaire en de drie kwantitatieve methodes zijn zeer significant. Daaruit leiden we af dat deze methodes dezelfde onderliggende risicobereidheid meten. De correlatie tussen de DOSPERT-methode en de drie kwantitatieve methodes is maar net significant op tienprocentniveau (tabel 1). Dat doet vermoeden dat deze methode risicobereidheid minder goed meet.

De hoogte van de correlaties tussen de kwantitatieve methodes en zowel de SOEP-methode als de DOSPERT-methode zijn beide een stuk lager dan de correlatie tussen de kwantitatieve methodes onderling. Dit geeft aan dat zelfs deze verschillende wetenschappelijk gevalideerde methodes een substantieel andere “risicobereidheid” meten. Gezien het grote verschil tussen kwantitatieve en kwalitatieve methodes is dit niet verrassend, maar het onderstreept wel het belang van een bewuste keuze wat betreft de methode voor het inventariseren van financiële risicobereidheid. Opnieuw worden mensen met inconsistente antwoorden niet meegenomen bij de vergelijking tussen de methodes. Statistische testen wijzen uit dat er voor de correlaties met de DOSPERT-methode geen significante verschillen zijn tussen de drie methodes. Op basis van deze tests kunnen we dus niet stellen dat de ene kwantitatieve methode beter zou zijn dan de ander. Voor de SOEP-vragenlijst zijn er wel verschillen: de CS-methode presteert het best, gevolgd door de HL- en de EG-methode.

Inconsistentie

Om te kunnen beoordelen hoe uitvoerbaar of feilbaar de methode is, is het belangrijk dat er niet te veel inconsistente antwoorden gegeven worden – iets dat bij feilbare methodes kan worden bepaald. Bij de HL-methode is 25 procent van de antwoorden inconsistent, en bij de CS-methode is dat 31 procent. Dit zijn maximum-schattingen, omdat sommige inconsistenties een verwaarloosbaar effect zullen hebben op de uiteindelijke risico-classificatie. Ook bleek dat respondenten vaker inconsistente antwoorden gaven naarmate ze

een bepaalde vraag later in de vragenlijst kregen. Dit vermoeidheidseffect speelt waarschijnlijk minder als de kwantitatieve test de enige vragen zijn die een consument moet beantwoorden. Respondenten die laag scoorden op een cognitieve reflectietest die aan het eind van de vragenlijst was toegevoegd, gaven ook vaker inconsistente antwoorden.

Conclusie

De twee geëvalueerde kwantitatieve risicobereidheidsinventarisaties die aan de FRAME-criteria voldoen, lijken een verbetering ten opzichte van de veelgebruikte inventarisaties uit de sector. Wij moedigen de sector dan ook aan om dit in de praktijk te onderzoeken. Een doorontwikkeling door middel van *test-learn-adapt* kan de methodes verder optimaliseren en simplificeren. Dit kan ook leiden tot een vermindering van de inconsistente beantwoording. Zie hiervoor ook de handleiding van het team consumentengedrag bij het expertisecentrum van de AFM (AFM, 2017).

In ieder geval hopen we dat er gebruikgemaakt gaat worden van wetenschappelijk gevalideerde methodes, en dat testen die niet meten wat ze behoren te meten uit het landschap verdwijnen. Ook de wetenschap kan een bijdrage leveren aan de zoektocht naar nauwkeuriger en op de praktijk gerichte methodes om de risicobereidheid van consumenten in kaart te brengen.

Literatuur

- AFM (2015) *Belangrijke inzichten over zelfstandige beleggers*, december. Rapport te vinden op www.afm.nl.
- AFM (2017) *Hoe voer ik een betrouwbaar gedragsexperiment uit? In 7 stappen meer inzicht in hoe consumenten zich gedragen*, mei. Publicatie te vinden op www.afm.nl.
- Alserda, G., B. Dellaert, L. Swinkels en F. van der Lecq (2016) *Pension risk preferences: a personalized elicitation method and its impact on asset allocation*. Netspar Industry Series Design Paper, 62.
- Barsky, R.B., F.T. Juster, M.S. Kimball en M.D. Shapiro (1997) Preference parameters and behavioral heterogeneity: an experimental approach in the health and retirement study. *The Quarterly Journal of Economics*, 112(2), 537–579.
- Blais, A.R. en E.U. Weber (2006) A domain-specific risk-taking (DOSPERT) scale for adult populations. *Judgment and Decision Making*, 1(1), 33–47.
- Charness, G., U. Gneezy en A. Imas (2013) Experimental methods: eliciting risk preferences. *Journal of Economic Behavior & Organization*, 87, 43–51.
- Crosetto, P. en A. Filippin (2013) The ‘bomb’ risk elicitation task. *Journal of Risk and Uncertainty*, 47(1), 31–65.
- Deck, C.A., J. Lee, J. Reyes en C. Rosen (2010) *Measuring risk aversion on multiple tasks: can domain specific risk attitudes explain apparently inconsistent behavior?* Paper te vinden op www.semanticscholar.org.
- Dohmen, T.J., A. Falk, D. Huffman et al. (2005) *Individual risk attitudes: new evidence from a large, representative, experimentally-validated survey*. DIW Discussion Paper, 511.
- Eckel, C.C. en P.J. Grossman (2002) Sex differences and statistical stereotyping in attitudes toward financial risk. *Evolution and Human Behavior*, 23(4), 281–295.
- Holt, C.A. en S.K. Laury (2002) Risk aversion and incentive effects. *The American Economic Review*, 92(5), 1644–1655.
- Holt, C.A., en S.K. Laury (2014) Assessment and estimation of risk preferences. In: M.J. Machina en W.K. Viscusi (red.) *Handbook of the Economics of Risk and Uncertainty*, Volume 1. Oxford: North-Holland, 99–133.
- Lejuez, C.W., J.P. Read, C.W. Kahler et al. (2002) Evaluation of a behavioral measure of risk taking: the Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, 8(2), 75–84.
- Markowitz, H. (1952) Portfolio selection. *The Journal of Finance*, 7(1), 77–91.
- Meeren, A.A. van der (2017) *Comparing risk preference elicitation methods for pensions*. Masterscriptie, Erasmus School of Economics, Erasmus Universiteit Rotterdam.
- Reynaud, A. en S. Couture (2012) Stability of risk preference measures: results from a field experiment on French farmers. *Theory and Decision*, 73(2), 203–221.
- Weber, E.U., A.R. Blais en N.E. Betz (2002) A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making*, 15(4), 263–290.