

De z-score: naar de geest of naar de letter?



Auteur
Hans Braker¹

In de wet voor de z-score wordt voor de rendementsberekeningen verwezen naar “daartoe opgestelde richtlijnen van de Vereniging van Beleggingsanalisten”. Het rekenwerk zit dus meestal wel goed. In de praktijk zijn er echter vaak grote interpretatieverschillen rondom datgene waaraan gerekend moet worden. Afhankelijk van de interpretatie behaalde een pensioenfonds een flinke outperformance dan wel een zware underperformance. Underperformance kan leiden tot een stroom van honderden miljoenen euro’s aan pensioenvermogen van pensioenfondsen naar verzekeraars. De economische impact van een ogenschijnlijk klein interpretatieverschil kan daarom bijzonder groot zijn. De dogma’s die leiden tot de grootste problemen zijn “de wet moet naar de letter worden gevolgd” en “de accountant heeft het resultaat goedgekeurd, dus de z-score is correct”.

Veel bedrijfstakingen kennen een pensioenfonds waar alle ondernemingen in deze bedrijfstak verplicht bij zijn aangesloten. In sommige takken bestaat een vrijwillig bedrijfstakpensioenfonds. De z-score² heeft betrekking op de verplichtgestelde bedrijfstakpensioenfondsen. Sinds 1998 is een vrijstellingsregeling van kracht voor deze fondsen. De regeling is enkele malen gewijzigd en is nu bekend als het Vrijstellings- en boetebesluit Wet Bpf 2000 (zie www.wetten.nl). De wetgever wil in beginsel de verplichtstelling en solidariteit handhaven. Een “Vrijstelling in verband met onvoldoende beleggingsrendement” kan echter aan een onderneming worden verleend wanneer uit een performance-toets, op grond van de jaarlijkse z-scores, blijkt dat het rendement over een periode van vijf jaren significant lager was dan het rendement van de eigen benchmark. De onderneming kan dan, onder voorwaarden, het pensioen van de werknemers elders onderbrengen.

De afgelopen jaren is vaker dan twintig maal een fonds gezakt voor de performancetoets. Na een enkel individueel “incident” in de periode tot en met het jaar 2007 zijn na 2008 vijftien en na 2009 zeven fondsen gezakt. Enkele fondsen zakten meerdere malen. Zie Molenkamp (2011) voor een actuele kritische blik op het instrument z-score. Op het moment van schrijven van dit artikel was nog niet bekend of er fondsen over de periode 2006 tot en met 2010 gezakt zijn. Het is niet algemeen bekend hoeveel ondernemingen zijn overgestapt. Waarschijnlijk betreft dit een handvol bedrijven.

— *De z-score toetst de uitvoering, niet het beleggingsbeleid* —

Met de z-score beoogt men de uitvoering van het beleggingsbeleid te toetsen. Dus niet het beleggingsbeleid zelf. Molenkamp en Zweekhorst (2004) geven voorbeelden van grijze gebieden tussen uitvoering en beleid. Een citaat uit het betreffende artikel: “De wet Bpf is summier met het geven van richtlijnen zodat er ruimte is voor een eigen interpretatie van de normportefeuille zonder dat daarbij de geest van de wet geweld wordt aangedaan. In tegendeel, door de wet niet naar de letter maar naar de geest te volgen ontstaat juist jurisprudentie waarbij het doel van de performancetoets: *het beoordelen van de kwaliteit van de uitvoering van beleggingsbeleid*, beter tot zijn recht komt.”

Eén van de waarborgen die is ingebouwd in de procedures rondom de z-score, is een accountantscontrole. Praktische ervaringen wijzen er echter op dat accountants regelmatig interpretaties naar de letter hanteren die conflicteren met de geest van de wet.

In het verleden zijn diverse technische onvolkomenheden in de berekeningsmethodiek aangegeven, zie Molenkamp en Zweekhorst (2002, 2003, 2004) en Jadnanansing (2001). We behandelen hier niet de details van de berekeningen van de z-score. Ook bij een eventuele andere toetsingsmethodiek die tot doel heeft de kwaliteit van de uitvoering te meten, komen dezelfde problemen namelijk naar

voren. In dit artikel laten we zien dat bij iedere toets op de uitvoering interpretatieverschillen kunnen leiden tot het ten onrechte wel of juist niet zakken. In vrijwel alle gevallen komt het probleem neer op het correct kunnen vaststellen van de normportefeuille.

Het probleem is het correct vaststellen van de normportefeuille

Rebalancing, een oud probleem

Ruim tien jaar na de introductie van de z-score en veel publicaties over dit onderwerp verder, zijn er nog altijd pensioenfondsen die in hun norm een andere rebalancing-strategie gebruiken dan in hun feitelijke portefeuille. Zie bijvoorbeeld Braker (2003) en Zweekhorst (2002). Een eerste voorbeeld is een pensioenfonds dat aan het begin van elke maand meet hoeveel de feitelijke wegingen in de portefeuille afwijken van de strategische wegingen. Wanneer de portefeuillegewingen duidelijk afwijken van de strategische wegingen, wordt de portefeuille bijgestuurd richting de strategische wegingen. Dit betekent dat er in principe maandelijkse rebalancing plaatsvindt, hoewel dit niet expliciet in het beleggingsplan is vermeld. Tegelijkertijd wordt door de externe performancemeter – zonder toelichting in de rapportage – een normportefeuille

gehanteerd die alleen aan het begin van het jaar de strategische wegingen weerspiegelt. Gedurende het jaar wordt er niets met die wegingen gedaan. Er vindt dus geen rebalancing plaats. Het jaarrendement van deze normportefeuille is simpelweg te berekenen door de begingewichten per categorie te combineren met de jaarrendementen van de categorie. Het jaarrendement van een wel bij de strategie passende normportefeuille volgt echter uit twaalf maandelijks berekeningen, waarbij steeds de strategische gewichten met maandrendementen worden gecombineerd. In Tabel 1 staat een voorbeeld van de mogelijke impact van de verschillen in normrendement.

Een tweede voorbeeld is dat van een pensioenfonds met twee vermogensbeheerders. De primaire beheerder, die het merendeel van het vermogen beheert, rapporteert over de gehele portefeuille. De secundaire beheerder heeft alleen een aandelenmandaat. De primaire beheerder is gewend om te werken met kwartaalrebalancing. In de overkoepelende rapportage wordt daarom een normportefeuille gebruikt waarbij ieder kwartaal de wegingen van het begin van het jaar worden hersteld. In de praktijk echter wordt alleen aan het begin van het jaar besloten of er geld wordt toegevoegd dan wel onttrokken aan het aandelenmandaat. De werkelijke weging van aandelen is dus door marktbevingen gedurende het jaar anders dan de weging aan het begin van het jaar. De normportefeuille past daarom niet bij de gevoerde strategie. Een significante fout zoals weergegeven in Tabel 1 kan het resultaat zijn.

Het is zorgelijk dat meerdere pensioenfondsen en hun controlerende accountants jarenlang zulke welbekende problemen niet hebben getraceerd.

Renterisico, een nieuw probleem

Pensioenfondsen hanteren voor de z-score uiteenlopende methodieken voor de afdekking van renterisico via een overlay. Zie Tabel 2 voor verschillende z-scores die voor één en dezelfde situatie kunnen worden bepaald met gebruikmaking van diverse methodieken.

Tabel 1: mogelijke effecten bij gebruik onjuiste rebalancing

	onjuiste rebalancing	juiste rebalancing
rendement portefeuille	10%	10%
rendement normportefeuille	10,6%	9,8%
outperformance	-0,6%	0,2%
z-score	-0,38	0,12

Tabel 2: mogelijke effecten bij verschillende methoden voor rente-overlay in benchmark

	methode 1: passende benchmark	methode 2: geen effect op z-score	"methode" 3a: tussentijdse aanpassing overlay niet in norm	"methode" 3b: onjuiste benchmark overlay
rendement portefeuille exclusief rente-overlay	6%	6%	6%	6%
effect rente-overlay in portefeuille	10%	10%	10%	10%
rendement portefeuille inclusief rente-overlay	16%	16%	16%	16%
rendement normportefeuille exclusief rente-afdekking	7%	7%	7%	7%
effect rente-overlay in normportefeuille	9%	10%	4%	6%
rendement normportefeuille inclusief rente-overlay	16%	17%	11%	13%
outperformance	0%	-1%	5%	3%
z-score	0	-0,63	3,13	1,88

1: Passende benchmark. De eerste methodiek heeft in het algemeen de voorkeur omdat de letter en de geest van de wet worden gevolgd. Deze methodiek wordt door weinig fondsen toegepast. Er wordt voor de rente-overlay zowel qua omvang als samenstelling een benchmark vastgelegd. De effecten van een afwijkende omvang van de rente-afdekking (allocatie-effect) en de specifieke invulling met instrumenten (selectie-effect) kunnen zo worden gemeten. Beide effecten komen tot uiting in de z-score. Er zijn echter complicaties: wanneer gedurende het jaar de strategische mate van rente-afdekking dient te worden aangepast, zijn formele stappen nodig om te voldoen aan het vrijstellingsbesluit. Onder andere moet de nieuwe normportefeuille worden bekendgemaakt.

2: Feitelijke rendementsbijdrage als benchmark. De gedachte bij de tweede methodiek is dat de rente-afdekking een pure balansbeslissing is, zonder relatie met de uitvoering van de beleggingen. Er wordt gedaan alsof de rente-afdekking geen onderdeel van de beleggingsportefeuille is en daarom niet onder de reikwijdte van de z-score valt. Het voordeel is dat de z-score eenvoudig wordt geneutraliseerd voor de rente-afdekking: de rente-afdekking is per definitie beleid en heeft nooit impact op de uitvoering. Ook kunnen de mate van rente-afdekking en de specifieke invulling snel en adequaat worden aangepast tijdens het jaar. De methodiek staat op gespannen voet met zowel de letter als de geest van de wet. Met de letter, onder andere omdat volgens de wet voor iedere categorie een passende benchmark dient te worden gekozen. Hier ontstaat snel een ingewikkelde discussie, omdat velen stellen dat er geen passende benchmarks bestaan voor de rente-afdekking. Voorstanders van een echte benchmark verwijzen naar de eerste methodiek, die toch door een aantal fondsen met succes wordt toegepast. Tegenstanders voeren bijvoorbeeld aan dat tegenwoordig om strategische redenen vaak swaptions worden ingezet. Deze instrumenten zijn niet goed te benchmarken. Ook jarenlange gebruikers van de eerste methodiek komen terug van de gedachte dat een honderd procent zuivere benchmark mogelijk is, juist omdat zij ook swaptions zijn gaan inzetten. Het probleem voor wat betreft de geest van de wet is dat de uitvoerder niet wordt afgerekend op twee mogelijk aanzienlijke effecten: een (al dan niet tactisch bewust) verschil in omvang van de rente-afdekking en een juiste implementatie binnen de gekozen omvang.

Er lijkt sprake van een trend naar het gebruik van de tweede methodiek. Al vele jaren hebben grote pensioenfondsen deze methodiek gebruikt en zijn hun accountants hiermee akkoord gegaan. In termen van Molenkamp en Zweekhorst (2004) leidt dit tot jurisprudentie: een rente-overlay behoort tot het beleid en niet tot de uitvoering.

Discutabele gevallen. De derde “methodiek” betreft alle gevallen waarbij de benchmark niet representatief is voor het gevoerde beleid. We

bespreken hier twee gevallen waarbij de uiteindelijke keuze, op instigatie van de accountant, leidt tot extreme z-scores.

3a: Benchmark niet aangepast aan nieuw beleid. In het eerste geval stelt een pensioenfonds halverwege het jaar de rente-afdekking bij. In het beleggingsplan stond een bepaalde mate van afdekking vermeld, inclusief passende benchmarks. Het bestuur wil de afdekking strategisch sterk vergroten vanwege de gedaalde dekkinggraad. Dus niet vanwege een tactische rentevisie. Het mandaat van de vermogensbeheerder wordt aangepast en de grotere omvang van de rente-afdekking wordt geïmplementeerd. Het bestuur stelt geen formele nieuwe normportefeuille vast en publiceert deze dus ook niet. Volgens de letter van de wet moet de oude mate van afdekking dan worden gebruikt. Na het vergroten van de afdekking daalt de rente nog verder. De afdekking levert een sterk positieve bijdrage aan het rendement. In de niet aangepaste benchmark is de positieve bijdrage beduidend kleiner. Het fonds behaalt een forse outperformance, puur dankzij de toegenomen rente-afdekking. Na afloop van het jaar wordt een discussie gevoerd met de accountant, die vaststelt dat wel naar de letter maar niet naar de geest is gehandeld. Mede vanwege de mooie outperformance wordt besloten de oude norm te gebruiken in de z-score. Onder andere door Molenkamp en Zweekhorst (2004) is beschreven dat pensioenfondsen in dergelijke gevallen een “vrije gok” kunnen nemen: als de rente juist zou zijn gestegen, had men mogelijk achteraf wel gehandeld naar de geest in plaats van de letter.

Pure ruis die zowel positief als negatief kan zijn

3b: Onjuiste benchmark. In het tweede geval heeft het bestuur op advies van een externe adviseur een benchmark vastgesteld voor de rente-afdekking, die wordt geïmplementeerd via een overlay met grote leverage. Gedurende het jaar blijkt echter dat de tracking error tussen de feitelijke rente-afdekking en de benchmark ongeveer 100% is. De vermogensbeheerder heeft zich ook al eerder gedistantieerd van de benchmark als maat voor de performance, omdat al voor aanvang van het jaar was gebleken dat de benchmark zich volkomen anders kon gedragen dan de gebruikte instrumenten. Uit nader onderzoek blijkt dat de benchmark niet passend is. Het verschil tussen benchmark en portefeuille geeft geen enkele informatie over de uitvoering. Dit verschil is pure ruis die zowel positief als negatief kan zijn. De accountant oordeelt echter dat de benchmark voorafgaand aan het jaar is vastgesteld en daarom ook dient te worden gebruikt in de bepaling van de z-score. De corresponderende

outperformance zou normaal gesproken alleen kunnen zijn behaald met een extreem risicovolle implementatie van het beleggingsbeleid.

Standaarden en feitelijk rendement

De wetstekst bij de rendementen van de normportefeuille en de portefeuille luidt: “Hierbij wordt bij het bepalen van het rendement gebruik gemaakt van daartoe opgestelde richtlijnen van de Vereniging van Beleggingsanalisten.” De VBA is in Nederland via de Commissie Investment Performance Measurement aanspreekpunt voor standaarden op het gebied van performancemeting. Momenteel is de Engelstalige GIPS (Global Investment Performance Standards) 2010 ook in Nederland de actieve standaard. Enkelen hebben voorgesteld om in de wet te verwijzen naar GIPS. Dat voorstel leidt echter inherent tot problemen, want GIPS verbiedt aanduidingen als “de berekeningen zijn in overeenstemming met GIPS” of “er wordt gerekend volgens GIPS behalve voor...”. Alleen een vermogensbeheerder die voldoet aan alle vereisten van GIPS mag refereren aan GIPS. Bovendien is GIPS een vrijwillige standaard met veel vrijheidsgraden en zonder diepgravende controles op berekeningen. Zelfs een organisatie die geverifieerd GIPS compliant is kan niet stellen dat berekeningsresultaten met zekerheid juist zijn.

Eén van de zaken die door alle standaarden wordt geëist, is dat de normportefeuille moet passen bij de portefeuille. Anders worden appels met peren vergeleken. Bij de bepaling van de z-score hebben de meeste problemen betrekking op de passendheid van de normportefeuille. Ook de feitelijke portefeuille kent echter aandachtspunten. Soms zijn de basisgegevens onjuist of niet aansluitend op het voorgaande jaar. Ook wordt wegens formulefouten nu en dan een incorrect rendement berekend. Een aandachtspunt is de behandeling van illiquide beleggingen zoals private equity, direct vastgoed en moeilijk verhandelbare deelnemingen. Diverse pensioenfondsen zijn bijvoorbeeld eigenaar van hun uitvoeringsorganisatie, die kan zijn opgenomen in de beleggingsportefeuille. Bij een taxatie van de marktwaarde, die bijvoorbeeld eenmaal per vijf jaar plaatsvindt, kan blijken dat op deze belegging een zeer hoog rendement is behaald. Hoe zou dit rendement in de z-score verwerkt moeten worden? Handhaaft men de boekwaarde, dan worden toekomstige rendementen op een te laag onderliggend vermogen behaald, wat kan leiden tot overstatement van z-scores. Het opnemen van de geschatte marktwaarde leidt tot ofwel een hoog rendement in het betreffende jaar ofwel tot het herzien van de gehele reeks van z-scores. Deze mogelijke effecten, en de invloed die kan worden uitgeoefend op de waarderingen van de illiquide beleggingen, zorgen voor een mate van stuurbaarheid van de z-scores: er kunnen bewust stille reserves worden gecreëerd die later kunnen worden aangesproken. Ook dit punt leidt typisch tot discussies met de accountant en tot berekeningen die niet passen bij de standaarden. Conform de letter van de wet worden illiquide beleggingen voor de z-score vaak afgezet tegen

een rente-index vermeerderd met een opslag. Op langere termijn zouden illiquide beleggingen het beter moeten doen dan een aan de rente gekoppelde belegging. Op kortere termijn kunnen ze juist fors achterblijven. Beide verschijnselen geven spanning met de geest van de wet. Een mogelijke oplossing bij de evaluatie van de uitvoering zou zijn om als benchmarkrendement het feitelijke rendement te hanteren, al is dat volgens de letter van de wet niet toegestaan. Voor illiquide instrumenten en bijvoorbeeld ook voor opties ligt deze oplossing echter voor de hand.

Conclusies

Op basis van één en hetzelfde beleggingsplan en één en dezelfde werkelijke gang van zaken tijdens een jaar komen verschillende partijen tot significant verschillende z-scores. Dit is een onwenselijke situatie die kan leiden tot willekeur. Sommige verschillen ontstaan doordat bepaalde onjuistheden niet worden gedetecteerd bij controles. Andere verschillen komen door een verschil in interpretatie tussen de partijen, waardoor beleidsmatige keuzes wel of niet worden toegerekend aan de kwaliteit van de uitvoering. Omdat beleidsmatige keuzes meestal meer rendementsimpact hebben dan de uitvoering van het beleid, leiden de verschillen tot substantiële afwijkingen in z-scores.

In de loop der jaren is jurisprudentie ontstaan die echter niet bij alle betrokkenen bekend is. Ten bate van een eenduidige werkwijze rondom de z-score zou een en ander in kaart moeten worden gebracht. Het is wenselijk om de betrokken partijen tot nadere afstemming te laten komen.

Literatuur

- Braker, J.G., 2003, Pensioenfondsen leggen de lat te hoog, Economisch Statistische Berichten, 4396: 103-104.
- Jadnanansing, A., 2001, Toets onder de loep, Economisch Statistische Berichten, 4319: 600-602.
- Molenkamp, J.B. en C. Zweekhorst, 2002, Prestaties bedrijfstakpensioenfondsen geanalyseerd: de z-score zegt niet alles, VBA Journaal, december 2002, nr. 3: 24-30.
- Molenkamp, J.B. en C. Zweekhorst, 2003, Z-scores bedrijfstakpensioenfondsen 1998-2002 geanalyseerd, VBA Journaal, zomer 2003, nr. 2: 26-34.
- Molenkamp, J.B. en C. Zweekhorst, 2004, Z-scores bedrijfstakpensioenfondsen 1999-2003 geanalyseerd, VBA Journaal, zomer 2004, nr. 2: 37-41.
- Molenkamp, J.B., 2011, Z-toets revisited, VBA Journaal Lustrum 50 jaar, mei 2011, pp. 46-49.
- Zweckhorst, C., 2002, Het rendement van de normportefeuille. In: Hendriks, C.J.G.M., red, Performancemeting en Benchmarking, Kluwer, Deventer, pp. 129-143.

Noten

- 1 Hans Braker is consultant bij en directeur/eigenaar van Braker Investment Consulting B.V.
- 2 De z-score in enig jaar wordt berekend als $z = [(Rf - k) - (Rb - kb)] / E$, waarbij:
 Rf = feitelijk rendement portefeuille
 k = beleggingsuitvoeringskosten
 Rb = rendement normportefeuille
 kb = normkosten (afhankelijk van beleggingsmix)
 E = risicocorrectie (elementaire correctie, afhankelijk van beleggingsmix)