door Miriam Loois

BACKTESTEN BINNEN SOLVENCY II – HET EFFECT VAN HET GEBRUIK VAN EEN ROLLENDE HORIZON

Dit is een samenvatting van de scriptie die Miriam Loois heeft geschreven toen zij als Actuaris/Financial Risk Manager bij Delta Lloyd werkte. Haar begeleider bij de UvA was Roger Laeven. Miriam werkt nu als balansrisicomanager bij PGGM.

Binnen Solvency II bestaat de mogelijkheid voor verzekeraars om een intern model te ontwikkelen. Op deze modellen dient een backtest te worden uitgevoerd. Omdat er vaak maar een beperkte data historie beschikbaar is, wordt er vaak een 'rollende horizon' gebruikt. Dit heeft effect op de uitkomsten van de backtest. Een voorbeeld gebaseerd op de aandelen stressparameter in het FTK zal laten zien dat indien er niet wordt gecorrigeerd voor het gebruik van de rollende horizon, stressparameters onterecht als te laag kunnen worden gezien.

Value-at-Risk modellen binnen Solvency II

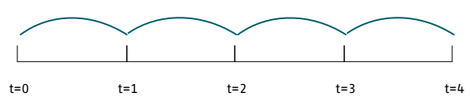
Indien een verzekeraar er voor kiest om een intern model te ontwikkelen, zal voor elk risico (bijvoorbeeld rente, aandelen, sterfte) een model moeten worden ontwikkeld. De Solvency II directive[1] schrijft voor dat voor het kalibreren van ieder van deze modellen een Value-at-Risk methodiek dient te worden gebruikt, met een 1-jaars horizon en een betrouwbaarheidsinterval van 99.5%. De stressparameter die uit zo'n model komt zou dus een '1-in-200 jaar' gebeurtenis moeten beschrijven. Gemiddeld genomen zou daarom eens in de 200 jaar een overschrijding van deze parameter voor mogen komen. Als voorbeeld zal in dit artikel aandelenrisico worden gebruikt. Binnen de Standaard Formule van Solvency II wordt een stress parameter van 39% voor aandelen gebruikt. Dit zou betekenen dat eens per 200 jaar de aandelen in één jaar tijd meer dalen van 39%.

Om dit met conventionele statistische technieken te backtesten heb je dus een erg lange data set nodig. Hierbij moet je denken aan minimaal 2000 jaar aan data. Hiervoor zijn onafhankelijke historische bewegingen nodig, dus bijvoorbeeld alle aandelen dalingen en stijgingen gemeten aan het einde van het jaar, ten opzichte van precies een jaar eerder. In de praktijk is het vaak al niet haalbaar om de relevante periode zelf, 200 jaar, aan historische data te verzamelen, laat staan een veelvoud hiervan.

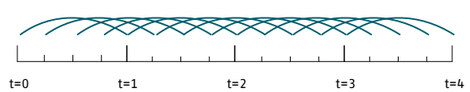
Backtesten met een rollende horizon

Om toch iets te kunnen zeggen over de geldigheid van het model, wordt de historische dataset vaak uitgebreid door een rollende horizon te gebruiken. Er wordt niet alleen gekeken naar jaarbewegingen gemeten aan het einde van het, maar ook naar bewegingen tussen januari en januari, februari en februari, enz. Hiermee wordt de dataset vergroot, echter, de datapunten zijn niet meer onafhankelijk. De historische bewegingen tussen januari en januari overlappen met de historische bewegingen tussen februari en februari. Daarnaast neemt de kans om een overschrijding van de stress parameter te vinden toe. Immers, een daling van de aandelen midden in het jaar kan aan het einde van het jaar weer hersteld zijn. Als men op een statistisch verantwoorde manier iets wil zeggen over de uitkomst van de backtest, zal er dus gecorrigeerd moeten worden voor het gebruik van deze rollende horizon.

Conventionele backtest (k=1)



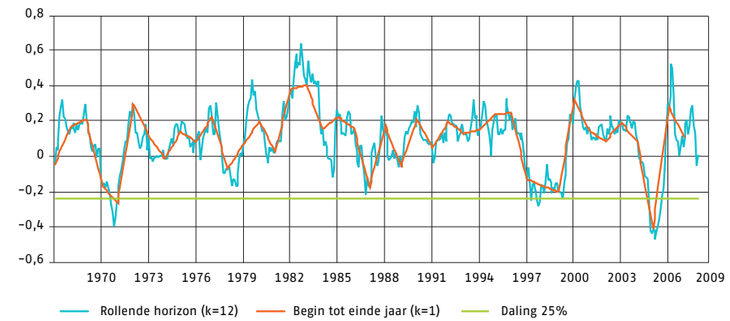
Backtesten met een rollende jaarhorizon (k=4)



Het effect van de rollende horizon

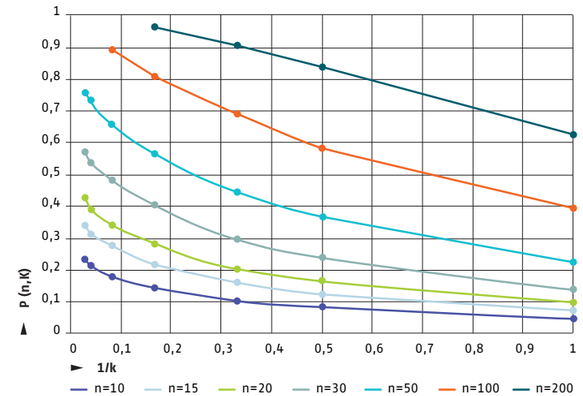
Om het effect van het gebruik van een rollende horizon te bepalen, is berekend wat de kans is om minimaal één overschrijding van de stress parameter te vinden in *n* jaar aan historische data, met een frequentie van *k*. Als aanname is genomen dat de historische aandelenbewegingen een Random Walk volgen. Indien *k=1* zijn er onafhankelijke data punten gebruikt. Indien *k=12* is er maanddata gebruikt (historische bewegingen tussen januari en januari, februari en februari, etc.). In het geval van *k=1* kan deze kans eenvoudig worden berekend. In n jaar aan data zijn *n-1* historische bewegingen te testen. De kans dat hier minimaal één overschrijding in voorkomt, *P(n,k)*, is gelijk aan 1 min de kans dat er geen overschrijdingen zijn. Indien *k* groter is kan deze kans worden gebaseerd op simulaties, of via het bepalen van een multivariaat normale verdeling die rekening houdt met de correlatie tussen de verschillende datapunten.

Historische veranderingen op jaarbasis in MSCI World index



Uit de resultaten blijkt dat de kans om een overschrijding te vinden substantieel toeneemt als een rollende horizon wordt gebruikt. Hoe korter de dataperiode en hoe hoger de frequentie van de data, hoe groter het effect. In figuur 'Random Walk' kan men zien dat als 30 jaar aan historische data wordt gebruikt, de kans 14% is om minimaal een overschrijding te vinden indien onafhankelijke jaardata wordt gebruikt (*n=30, k=1*). Indien een rollende horizon wordt gebruikt op basis van maanddata (*n=30, k=12*), stijgt deze kans naar 48%. Zomaar concluderen dat in 30 jaar aan data geen overschrijding mag plaatsvinden van een 1-in-200 jaar stress kan dus leiden tot het onterecht verwerpen van een stress parameter.

Random walk



Hoe dan wel correct backtesten met een rollende horizon?

Om te backtesten op basis van een rollende horizon, zal er gecorrigeerd moeten worden voor het gebruik van de rollende horizon. Indien er *n* jaar aan historische data wordt gebruikt met een frequentie *k* kan er door middel van simulaties worden bepaald wat de theoretische kansverdeling is van het aantal overschrijdingen in de data. Door de theoretische kans te berekenen dat er evenveel of meer overschrijdingen plaats vinden als waargenomen in de historische data, kan een *p-waarde* worden berekend. Indien deze *p-waarde* te laag is, bijvoorbeeld kleiner dan 5%, kan het model verworpen worden.[2]

De aandelen stress parameter in het FTK

In het Financieel Toetsingskader voor pensioenfondsen wordt uitgegaan van een aandelen stress parameter van 25%. Ook binnen het FTK wordt de Value-at-Risk methodiek gebruikt, echter met een betrouwbaarheidsinterval van 97.5% in plaats van 99.5% in Solvency II. Een 25% daling zou dus moeten corresponderen met een 1-in-40 jaar gebeurtenis. Onlangs is deze parameter kritisch bekeken in de evaluatie van het FTK. De commissie[3] die deze evaluatie heeft uitgevoerd, concludeert dat er sinds 1970 zich vier keer een daling heeft voorgedaan die groter is dan 25%, waarvan twee in de afgelopen zeven jaar, en dat dit hoger is dan op basis van de kansverdeling mag worden verwacht.

Als je alleen naar veranderingen in de aandelenindex kijkt tussen het begin en einde van het jaar, heeft er zich niet vier, maar slechts twee keer een aandelendaling groter dan 25% voorgedaan. In ruim 40 jaar aan data is dit statistisch gezien geen reden om de hoogte van de stress parameter te verwerpen. Het correspondeert met een *p-waarde* van 0.26. Indien een rollende horizon wordt gebruikt op basis van maanddata, vindt men 21 keer een daling groter dan 25%. Veel van deze overschrijdingen zijn geclusterd. Indien men de backtest methodiek toepast zoals hiervoor beschreven, die corrigeert voor het gebruik van een rollende horizon, vindt men een *p-waarde* van 0.14. Op basis hiervan zou de stress parameter van 25% dus ook niet verworpen hoeven te worden.

De hiervoor genoemde commissie die het FTK heeft geëvalueerd heeft niet alleen gekeken naar het aantal overschrijdingen, maar ook naar de grootte van de overschrijdingen. Het is dus niet zo dat dit artikel wil beargumenteren dat de hoogte van stressparameter juist is. Echter, men kan concluderen dat bij het backtesten statistisch gezien snel verkeerde conclusies kunnen worden getrokken indien men niet corrigeert voor het gebruik van een rollende horizon.

Conclusies

Indien men backtest op basis van een rollende horizon, kan intuïtie snel misleiden, wat kan leiden tot het onterecht verwerpen van een model of de hoogte van een stress parameter. Indien men toch een rollende horizon wil gebruiken vanwege een beperkte datahistorie, kan een backtest methodiek gebruikt worden die hiervoor corrigeert.

1 – [1] DIRECTIVE 2009/138/EC, 2009

2 – Voor een uitgebreidere beschrijving van deze methodiek kan informatie worden opgevraagd bij de auteur van dit artikel.

3 – De commissie is ingesteld door de toenmalig minister van Sociale Zaken en Werkgelegenheid Donner met als taak het FTK te evalueren. Het onderzoek is uitgevoerd door de Nederlandse Bank en het Ministerie van Sociale Zaken en Werkgelegenheid.