CPB Memorandum

Sector	:	2
Afdeling/Project	:	Conjunctuur
Samensteller(s)	:	Lennart Dek
Nummer	:	253
Datum	:	16 december 2010

BBP-groei voorspellen met leading indicators: het frequentieverschil gemodelleerd

Het voorspellen van BBP-groei met leading indicators kent een aantal moeilijkheden. In dit onderzoek wordt gekeken naar het probleem van gemengde beschikbaarheid: BBP-cijfers worden op kwartaalbasis gepubliceerd, terwijl veel leading indicators elke maand uitkomen. Het recent voorgestelde MIDAS-model, dat met deze gemengde beschikbaarheid om kan gaan, wordt hier vergeleken met een aantal nieuw voorgestelde alternatieven. Daarnaast wordt onderzocht of poolen over meerdere voorspellingen de voorspelprestatie verbetert. Het onderzoek laat zien dat het conventionele MIDAS-model met exponential Almon vertragingsfunctie goed presteert, maar dat het hier voorgestelde gebruik van modelselectie criteria de voorspelprestatie nog verder kan verbeteren. Ook wordt aangetoond dat poolen over voorspellingen gedaan door meerdere multivariate modellen de beste resultaten oplevert en dat deze methode tot 66 procent beter presteert dan een naïeve benchmark.

Steekwoorden: BBP, voorspellen, leading indicators, out-of-sample, frequentieverschil, poolen

Inhoudsopgave

1 Inleiding

Het bruto binnenlands product (BBP) wordt gezien als dé maatstaf voor geaggregeerde economische activiteit. Er zijn meerdere mogelijkheden om de ontwikkeling van het volume van het BBP te voorspellen. Een daarvan is het gebruik van een groot macro-economisch model, zoals SAFFIER.[1] Een dergelijk model probeert de belangrijkste aspecten van de economie mee te nemen en wordt dan ook niet louter benut om BBP te voorspellen, maar levert voorspellingen op voor een breed scala aan economische variabelen.

Een andere methode behelst het gebruik van leading indicators. Dit zijn variabelen waarvan wordt verondersteld dat ze voorlopen op de geaggregeerde economische activiteit, i.e. een bepaalde ontwikkeling van een dergelijke variabele is een indicatie dat de geaggregeerde economische activiteit een soortgelijke (of tegenovergestelde) ontwikkeling te wachten staat. Een veelgebruikt voorbeeld voor een leading indicator is het aantal bouwaanvragen. Een stijging van deze grootheid is een indicatie dat de activiteit in de bouwsector gaat toenemen en daarmee ceteris paribus ook de geaggregeerde economische activiteit. Hieruit kan worden afgeleid dat leading indicators informatie bevatten over toekomstige economische activiteit en daarmee ook over het toekomstige BBP, wat het gebruik van leading indicators om BBP te voorspellen rechtvaardigt.

De benadering om leading indicators te gebruiken om BBP te voorspellen is compleet verschillend van die van grootschalige modellen als SAFFIER. Laatstgenoemde gaat op zoek naar structurele economische verbanden die een stevige basis hebben in de economische theorie. Op basis van de schattingen van deze relaties worden vervolgens voorspellingen geconstrueerd. Modellen die leading indicators gebruiken om te voorspellen houden zich niet bezig met structurele relaties en missen de theoretische basis. Deze modellen proberen louter zoveel mogelijk informatie over het toekomstige BBP te halen uit een gegeven dataset. Variabelen worden in deze dataset opgenomen omdat wordt verondersteld dat ze informatie over toekomstige economische omstandigheden bevatten, niet vanwege een causale relatie met het BBP. Zo zijn dalende beurskoersen een teken van teruglopend vertrouwen in de prestaties van beursgenoteerde bedrijven en daarmee een indicatie dat economische activiteit af zou kunnen gaan nemen. Echter, het is niet waarschijnlijk dat er een causale relatie bestaat tussen de beurskoers en BBP, in die zin dat dalende beurskoersen een dalend BBP veroorzaken.

Niet alleen is de insteek van de twee aanpakken anders, ze gebruiken ook verschillende informatie. Het SAFFIER model gebruikt bijvoorbeeld honderden ‘harde’ economische variabelen. Het aantal variabelen dat in aanmerking komt om dienst te doen als leading indicator is veel kleiner en komt in de meeste onderzoeken niet boven de vijftig uit. Bovendien

[1] Zie Kranendonk en Verbruggen (2006) voor een beschrijving van SAFFIER of CPB (2010) voor een beschrijving van SAFFIER II.

zijn deze cijfers vaak 'zacht', e.g. de uitkomsten van enquêtes onder consumenten en producenten, en slechts indirect verbonden met BBP, zoals het hierboven gegeven voorbeeld van het aantal bouwaanvragen. Het CPB beschikt sinds 1990 over de CPB-conjunctuurindicator waarin informatie van zo'n 25 variabelen wordt samengevat in een barometer die de op- en neergangen van de conjunctuur beschrijft.[2]

Deze verschillen betekenen dat de ene aanpak niet 'beter' kan zijn dan de andere, aangezien het doel, onderbouwing en gebruikte informatie erg verschilt. Het maakt de twee benaderingswijzen juist complementair: de gecombineerde voorspellingen van de twee methoden bevatten meer informatie dan de afzonderlijke voorspellingen. Daarom kunnen de aanpakken zeer goed naast elkaar worden gebruikt, zoals het CPB ook daadwerkelijk doet.[3]

Het gebruik van leading indicators om economische omstandigheden te voorspellen kent een lange historie, die teruggaat tot het pionierswerk van Mitchell en Burns (1938). Dat betekent echter niet dat er intussen een consensus is bereikt over de welke methode het best gebruikt kan worden als wordt voorspeld met leading indicators. Er zijn vele methoden voorgesteld in de literatuur en de keuze voor een van hen hangt af van de exacte toepassing waarvoor leading indicators worden gebruikt.

Hier is het doel om de volumegroei van het BBP te voorspellen. Dit doel kent een aantal complicaties, zoals het feit dat cijfers over het BBP op kwartaalbasis worden gepubliceerd, terwijl leading indicators maandelijks of nog frequenter uitkomen. Dit is een probleem aangezien standaardtechnieken ervan uitgaan dat de afhankelijke en de verklarende variabelen met dezelfde frequentie worden waargenomen.

Het hier beschreven onderzoek concentreert zich op dit probleem en onderzoekt verschillende methoden om het frequentieverschil te modelleren. De aandacht richt zich op het MIDAS-model: een recent voorgesteld model dat gebruikt kan worden als er een frequentieverschil is tussen de afhankelijke en verklarende variabelen. In dit onderzoek wordt een aantal alternatieven voorgesteld voor de keuzes die in de huidige literatuur over het MIDAS-model worden gemaakt. De nieuwe en bestaande alternatieven worden met elkaar vergeleken op basis van theoretische argumenten en de resultaten van een out-of-sample voorspelexperiment. Vervolgens wordt onderzocht hoe de gebruikte leading indicators en alternatieven het best kunnen worden ingezet om te komen tot een enkele puntvoorspelling voor BBP-groei door middel van het poolen van voorspellingen.

Dit document is als volgt opgebouwd. De volgende paragraaf bespreekt de problemen die gepaard gaan met het gebruik van leading indicators om te voorspellen. In paragraaf 3 komen de alternatieve benaderingen aan bod die kunnen worden gebruikt om het frequentieverschil te

[2] Zie Kranendonk, Bonenkamp, Verbruggen (2003) voor een beschrijving van de CPB-conjunctuurindicator. Met ingang van het Centraal Economisch Plan 2011 wordt een vernieuwde CPB-conjunctuurindicator in gebruik genomen die beschreven is in Muns en Kranendonk (2010).

[3] Zie paragraaf 3.2 van Kranendonk, Bonenkamp en Verbruggen (2003).

modelleren. Paragraaf 4 staat in het teken van de opzet van het out-of-sample voorspelexperiment en behandelt ook de data die bij dit experiment is gebruikt. De resultaten van dit experiment zijn het onderwerp van paragraaf 5. Paragraaf 6 gaat dieper in op de verschillende methoden om tot een enkele voorspelling te komen met behulp van poolen. Tot slot worden de conclusies besproken in paragraaf 7.

2 Voorspellen met behulp van leading indicators

Leading indicators worden vaak gebruikt om economische ontwikkelingen te voorspellen. Om de staat van de economie te voorspellen is het nodig om deze te kunnen meten. Het bruto binnenlands product wordt gezien als dé maatstaf voor de geaggregeerde economische activiteit. Daarom wordt het BBP vaak gebruikt als doelvariabele, i.e. het voorspellen van de economische condities wordt gedaan door de volume-ontwikkeling van het BBP te voorspellen.

In de bestaande literatuur over leading indicators wordt in de plaats van BBP vaak geopteerd voor het gebruik van een zogenoemde samenvallende indicator, zie bijvoorbeeld Stock en Watson (1989) en Forni et al. (2001). Deze keuze wordt gemaakt omdat er aan het gebruik van BBP als doelvariabele een aantal nadelen kleeft. Ten eerste gaat het BBP gebukt onder vaak grote dataherzieningen. Hierdoor moet rekening worden gehouden met meetfouten als er voorlopige versies van de cijfers worden gebruikt. Ten tweede wordt een eerste schatting voor het gerealiseerde BBP pas 40 dagen na het eind van het kwartaal bekend gemaakt. Als deze schatting benodigd is om tot een nieuwe voorspelling te komen, duurt het ook ten minste 40 dagen voordat zo'n voorspelling kan worden gemaakt. Een derde probleem is dat BBP slechts één keer per kwartaal wordt berekend, terwijl veel leading indicators frequenter (i.e. op maandbasis) beschikbaar zijn. Econometrische standaardtechnieken kunnen bij een dergelijke gemengde beschikbaarheid niet worden gebruikt.

De dataherzieningen impliceren dat men te maken heeft met meetfouten die voortkomen uit de revisieprocedure. De meeste macro-economische reeksen worden in meerdere rondes gepubliceerd: aanvankelijk worden deze aangekondigd als eerste schatting, waarna er nog enkele revisies plaatsvinden voordat het uiteindelijke getal vaststaat. Er is een aantal oorzaken aan te wijzen voor deze herzieningen. Ten eerste kan het het geval zijn dat niet alle benodigde data beschikbaar is op het moment dat de macro-economische reeks wordt geconstrueerd. Een tweede mogelijkheid is dat de procedure om seizoenspatronen uit de data te verwijderen, een standaardprocedure bij het publiceren van macro-economische data, niet robuust is voor de toevoeging van nieuwe gegevens. Ten derde kunnen definities van reeksen door de tijd veranderen, waardoor de wijze waarop deze worden berekend ook verandert.

Deze drie mogelijkheden zorgen ervoor dat de voorlopige versies van een gepubliceerde reeks erg kunnen verschillen van de slotversie. Echter, vaak worden de herziene data gebruikt om geschikte verklarende reeksen te identificeren, het model te specificeren en te schatten, terwijl het voorspellen in de praktijk plaatsvindt met voorlopige versies van de data. Dit kan de voorspelresultaten op meerdere manieren beïnvloeden. Allereerst, en meest voor de hand liggend, veranderen de herzieningen de input van het voorspelmodel. Een andere mogelijkheid is dat de parameterschattingen veranderen door het gebruik van aangepaste data. Ten slotte kan

het zo zijn dat uit het modelselectieproces een andere specificatie naar voren komt bij het gebruik van de nieuwere versies van de data.

Meerdere onderzoeken laten zien dat het negeren van het probleem ertoe kan leiden dat de voorspelprestaties slechter worden, zie Croushore (2006) voor een overzicht. Als het gaat om leading indicators hebben Diebold en Rudebusch (1991) gedemonstreerd dat hoewel de samengestelde leading indicator van de Amerikaanse Conference Board voorspellende waarde heeft als er finale data wordt gebruikt, deze voorspelkracht sterk achteruit gaat bij het gebruik van de eerste versie van de data.

Het is niet gemakkelijk om het effect van dataherzieningen op voorspelprestatie te kwantificeren. Echter, er zijn redenen die het aannemelijk maken dat het probleem in de huidige toepassing relatief klein is. Ten eerste is hier het streven om de volumegroei van het BBP te voorspellen en niet het niveau. Het onderzoek van Howrey (1996) wijst erop dat meetfouten een groter probleem vormen wanneer er niveaus worden voorspeld dan bij het voorspellen van groeivoeten. Ten tweede impliceert het werk van Stark en Croushore (2002) dat voorspellingen voor inflatie gevoeliger zijn voor dataherzieningen dan outputvoorspellingen. Een mogelijke verklaring hiervoor is dat eerstgenoemde persistenter is. Ten derde is de doelvariabele in het onderzoek van Diebold en Rudebusch (1991) de samengestelde leading indicator van de Conference Board. Deze gaat gebukt onder forse herzieningen: niet alleen worden statistische meetfouten hierin opgenomen die voortkomen uit het gebruik van meer en actuelere data, ook veranderen zijn componenten regelmatig (Klein, 2001). Dit laatste geldt niet voor het BBP. Ten slotte zijn herzieningen volgens Kuzin et al. (2009) sinds het jaar 2000 over het algemeen klein. Zij verwijzen naar Marcellino en Musso (2010), die vinden dat de correlatiecoëfficiënt tussen de eerste en uiteindelijke versie van de data 0,98 bedraagt.

Door deze redenen denken wij dat het probleem van dataherzieningen bij het voorspellen van de BBP-groei relatief klein is. Daardoor zijn we van mening dat het gerechtvaardigd is om de meetfouten niet expliciet mee te nemen in dit onderzoek. Wel is het effect van het gebruik van voorlopige dataversies op de voorspelprestatie een interessant onderwerp voor vervolgonderzoek.

De twee andere problemen, gemengde beschikbaarheid van data en de vertraagde publicatie van data over BBP, stellen duidelijke eisen aan het voorspelmodel. Dit model moet in staat zijn om te gaan met data van een gemengde frequentie. Bovendien heeft een model dat niet de volledige data nodig heeft om tot een nieuwe voorspelling te komen de voorkeur boven een model dat dat wel heeft.

Het onderzoek waarvan dit document verslag doet heeft met name betrekking op het probleem van gemengde beschikbaarheid van data; alle hier beschouwde modellen voldoen aan

de tweede eis, zodat daar verder geen aandacht aan wordt besteed. In de volgende paragraaf worden de modellen besproken die in dit onderzoek zijn gebruikt.

3 Frequentieverschil modelleren

Het probleem van gemengde beschikbaarheid van data is niet onopgemerkt gebleven in de literatuur. Er is reeds een aantal oplossingen voorgesteld. Onder deze oplossingen bevinden zich de Mixed Frequency Vector Autoregressive (MF-VAR) modellen, *linkage models* en *bridge equations*. De MF-VAR modellen hebben hun oorsprong in Zadrozny (1988) en zijn later uitgebreid door onder meer Zadrozny (1990) en Mariano en Murasawa (2003 en 2004). In deze aanpak wordt de frequenter geobserveerde variabele samen met de niet geobserveerde frequente tegenhanger van de lage frequentie variabele gemodelleerd in een VAR-proces. Het model wordt geschat met behulp van een Kalman filter. Greene et al. (1985) waren de eersten die een methode voorstelden om gemengde data met elkaar te verbinden. Het idee van deze zogenoemde *linkage models* is om de frequente en minder frequente reeksen afzonderlijk te modelleren en hun voorspellingen daarna samen te voegen. *Bridge models* worden onder andere gebruikt door Klein en Sojo (1989), Parigi en Schlitzer (1995) en Baffigi et al. (2004). Deze modellen gebruiken de informatie in tijdig gepubliceerde (frequente) indicatoren om de (lage frequentie) variabelen te voorspellen die met vertraging bekend worden gemaakt. In deze methode worden ARIMA-modellen gebruikt om tot afzonderlijke voorspellingen te komen voor de indicatoren, welke vervolgens worden geaggregeerd tot de frequentie van de te voorspellen reeks. De voorspelde waardes van de indicatoren worden dan in de *bridge equation* gesubstitueerd om een voorspelling voor de afhankelijke variabele te verkrijgen.

Een andere methode die recent populair is geworden is het MIDAS-model. MIDAS staat voor MIxed DAta Sampling en is geïntroduceerd door Ghysels et al. (2002 en 2005). MIDAS-modellen werden oorspronkelijk gebruikt in financiële toepassingen, maar bleken al snel ook benut te kunnen worden voor macro-economische doeleinden, e.g. Clements en Galvào (2008 en 2009). Het MIDAS-model projecteert de afhankelijke variabele direct op de verklarende variabelen, waardoor het mogelijk is meerdere frequenties te combineren in één model. Dit document zal zich op deze methode richten, aangezien dit de nieuwste en minst onderzochte methode is om te voorspellen met gemengde frequentie data. Bovendien blijkt uit recent onderzoek dat MIDAS beter presteert in termen van voorspelnauwkeurigheid dan state-space modellen wanneer de nabije toekomst wordt voorspeld, zie Kuzin en al. (2009) en Bai et al. (2010).

Dit document onderzoekt verschillende kwesties gerelateerd aan het probleem van voorspellen met data van gemengde frequentie. Het evalueert het MIDAS-model langs drie dimensies: het model zelf, de structuur die wordt meegegeven aan vertragingen en de wijze waarop de uiteindelijke specificatie wordt geselecteerd. Het doel bij elk van deze dimensies is om te komen tot een methode die zo goed mogelijk voorspelt. Hoewel in de bestaande literatuur de

voorspelprestatie van het MIDAS-model wel is vergeleken met andere modellen, is er nog niet eerder onderzoek gedaan naar de andere twee dimensies.

Het vervolg van deze paragraaf kent de volgende opbouw. De eerste subparagraaf gaat dieper in op het MIDAS-model en introduceert het hier voorgestelde alternatief: het *autoregressive distributed lag model*. Subparagraaf 2 bespreekt de verschillende methoden om structuur mee te geven aan de vertragingen. In subparagraaf 3 worden diverse alternatieven vergeleken om een specificatie te selecteren.

3.1 Model

Het MIDAS-model lijkt enigszins op het veel bekendere *distributed lag* (DL)-model. Het DL-model veronderstelt dat de afhankelijke variabele (y_t) niet alleen gerelateerd is aan de verklarende variabele (x_t), maar ook aan waardes van deze variabele van een aantal tijdsperioden geleden (x_{t-l}). Een standaardvoorbeeld van het DL-model kan worden uitgedrukt als:

$$y_t = \beta_0 + B(L)x_t + \varepsilon_t. \tag{1}$$

Hier is $B(L)$ de vertragingsfunctie en L de vertragingsoperator zodanig dat $Lx_t=x_{t-1}$. Vergelijking (1) kan hier natuurlijk niet worden gebruikt, omdat het gelijke beschikbaarheid van y_t en x_t veronderstelt. In het geval dat in dit document centraal staat, wordt de verklarende variabele x_t (de leading indicator) vaker waargenomen dan de afhankelijke variabele y_t (BBP-groei). Bij maandelijkse waargenomen verklarende variabelen en een afhankelijke variabele die op kwartaalbasis wordt geobserveerd, is het frequentieverschil drie. Om dit verschil duidelijk te maken in de formules, wordt de verklarende variabele daarom genoteerd als y_t . Met behulp van deze notatie kan een gestileerde versie van het MIDAS-model worden aangeduid als:

$$y_t = \beta_0 + B(L)x_t^{(3)} + \varepsilon_t^{(3)}. \tag{2}$$

Hoewel de twee uitdrukkingen (1) en (2) erg op elkaar lijken, is er ook een aantal verschillen tussen de twee modellen. Een voorbeeld hiervan is dat het DL-model gemakkelijk kan worden uitgebreid met een autoregressief element, i.e. met een vertraagde waarde van de afhankelijke variabele. Dit is niet zo gemakkelijk in het MIDAS-model. Clements en Galvào proberen het probleem op te lossen, wat resulteert in het autoregressieve MIDAS (AR-MIDAS) model:

$$y_t = \beta_0 + \lambda y_{t-1} + B(L)(1-\lambda L)x_t^{(3)} + \varepsilon_t^{(3)}.$$

In deze uitbreiding wordt dus een gemeenschappelijke factor (1-λL) verondersteld tussen de verklarende en afhankelijke variabele.

Het autoregressieve DL (ADL)-model heeft deze veronderstelling niet nodig: het legt geen restricties op aan de coëfficiënten. Helaas kan het (A)DL-model in zijn vorm als in (1) niet worden geschat, omdat het veronderstelt dat de verklarende en afhankelijke variabele even vaak worden geobserveerd. Echter, dat BBP-groei niet maandelijks wordt geobserveerd betekent niet dat er niet zo iets bestaat als maandelijkse BBP-groei. Wanneer deze grootheid wordt aangeduid als $y_t^{(3)}$, veronderstelt een ADL-model de volgende relatie tussen maandelijkse BBP-groei en de leading indicator:

$$y_t^{(3)} = \beta_0 + \alpha y_{t-1/3}^{(3)} + B(L)x_t^{(3)} + \varepsilon_t^{(3)}. \tag{3}$$

In deze vorm is er sprake van het ADL-model. Wanneer de restrictie dat α=0 wordt opgelegd, is er sprake van het DL-model. Zoals gezegd kan vergelijking (3) niet direct worden geschat, aangezien maandelijkse BBP-groei $y_t^{(3)}$ niet wordt waargenomen. Wanneer echter enige wiskundige manipulatie wordt toegepast, kan (3) worden herschreven als[4]:

$$y_t = 3(1+\alpha+\alpha^2)\beta_0 + \alpha^3 y_{t-1} + \\ (\tfrac{1}{3} + \tfrac{2}{3}L^{1/3} + L^{2/3} + \tfrac{2}{3}L + \tfrac{1}{3}L^{4/3})(1+\alpha L^{1/3} + \alpha^2 L^{2/3})(B(L)x_t^{(3)} + \varepsilon_t^{(3)}). \tag{4}$$

Hoewel deze uitdrukking langer en ingewikkelder is, heeft hij één groot voordeel ten opzichte van (3): er zit geen niet-geobserveerde term $y_t^{(3)}$ meer in. Dit betekent dat deze uitdrukking wel kan worden geschat en dus dat het ADL-model kan worden gebruikt wanneer sprake is van gemengde beschikbaarheid van data.

Een vergelijking van het (AR-)MIDAS-model met het (A)DL-model laat zien dat beide zowel voor- als nadelen hebben. Zoals eerder aangegeven is een voordeel van laatstgenoemde dat de uitbreiding met een autoregressief component gemakkelijk gaat en geen extra aannames vereist. Een nadeel is echter zijn ingewikkelde vorm zoals getoond in (4). Dit maakt het schatten van de vergelijking een stuk ingewikkelder dan bij het MIDAS-model. Bovendien kan het MIDAS-model ook worden toegepast als er meer dan twee verschillende frequenties in de data zitten, terwijl het ADL-model dat niet toestaat. Een voordeel van het ADL-model is daarentegen dat de achterliggende intuïtie eleganter is. De veronderstelling dat er een direct verband is tussen maandelijkse BBP-groei en een maandelijkse leading indicator lijkt namelijk natuurlijker dan de veronderstelling dat er verband is tussen maandelijkse leading indicators en kwartaalgroei.

[4] Zie Dek (2010) voor de volledige afleiding.

In deze subparagraaf zijn twee alternatieve modellen besproken: het conventionele (AR-)MIDAS-model en het in deze toepassing nieuw voorgestelde (A)DL-model. Duidelijk is geworden dat beide modellen hun voor- en nadelen hebben. Het is daarom niet mogelijk om de twee te rangschikken op basis van louter theoretische argumenten. Daarom zullen beide later in dit document worden vergeleken op basis van hun vermogen om BBP-groei accuraat te voorspellen.

3.2 Vertragingsfunctie

Beide modellen die in dit document worden beschouwd, veronderstellen dat BBP-groei niet alleen gerelateerd is aan de waarde van de leading indicators in de afgelopen maand, maar mogelijk ook aan zijn waardes in de maanden ervoor. Dit betekent bijvoorbeeld dat BBP-groei in het derde kwartaal niet alleen kan worden verklaard aan de hand van de beurskoers in juni, maar ook die van mei en wellicht zelfs de maand(en) daarvoor.

In de vorige subparagraaf is reeds de vertragingsfunctie B(L) geïntroduceerd. Deze functie is zo gedefinieerd dat het de gewichten toekent aan de vertragingen van de leading indicators. Stel bijvoorbeeld dat het effect van de beurskoers van twee maanden geleden half zo groot is als het effect van één maand geleden, dan zal de vertragingsfunctie ervoor zorgen dat er aan de beurskoers van twee maanden geleden een tweemaal zo kleine coëfficiënt wordt meegegeven als aan de beurskoers van de afgelopen maand.

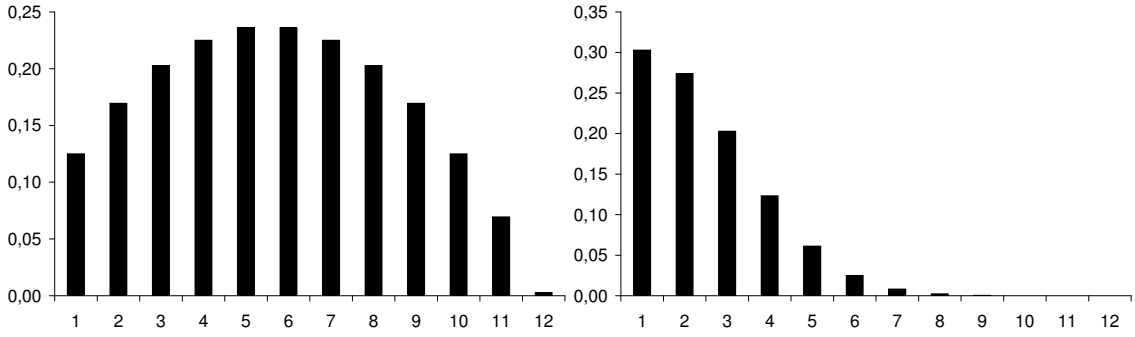
Het uiteindelijk toegekende gewicht hangt van twee factoren af: de gebruikte vertragingsfunctie en de geschatte parameters. Van tevoren kan de econometrist beslissen of hij zelf al enige structuur wil opleggen aan de gewichten. Als dat het geval is, kan hij kiezen voor een vertragingsfunctie die ervoor zorgt dat de uiteindelijke gewichten voldoen aan deze structuur voor welke geschatte parameterwaardes dan ook. Zo kan hij er bijvoorbeeld voor zorgen dat de gewichten afnemen met de tijd of dat ze allemaal positief zijn. Een andere mogelijkheid is om de vertragingsfunctie volledig vrij te laten. In dit geval zullen alle gewichten afzonderlijk worden geschat.

In zijn algemeenheid geldt dat voor functies met veel structuur weinig parameters hoeven te worden geschat en dat deze functies ‘mooiere’ gewichtenverdelingen opleveren. Dit laatste is ook logisch, als van tevoren de veronderstelde structuur wordt opgelegd, zal deze er ook uitkomen. Een nadeel van het opleggen van veel structuur is dat er aannames worden gemaakt over de gewichten. Als deze aannames worden overtreden zullen de gewichten onzuiver of inconsistent worden geschat.

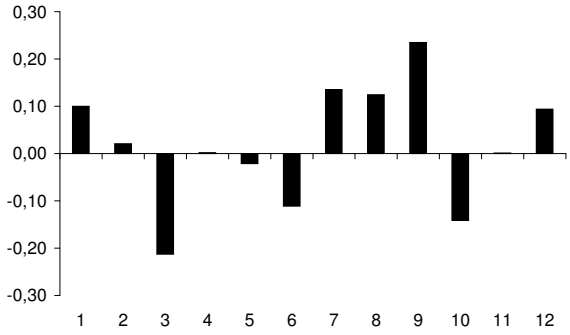
Er zijn vele mogelijkheden voor een vertragingsfunctie. In dit document worden drie alternatieven onderzocht. Het eerste is de *exponential* Almon functie. Deze wordt in de bestaande literatuur vaak gebruikt in combinatie met het MIDAS-model. De *exponential* Almon functie is zeer flexibel: zelfs met een beperkt aantal parameters kan hij vele vormen aannemen. De functie is recent geïntroduceerd (zie Ghysels et al, 2007), maar is losjes gebaseerd op de Almon functie die al veel langer bekend is (Almon, 1965). Gewichten die het resultaat zijn van de Almon functie liggen op een polynoom. Zowel de Almon als *exponential* Almon functie zijn voorbeelden van vertragingsfuncties die de gewichten een bepaalde structuur meegeven. Als derde alternatief wordt derhalve gekeken naar de *unconstrained* vertragingsfunctie. Zoals de naam al zegt, laat deze functie de gewichten volledig vrij. Voor de exacte functies wordt hier verwezen naar Dek (2010). In de figuren 3.1 en 3.2 wordt van elk van de functies een voorbeeld gegeven hoe de gewichten eruit kunnen zien.

In de figuren 3.1 en 3.2 zijn drie verschillende vertragingsstructuren te zien. In alle drie de figuren zijn twaalf vertragingen opgenomen: bij maandelijkse leading indicators betekent dit één jaar aan vertragingen. In de figuren geeft de hoogte van de balk het gewicht aan. Figuur 3.1 (links) geeft een voorbeeld van de Almon vertragingsfunctie weer. De gewichten bevinden zich in dit geval op een polynoom van tweede orde. Figuur 3.1 (rechts) laat een vertragingsstructuur zien die kan voortkomen als voor de *exponential* Almon functie wordt gekozen. In deze figuur is te zien dat de gewichten exponentieel afnemen. De vertragingen in figuur 3.2 zijn ten slotte zeer afwisselend. Hier is sprake van de *unconstrained* functie.

Figuur 3.1 Vertragingsfuncties, Almon en *exponential* Almon



Figuur 3.2 Vertragingsfunctie unconstrained



Evenals in het geval van de modellen hebben ook alle hier besproken functies hun voor- en nadelen. De Almon en *exponential* Almon functie zijn spaarzaam in dat slechts een beperkt aantal parameters benodigd is om alle gewichten vast te leggen. Dit is het tegenovergestelde van de *unconstrained* functie, welke het schatten van evenveel parameters als vertragingen vereist. Daar staat tegenover dat de *unconstrained* functie geen a priori restricties oplegt aan de gewichten, iets wat wel het geval is voor de andere functies. De *exponential* Almon functie is weliswaar zeer flexibel, maar legt wel op dat alle gewichten hetzelfde teken hebben. Gewichten gegenereerd door de Almon functie liggen allen op een polynoom. Als deze restricties worden overtreden, kan dit leiden tot onzuivere of inconsistente schattingen voor de gewichten. Tot slot is de *exponential* Almon functie zeer ingewikkeld. Hierdoor is het schatten van modellen met de *exponential* Almon vertragingen veel moeilijker dan het schatten van modellen met Almon of *unconstrained* vertragingen.

Dit betekent dat elk van de vertragingsfuncties goede en mindere eigenschappen heeft. Ook hier zal het vermogen van de functies om BBP-groei accuraat te voorspellen de doorslag moeten geven.

3.3 Specificatieselectie

Wanneer de keuze is gemaakt voor een bepaald model en vertragingsfunctie, blijven nog twee vragen over. De eerste behelst de vraag hoeveel vertragingen van de leading indicators moeten worden opgenomen in het model. Een tweede kwestie is de beslissing wat betreft het aantal parameters waarmee de spaarzame functies moeten worden geschat.

In de bestaande literatuur worden deze twee keuzes vaak min of meer ad hoc gemaakt. Zo kiezen Clements en Galvào (2008) voor een exponential Almon vertragingsfunctie met twee parameters en nemen ze steeds 24 vertragingen oftewel twee jaar van de verklarende variabelen op in hun modellen. Echter, de ad hoc keuze lijkt moeilijk te rechtvaardigen. Zo blijkt uit de literatuur over de Almon en unconstrained vertragingsfuncties dat als een verkeerd aantal vertragingen of parameters wordt gekozen, dit in het algemeen leidt tot onzuivere of inefficiënte

schatters voor de gewichten. Hieruit kan worden afgeleid dat het de voorkeur heeft de specificatie met zorg te kiezen.

Er zijn twee statistische manieren om te specificatie te selecteren. De eerste is het uitvoeren van een serie genestelde toetsen. Zo'n procedure kan er als volgt uit zien. Als eerste stap moet het maximale aantal vertragingen worden bepaald dat men in overweging wilt nemen. Hierna wordt het model geschat met dit aantal vertragingen. Op basis van de schattingsresultaten kan worden getoetst of de laatste vertraging een coëfficiënt van nul heeft. Als de hypothese van een gewicht van nul niet wordt verworpen, impliceert dit dat deze vertraging ten onrechte is opgenomen in het model. Daarom wordt bij het achterwege blijven van zo'n verwerping het model wederom geschat, dit maal met één vertraging minder. De procedure gaat door totdat de hypothese wel wordt verworpen. Het aantal vertragingen in het laatst geschatte model is dan het aantal dat wordt geselecteerd door de procedure.

De tweede mogelijkheid is het gebruik van modelselectie criteria. Deze vergelijken concurrerende modellen op basis van hun fit waarbij er een straf wordt toegekend voor de toevoeging van meer parameters. Op basis van deze vergelijking komt één van de concurrerende modellen als beste uit de bus. Het gebruik van modelselectie criteria vereist wel dat alle modellen die men in overweging wil nemen moeten worden geschat.

Voor het geval van de Almon vertragingsfunctie zijn de twee benaderingen met elkaar vergeleken door Teräsvirta en Mellin (1983). De bevindingen van dit onderzoek wijzen erop dat het gebruik van modelselectie criteria betere voorspelresultaten oplevert dan de sequentiële toetsprocedures. Voor zover bij ons bekend is er geen vergelijkbaar onderzoek gedaan voor de *exponential* Almon en *unconstrained* vertragingsfunctie. Daarom wordt de keus voor een van de twee alternatieven voornamelijk gebaseerd op de resultaten van het onderzoek van Teräsvirta en Mellin (1983) en opteren we voor het gebruik van modelselectie criteria.

Ook binnen de klasse van de modelselectie criteria zijn nog verschillende alternatieven mogelijk. Uit het onderzoek van Teräsvirta en Mellin (1983) komt naar voren dat het Bayesian Information Criterion (BIC) het beste presteert. Behalve deze wordt hier ook het veelgebruikte Akaike's Information Criterion (AIC) getest.

Op voorhand lijkt er weinig tot niets te pleiten voor de ad hoc methode om een specificatie te selecteren. Echter, zoals hierboven aangegeven, moeten voor het gebruik van modelselectiecriteria alle concurrerende specificaties worden geschat. Bij een model met één verklarende variabele kan dit al gaan om veertig verschillende specificaties. Wanneer er meerdere variabelen in het model zijn opgenomen, wordt het schatten van alle verschillende mogelijkheden al snel onhaalbaar.

In de bestaande literatuur wordt echter vaak gebruik gemaakt van modellen met slechts één verklarende variabelen. In dit geval lijkt het gebruik van modelselectie criteria te prefereren

boven de ad hoc methode. De resultaten van voorspelexercitie moeten duidelijk maken of dit inderdaad het geval is.

In deze paragraaf is aandacht besteed aan de problemen die gepaard gaan met het modelleren van data met een gemengde frequentie. Allereerst zijn er verschillende modellen gepresenteerd die kunnen worden gebruikt in deze ongewone situatie. Ten tweede is er gekeken naar de vertragingsfunctie. Tot slot is een aantal methodes om een passende specificatie te selecteren de revue gepasseerd. Steeds zijn er een of meerdere alternatieven gepresenteerd voor de keuzes die in de bestaande literatuur worden gemaakt. Er is duidelijk geworden dat er in de meeste gevallen geen eenduidige rangschikking kan worden gemaakt van de alternatieven op basis van louter theoretische argumenten. Daarom zal een experiment worden uitgevoerd om te bepalen of de verschillende opties geschikt zijn om BBP-groei te voorspellen. De volgende paragraaf bespreekt de opzet van het experiment en de data waarmee het wordt uitgevoerd.

4 Opzet en data

Om de verschillende alternatieven voor model, vertragingsfunctie en specificatieselectiemethode te vergelijken, wordt een recursieve out-of-sample voorspelexercitie uitgevoerd. Dit behelst een aantal stappen. Allereerst worden alle data tot een bepaald punt in het verleden verzameld. Vervolgens wordt vanuit het perspectief van dit punt BBP-groei voor de volgende kwartalen voorspeld aan de hand van deze data. Hierna worden de observaties voor het volgende kwartaal toegevoegd aan de dataset en worden wederom voorspellingen gedaan. Het proces wordt herhaald totdat het eind van de steekproef is bereikt. De gedane voorspellingen kunnen daarna worden vergeleken met de gerealiseerde BBP-groei.

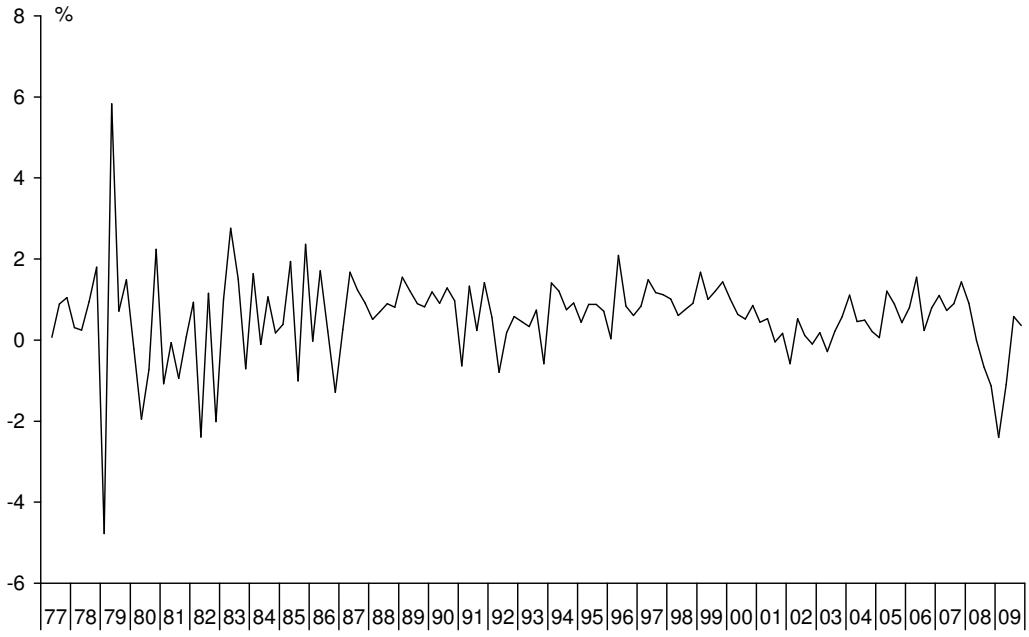
Het is belangrijk op te merken dat bij deze voorspelexercitie geen gebruik wordt gemaakt van kennis uit de toekomst, dat wil zeggen dat de voorspellingen worden uitgevoerd alsof we op dat punt in het verleden leven. De enige uitzondering vormt het gebruik van finale data: deze zou niet volledig beschikbaar zijn geweest op dat moment.

De voorspelexercitie wordt afzonderlijk uitgevoerd met elk van de leading indicators als enige verklarende variabele. Deze aanpak om verschillende alternatieven te vergelijken komt overeen met e.g. Marcellino et al. (2006), Clements en Galvào (2008) en Kuzin et al. (2009). Telkens wordt BBP-groei één, twee en drie kwartalen vooruit voorspeld.

Voor het experiment zijn data beschikbaar over BBP-groei en leading indicators vanaf het eerste kwartaal van 1977 tot en met het vierde kwartaal van 2009. Er wordt hier gebruik gemaakt van seizoensgecorrigeerde data. Voor BBP-groei is het eerste verschil[5] van de logaritmes van BBP genomen om een stationaire reeks te verkrijgen. Figuur 4.1 geeft de historische ontwikkeling van BBP-groei weer. Voor de leading indicators wordt een soortgelijke exercitie ondernomen als een *augmented* Dickey–Fuller toets de nulhypothese van stationairiteit verwerpt. In totaal worden er 27 potentiële leading indicators gebruikt; een overzicht van deze reeksen en de transformaties die benodigd zijn om ze stationair te maken staat afgedrukt in bijlage A.

[5] Met het eerste verschil wordt hier het verschil tussen twee aaneengesloten waargenomen waardes bedoeld. In het geval van BBP betekent dit dat van het logaritme van BBP het logaritme van BBP uit het kwartaal ervoor wordt afgetrokken.

Figuur 4.1 **Historische ontwikkeling van BBP-groei vanaf het tweede kwartaal van 1977 tot en met het vierde kwartaal van 2009**



De steekproefperiode voor de exercitie loopt van het tweede derde kwartaal van 2003 tot en met het vierde kwartaal van 2009. Dat betekent dat in eerste instantie alle data tot en met het tweede kwartaal van 2003 worden genomen om BBP-groei voor het derde en vierde kwartaal van 2003 en eerste kwartaal van 2004 te voorspellen. Daarna wordt de dataset uitgebreid met data uit het derde kwartaal van 2003 en worden het vierde kwartaal van 2003 en de eerste twee kwartalen van 2004 voorspeld. Dit gaat door tot en met het derde kwartaal van 2009, waar alleen nog BBP-groei voor het vierde kwartaal van 2009 wordt voorspeld. Twee en drie kwartalen vooruit voorspellen heeft nu geen toegevoegde waarde, aangezien deze voorspellingen niet meer kunnen worden vergeleken met de realisaties. Hieruit volgt dat de voorspelexercitie 26 voorspellingen oplevert voor één kwartaal vooruit, 25 voor twee kwartalen vooruit en 24 voor drie kwartalen vooruit.

De periode van 2003 tot en met 2009 is niet de meeste rustige uit de geschiedenis, aangezien de wereldwijde kredietcrisis hier in valt: één van de grootste crises van de afgelopen tachtig jaar. Aangezien we denken (en hopen) dat deze periode niet representatief is voor zowel het verleden als de toekomst, wordt er bij het vergelijken van de voorspellingen onderscheid gemaakt tussen de periode voor en tijdens de crisis. Er bestaat geen officiële startdatum voor de crisis, maar afgaande op figuur 4.1 kan worden gesteld dat deze vanaf het eerste halfjaar van 2008 zijn weerslag begon te krijgen op de reële economie. Economische groei kwam in het tweede kwartaal tot stilstand (de groei was vrijwel nul in dit kwartaal) zodat dit kwartaal de grens zal vormen tussen de twee periodes. Hierdoor vallen, afhankelijk van de voorspelhorizon,

zeventien tot negentien voorspellingen in de periode voor de crisis en zeven in de tijdens crisisperiode.

Om enige intuïtie te verschaffen over de voorspelprestatie van de verschillende alternatieven, zullen deze worden vergeleken met een naïeve benchmark. Hier worden twee verschillende kandidaten onderzocht om dienst te doen als benchmark. De eerste is het steekproefgemiddelde, de tweede de voorspellingen die worden gegenereerd door een simpel autoregressief (AR) proces. In een AR-proces wordt toekomstige BBP-groei verklaard aan de hand van uitsluitend vertraagde waarden van zichzelf en een constante. Ook hier worden de twee modelselectiecriteria AIC en BIC gebruikt om te bepalen hoeveel vertragingen van BBP-groei moeten worden opgenomen.

De resultaten van de voorspelexercitie met naïeve voorspelmethoden staan weergegeven in tabel 4.1. In deze tabel worden de verschillende voorspellers met elkaar vergeleken op basis van hun gemiddelde gekwadrateerde voorspelfout (MSFE, mean square forecasting error), logischerwijs gedefinieerd als het gemiddelde van het gekwadrateerde verschil tussen voorspelling en realisatie. De tabel laat voor elk van de naïeve voorspellers de voorspelprestatie zien voor drie voorspelhorizonnen, aangeduid met de letter h, en de twee periodes.

Tabel 4.1 Voorspelprestaties van naïeve voorspellers[a]

	Voor crisis			Tijdens crisis		
	Steekproefgem.	AR: AIC	AR: BIC	Steekproefgem.	AR: AIC	AR: BIC
$h = 1$	0,188	0,192	0,192	2,42	2,21	2,13
$h = 2$	0,192	0,188	0,203	2,45	2,93	2,78
$h = 3$	0,202	0,198	0,208	2,46	2,69	2,79

[a] In deze tabel staan de resultaten van drie naïeve voorspellers: het steekproefgemiddelde en een AR-proces waarvan de orde is bepaald door zowel Akaike's Information Criterion en Schwartz Information Criterion. De afgedrukte getallen zijn de gemiddelde gekwadrateerde voorspelfoutem in procenten. In de tabel wordt onderscheid gemaakt tussen de periode voor (2003Q3-2008Q1) en tijdens de crisis (2008Q2-2009Q4) en drie voorspelhorizonnen: één, twee en drie kwartalen vooruit.

Uit de tabel kan een aantal conclusies worden getrokken. Ten eerste worden de voorspellingen in zijn algemeenheid minder goed naarmate de voorspelhorizon toeneemt, zoals verwacht kon worden. Een tweede waarneming is dat de RMSE'en van de verschillende naïeve voorspellers weinig voor elkaar onderdoen. Dit kan erop duiden dat de vertraagde waarden van BBP-groei weinig voorspellende waarden bezitten voor toekomstige BBP-groei, aangezien het AR-proces gelijk is aan het steekproefgemiddelde als alle AR-coëfficiënten nul zijn. Ten derde is de RMSE meer dan tien keer zo groot tijdens de crisis dan ervoor. Dit benadrukt het afwijkende karakter van de crisisperiode en laat zien dat de crisis moeilijk te voorspellen was door deze naïeve voorspelmethoden.

In het vervolg van dit document zullen alle voorspelresultaten worden uitgedrukt in termen van hun relatieve RMSE. Deze is gedefinieerd als de RMSE van de voorspelling gedeeld door de RMSE van de best presterende naïeve benchmark voor de corresponderende periode en voorspelhorizon. Hieruit volgt dat een relatieve RMSE van groter dan 1 impliceert dat de betreffende voorspeller het minder doet dan het naïeve alternatief, terwijl een relatieve RMSE kleiner dan 1 betekent dat de voorspeller een verbetering is ten opzichte van de naïeve benchmark.

In deze paragraaf zijn de data en opzet van de voorspelexercitie besproken, welke als doel heeft om de voorspelprestaties van verschillende voorspellers met elkaar te vergelijken. Hiermee is een stevige basis gelegd om tot een conclusie te komen welke combinatie van model, vertragingsfunctie en specificatieselecteur te prefereren valt om BBP-groei te voorspellen. In de volgende paragraaf worden de resultaten van de voorspelexercitie gepresenteerd en besproken.

5 Resultaten

In paragraaf 4 is een aantal verschillende modellen, vertragingsfuncties en modelselectiemethoden besproken. Uit de theoretische overwegingen bleek dat alle alternatieven zowel voor- als nadelen hebben, zodat het a priori niet mogelijk is te bepalen welk model het meest geschikt is om BBP-groei te voorspellen. Daarom is een *out-of-sample* voorspelexercitie uitgevoerd om de modellen te beoordelen op hun voorspelkwaliteiten. De exacte opzet van dit experiment en de data die hierbij zijn gebruikt, zijn reeds aan de orde gekomen in de vorige paragraaf. In deze paragraaf worden de resultaten besproken.

De resultaten van het vergelijken van de modellen staan weergegeven in tabel 5.1. De tabel kent dezelfde opzet als tabel 4.1: er is wederom onderscheid gemaakt tussen twee periodes en drie voorspelhorizonnen. De getallen in de tabel zijn de gemiddelde gekwadrateerde voorspelfouten, relatief ten opzichte van de best presterende naïeve voorspeller. De resultaten in de tabel zijn gemiddeld over alle 27 leading indicators, drie vertragingsfuncties en de twee modelselectie criteria.

Tabel 5.1 Vergelijking van voorspellingen van verschillende modellen[a]

	Voor crisis				Tijdens crisis			
	MIDAS	AR-MIDAS	ADL	DL	MIDAS	AR-MIDAS	ADL	DL
$h = 1$	1,06	1,14	1,47	1,06	0,94	1,09	1,82	0,98
$h = 2$	1,11	1,10	1,32	1,08	0,97	1,07	0,89	1,01
$h = 3$	1,12	1,18	1,16	1,08	1,01	1,06	1,09	1,03

[a] Deze tabel geeft de relatieve RMSE weer van alle vier modellen over drie voorspelhorizonnen en twee periodes. De nummers zijn het gemiddelde over alle leading indicators, vertragingsfuncties en modelselectie criteria.

Tabel 5.1 laat zien dat de modellen zonder autoregressief component beter presteren dan de modellen met autoregressief component: het MIDAS en het DL-model leveren voorspellingen op die dichter bij de gerealiseerde BBP-groei zitten dan het AR-MIDAS en ADL-model. Dit geldt zowel voor als tijdens de crisis, een enkele uitzondering daargelaten. Als het MIDAS en DL-model onderling worden vergeleken is duidelijk dat deze zeer aan elkaar gewaagd zijn. Voor de crisis presteert het DL-model iets beter, terwijl MIDAS tijdens de crisis betere voorspellingen oplevert. De verschillen zijn echter klein.

De resultaten van tabel 5.1 duiden erop dat het opnemen van vertraagde waardes van BBP-groei geen toegevoegde waarde heeft. Dit bleek ook al uit de evaluatie van de naïeve voorspellers: het steekproefgemiddelde als voorspelling presteert ongeveer even goed als een autoregressief proces. Dit alles wijst erop dat in gerealiseerde BBP-groei geen informatie over toekomstige BBP-groei zit verscholen.

De resultaten van het vergelijken van de vertragingsfunctie staan weergegeven in tabel 5.2. De tabel heeft dezelfde opzet als tabel 5.1, zij het dat hier is uitgemiddeld over alle leading indicators, modellen en modelselectie criteria.

Tabel 5.2 Vergelijking van voorspellingen van verschillende vertragingsfuncties[a]

	Voor crisis			Tijdens crisis		
	Unconstrained	Almon	Exp. Almon	Unconstrained	Almon	Exp. Almon
$h = 1$	1,29	1,14	1,13	1,31	1,22	1,09
$h = 2$	1,22	1,13	1,10	0,99	1,00	0,96
$h = 3$	1,18	1,12	1,10	1,03	1,07	1,04

[a] Deze tabel geeft de relatieve RMSE weer van alle drie de vertragingsfuncties over drie voorspelhorizonnen en twee periodes. De nummers zijn het gemiddelde over alle leading indicators, modellen en modelselectie criteria.

Uit tabel 5.2 kan worden afgeleid dat de functies die enige structuur meegeven aan de gewichten van de vertragingen beter presteren dan de functie die dat niet doet: vooral voor de crisis leveren de Almon en exponential Almon vertragingsfunctie betere resultaten op dan de *unconstrained* functie. Tijdens de crisis is dit minder duidelijk. Er moet echter meer waarde worden gehecht aan de resultaten van voor de crisis. Niet alleen zitten er meer voorspellingen in deze periode, ook is hij meer representatief voor het verleden en (wellicht) de toekomst.

Wanneer de twee spaarzame vertragingsfuncties worden vergeleken, komt de *exponential* Almon functie iets beter voor de dag. Voor de crisis is het verschil klein, maar tijdens de crisis neemt dit toe. De verschillen blijven echter dusdanig klein dat deze resultaten geen doorslaggevend bewijs opleveren welke van de twee spaarzame functies moet worden geprefereerd.

In tabel 5.3 zijn de verschillende methodes om een specificatie te selecteren met elkaar vergeleken. Ook deze tabel volgt de inmiddels bekende opzet. In dit geval zijn de gegeven relatieve RMSE de gemiddelden over alle leading indicators, modellen en vertragingsfuncties.

Uit tabel 5.3 kan worden afgeleid dat het gebruik van modelselectie criteria voor de crisis resulteerde in betere voorspellingen dan de ad hoc methode. Er kan in deze periode een duidelijk rangschikking worden gemaakt: het Bayesian Information Criterion presteert het best, gevolgd door Akaike's Information Criterion, terwijl de ad hoc methode de minste resultaten oplevert. Tijdens de crisis zijn deze rollen precies omgedraaid: nu resulteert de ad hoc methode in de beste resultaten en is BIC het minst. De verschillen zijn hier echter zeer klein en, zoals hierboven is toegelicht, moet bovendien meer waarde worden gehecht aan de resultaten van voor de crisis. Hieruit volgt dat de resultaten erop wijzen dat het gebruik van modelselectie criteria te prefereren valt boven het maken van ad hoc keuzes.

Tabel 5.3 Vergelijking van voorspellingen van verschillende specificatieselecteurs[a]

	Voor crisis			Tijdens crisis		
	AIC	BIC	Ad hoc	AIC	BIC	Ad hoc
$h = 1$	1,21	1,16	1,26	1,17	1,24	1,22
$h = 2$	1,18	1,12	1,15	0,98	0,99	0,97
$h = 3$	1,16	1,11	1,19	1,05	1,05	1,04

[a] Deze tabel geeft de relatieve RMSE weer van alle drie de specificatieselecteurs over drie voorspelhorizonnen en twee periodes. De nummers zijn het gemiddelde over alle leading indicators, modellen en vertragingsfuncties.

In alle gevallen zijn controles uitgevoerd om na te gaan of de resultaten wel robuust zijn. Hieruit is gebleken dat het niet meenemen van de slechtst presterende leading indicators, modellen, vertragingsfuncties of specificatieselectiemethoden de resultaten kwalitatief niet veranderen.

In deze paragraaf zijn de resultaten aan bod gekomen van het vergelijken van de verschillende alternatieven wat betreft model, vertragingsfunctie en specificatieselectiemethode. Er is gebleken dat de modellen zonder autoregressief component betere voorspellingen voor BBP-groei opleveren dan de modellen met zo'n component. Ook impliceren de resultaten dat vertragingsfuncties die structuur opleggen aan de gewichtenverdeling moet worden geprefereerd boven de functie die dat niet doet. Ten slotte kan het gebruik van modelselectie criteria tot betere resultaten leiden dan als de specificatie op een ad hoc wijze wordt gekozen.

Uit de tabellen die in deze paragraaf zijn gepresenteerd kan echter ook worden afgeleid dat de voorspelprestaties gemiddeld genomen niet erg indrukwekkend zijn. De meeste gemiddelde relatieve RMSE'en liggen rond of boven de 1, wat erop wijst dat de voorspellingen niet beter of zelfs minder zijn dan die van de naïeve benchmarks. Hoewel hierbij kan worden aangetekend dat in deze gemiddeldes ook de slecht presterende leading indicators, modellen, vertragingsfuncties en specificatieselectiemethoden zijn meegenomen, zijn de voorspelprestaties hoe dan ook matig te noemen. Hoewel het verwijderen van de slechtst presterende alternatieven zou leiden tot betere voorspelresultaten, zou dit tegen het buiten-de-steekproef karakter van het ondernomen experiment ingaan. Bovendien is er ook een andere mogelijkheid om te proberen de voorspelprestatie te verbeteren, namelijk poolen. De volgende paragraaf zal hier dieper op ingaan.

6 Poolen

In de vorige paragraaf is naar voren gekomen dat de voorspellingen die worden gedaan door modellen waarin één leading indicator wordt gebruikt, gemiddeld genomen matige voorspelresultaten opleveren. Een verklaring hiervoor zou kunnen zijn dat er geen leading indicators zijn die vooruitlopen op alle facetten van de economie. Het samenbrengen van de informatie die leading indicators bevatten zou de voorspelprestatie kunnen verbeteren. Als elk van de afzonderlijke leading indicators vooruitloopt op een deel van de economie, kan uit hun gecombineerde informatie wellicht een betere voorspelling worden gedestilleerd dan de geïsoleerde voorspellingen.

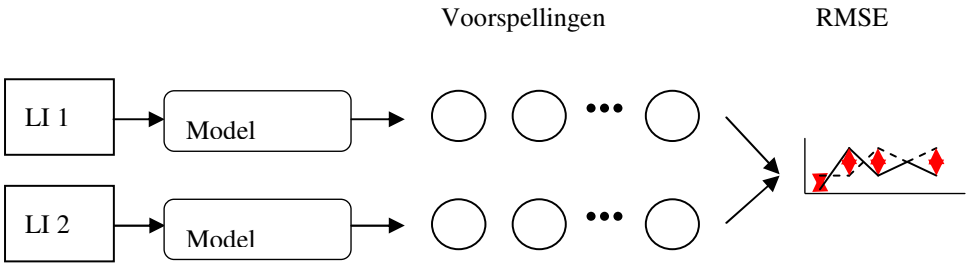
Een methode die vaak wordt gebruikt om informatie uit verschillende voorspellingen te combineren is poolen, e.g. het nemen van (gewogen) gemiddelde van alle voorspellingen. Er is veel empirisch bewijs dat deze aanpak goede resultaten oplevert, zie bijvoorbeeld Timmermann (2006) voor een overzicht. Er is echter ook een geheel andere aanpak om de informatie in de leading indicators te combineren, namelijk door meerdere leading indicators op te nemen in één model.

Clements en Galvào (2009) gebruiken deze twee benaderingswijzen door twee methoden te onderzoeken om informatie te combineren. De eerste manier is poolen over de voorspellingen van de afzonderlijke leading indicators. De tweede optie is het opnemen van meerdere leading indicators in één model en alleen met dit model te voorspellen. Zij bevinden dat de eerste methode betere voorspelresultaten oplevert: de resultaten van poolen over meerdere modellen met een enkele leading indicator zijn beter dan de resultaten van het gebruik van een enkel model waarin alle leading indicators zijn opgenomen. De twee benaderingswijzen worden schematisch weergegeven in de figuur 6.1a respectievelijk 6.1b.

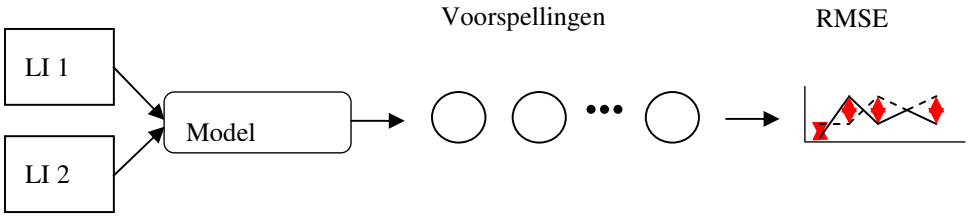
In dit onderzoek wordt bekeken of dit resultaat kan worden bevestigd voor de huidige dataset. Ook wordt er een derde mogelijkheid onder de loep genomen: poolen over de voorspellingen van meerdere modellen met meerdere indicatoren. Deze methode zou beter kunnen werken omdat hij het beste van de twee werelden combineert: de robuustheid van poolen met het informatiegehalte van modellen met meerdere leading indicators.

Figuur 6.1 Schematische weergave van verschillende methoden om tot een voorspelling te komen[a]

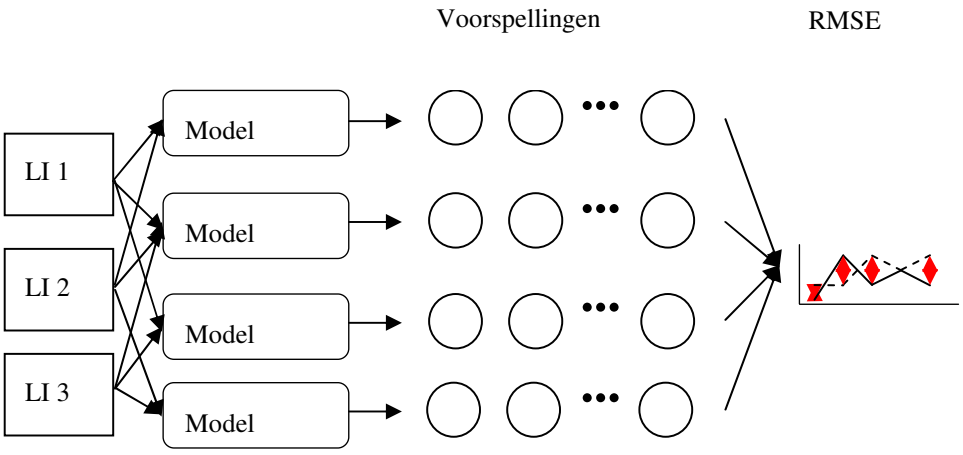
6.1 a poolen van modellen met 1 variabele



6.1 b alle leading indicators in 1 model



6.1 c poolen van modellen met meerdere leading indicators



[a] De bovenste subfiguur geeft schematisch weer hoe telkens één leading indicator wordt gebruikt om een voorspelling te construeren, waarna wordt gepoold. In het middelste subfiguur worden alle leading indicators opgenomen in een enkel model. Het onderste subfiguur laat zien hoe er meerdere combinaties van meerdere leading indicators kunnen worden gemaakt om tot meerdere voorspellingen te komen. Over deze voorspellingen wordt wederom gepoold.

Figuur 6.1 geeft de drie benaderingswijzen schematisch weer. Wel dient er te worden opgemerkt dat het aantal leading indicators in de figuur is gereduceerd tot twee of drie en er steeds maar één model wordt gebruikt om voorspellingen te verkrijgen. Hier worden er meer dan drie leading indicators meegenomen en worden voorspellingen verkregen uit meerdere combinaties van model, vertragingsfunctie en modelselectie criterium.

Er wordt per combinatie van model, vertragingsfunctie en modelselectie criterium bekeken welke van de drie aanpakken het beste werkt. Aangezien er 24 van zulke combinaties zijn, zou dit een betrouwbaar beeld moeten opleveren van welke methode het beste is. Tot slot wordt een laatste alternatief beschouwd: poolen over verschillende specificaties. Hierbij wordt gekeken of poolen over de 24 combinaties van model, vertragingsfunctie en modelselectie criterium beter presteert dan de afzonderlijke alternatieven.

Bij de voorspelexercitie van paragraaf 6 is gebruik gemaakt van 27 leading indicators. Dit aantal is te groot om in een enkel model op te nemen. Bovendien kan uit de resultaten worden afgeleid dat enkele van deze reeksen weinig of geen voorspellende capaciteiten hebben voor BBP-groei. Derhalve dient een selectie te worden gemaakt uit deze 27 leading indicators. Het is een mogelijkheid om deze selectie te baseren op de prestaties van de leading indicators in de één-op-één exercitie van de vorige paragraaf. Dit heeft echter een belangrijk nadeel: het is namelijk zeer waarschijnlijk dat de reeksen die geïsoleerd goed presteren sterk gecorreleerd zijn. Het gebruik van sterk gecorreleerde reeksen in de multivariate analyse kan leiden tot collineairiteitsproblemen als onnauwkeurige schattingen. Bovendien is het waarschijnlijk dat de reeksen veel dezelfde informatie bevatten. De reeksen combineren levert in dat geval geen of weinig winst op.

Hieruit volgt dat op zoek moet worden gegaan naar reeksen die voorspelkracht hebben voor BBP-groei en complementaire informatie bevatten. De selectie van leading indicators voor de multivariate analyse is op drie aspecten gebaseerd. Het eerste is de voorspelprestatie van de reeks in de voorspelexercitie van de vorige paragraaf. Het tweede is dat de onderlinge correlatie van de reeksen, welke laag moet zijn. De derde eis heeft betrekking op de soort reeksen. De leading indicators kennen verschillende origine, zoals financiële reeksen en de resultaten van enquêtes. Om robuustere voorspellingen te verkrijgen, worden hier reeksen uit verschillende bronnen geselecteerd.

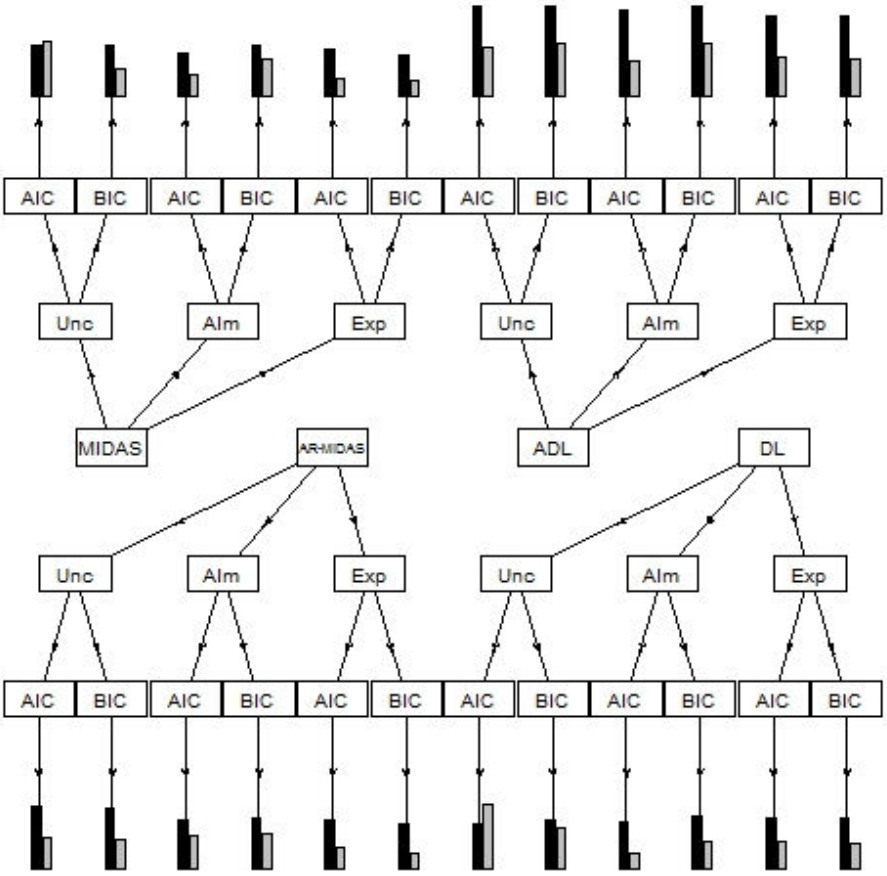
Dit leidt tot een selectie van zes leading indicators: bedrijvigheid, beurskoers, koopbereidheid, geldhoeveelheid, bouwvergunningen en bedrijvigheid in de woningbouw. Een overzicht van de prestaties, het soort bron en de correlaties staat afgedrukt in bijlage B.

Om de verschillende benaderingswijzen te vergelijken wordt dezelfde out-of-sample voorspelexercitie uitgevoerd als in paragraaf 5 met één verschil: er wordt niet langer

onderscheid gemaakt tussen de periode voor en tijdens de crisis. De reden hiervoor is dat de techniek van het poolen relatief juist geschikt is voor plotselinge afwisselingen.

Als een eerste stap om de verschillende benaderingen te vergelijken, worden de verschillende combinaties van model, vertragingsfunctie en selectiemethode geschat met alle zes geselecteerde leading indicators. Deze schattingen worden gebruikt om voorspellingen te verkrijgen, welke vergeleken kunnen worden met de gerealiseerde BBP-groei om de relatieve gemiddelde gekwadrateerde voorspelfout te berekenen. Deze statistieken kunnen vervolgens worden afgezet tegen de resultaten van het poolen over voorspellingen gedaan met één leading indicator per keer. Hierbij worden uitsluitend de resultaten gebruikt van de geselecteerde reeksen. Beide methoden leveren voor elk van de 24 combinaties van model, vertragingsfunctie en selectiemethode een RMSE op, zoals schematisch is weergegeven in figuur 6.2.

Figuur 6.2 Schematische weergave vergelijking poolen[a]



[a] Voor de twee aanpakken weergegeven in figuur 6.2 a en 6.2 b wordt voor alle combinaties van model, vertragingsfunctie en selectiemethode de RMSE berekend. De zwarte balk geeft de RMSE weer van het gebruik van één leading indicator, de grijze wordt verkregen wanneer alle leading indicators in een enkele vergelijking worden opgenomen, beide voor één kwartaal vooruit.

Figuur 6.2 laat zien hoe de twee aanpakken worden vergeleken. Elke combinatie van model, vertragingsfunctie en selectiemethode levert aparte voorspellingen op. In de vorige paragraaf is dit uitsluitend gedaan met één leading indicator per keer. Hier wordt ook de mogelijkheid meegenomen om zes leading indicators in één keer op te nemen. De resultaten van deze aanpak worden vergeleken met poolen over voorspellingen gedaan met behulp van één leading indicator in figuur 6.2. In deze figuur worden alleen de resultaten voor één kwartaal vooruit weergegeven. Hier geeft de zwarte balk de relatieve RMSE van laatstgenoemde aanpak aan, terwijl de grijze balk de relatieve RMSE van het gebruik van het multivariate model laat zien.

Uit figuur 6.2 kan worden afgeleid dat het gebruik van alle leading indicators in één model beter werkt dan poolen over de één-op-één modellen. In een grote meerderheid van de gevallen is de grijze balk namelijk korter dan de zwarte balk, hetgeen betekent dat de multivariate aanpak leidt tot een kleinere RMSE dan de poolaanpak. De figuur zegt echter niks over de absolute prestatie, i.e. hoe de aanpakken zich verhouden ten opzichte van de naïeve benchmark, aangezien de schaal ontbreekt. Hiervoor biedt tabel 6.1 uitkomst.

Tabel 6.1 Vergelijking van de relatieve RMSEen van poolen en het meerdere regressoren model[a]

$h = 1$		Unconstrained		Almon		Exponential Almon	
		AIC	BIC	AIC	BIC	AIC	BIC
MIDAS	Pool	0,84	0,84	0,73	0,84	0,78	0,70
	MRM	0,91	0,46	0,36	0,63	0,28	0,26
AR-MIDAS	Pool	1,09	1,06	0,84	0,89	0,85	0,77
	MRM	0,54	0,50	0,58	0,61	0,39	0,28
ALD	Pool	1,66	1,84	1,45	1,71	1,36	1,37
	MRM	0,83	0,87	0,60	0,88	0,64	0,63
DL	Pool	0,77	0,85	0,81	0,91	0,87	0,87
	MRM	1,10	0,72	0,29	0,48	0,46	0,43

[a] De tabel laat per combinatie van model, vertragingsfunctie en selectiemethode de relatieve RMSE zien voor de twee aanpakken. De eerste, hier aangeduid met "pool", is het poolen over de voorspellingen gemaakt door het gebruik van één leading indicator per keer. De tweede, waarvoor de afkorting MRM wordt gebruikt, benut alle leading indicators in een enkele vergelijking.

Tabel 6.1 laat alle relatieve RMSE'en zien van de verschillende combinaties van model, vertragingsfunctie en selectiemethode bij het gebruik van één model met alle leading indicatoren en het poolen over de voorspellingen gemaakt met één leading indicator per keer. Ook in tabel 6.1 staan uitsluitend de resultaten voor één kwartaal vooruit afgedrukt.

Uit tabel 6.1 kan worden afgeleid dat er behoorlijke spreiding zit in de resultaten. Het gebruik van het MIDAS-model met exponential Almon vertragingsfunctie, waarin de specificatie is geselecteerd door BIC, levert voorspellingen op die het in termen van RMSE 74% beter doen dan de naïeve benchmark wanneer alle leading indicators tegelijk worden gebruikt. Poolen over de voorspellingen gedaan door het ADL-model met *unconstrained*

vertragingsfunctie en AIC waarin telkens één leading indicator wordt gebruikt, leidt echter tot voorspellingen die 84 procent slechter presteren dan de naïeve benchmark. Als het poolen wordt vergeleken met het gebruik van een enkel model, kan uit tabel 6.2 worden geconcludeerd dat laatstgenoemde substantieel betere resultaten oplevert.

Figuur 6.2 en tabel 6.1 bieden echter geen uitsluitsel voor de prestaties wanneer er twee en drie kwartalen vooruit wordt voorspeld. Tabel 6.2 geeft een aantal samenvattende statistieken weer voor alle drie de voorspelhorizonnen.

Tabel 6.2 Poolen over voorspellingen van modellen met één leading indicator vergeleken met de voorspellingen van een enkel model met alle leading indicators[a]

Voorspelhorizon	Poolen beter (%)	Gemiddeld verschil
$h = 1$	8,3	0,46
$h = 2$	8,3	0,18
$h = 3$	12,5	0,02

[a] De tabel geeft het percentage van de combinaties van model, vertragingsfunctie en selectiemethode weer waarin poolen het beter doet dan het combineren van de informatie in één model, alsmede het gemiddelde verschil tussen de twee benaderingen.

Tabel 6.2 geeft per voorspelhorizon twee statistieken weer. De eerste is het percentage van de 24 combinaties van model, vertragingsfunctie en selectiemethode, waarin poolen over de zes voorspellingen van de afzonderlijke leading indicators het beter doet dan het enkele model dat alle leading indicators gebruikt. De tweede statistiek is het gemiddelde verschil van de twee methode, berekend als de gemiddelde over de relatieve RMSEen van het poolen minus de relatieve RMSEen van de multivariate benadering.

Uit tabel 6.2 blijkt dat het in dit geval beter is om de informatie in de leading indicators te benutten door ze allemaal op te nemen in een enkel model. Voor elke voorspelhorizon geldt dat in meer dan 85% van de combinaties van model, vertragingsfunctie en selectiemethode deze aanpak betere resultaten oplevert dan het poolen over modellen die één leading indicator gebruiken. Het gemiddelde verschil is ook telkens in het voordeel van de multivariate aanpak; in termen van relatieve RMSE kan dit verschil oplopen tot 0,46.

Dit resultaat staat haaks op de bevindingen van Clements en Galvào (2006 en 2008). Uit dit onderzoek blijkt namelijk dat poolen over de voorspellingen van modellen met één verklarende variabele betere voorspelresultaten oplevert dan het gebruik van één model waarin alle verklarende variabelen zijn opgenomen. Blijkbaar is dit resultaat niet robuust voor de exacte toepassing. Wel blijkt uit nadere inspectie dat de resultaten van poolen wat robuuster zijn: de variantie van de relatieve RMSE'en verkregen door poolen is lager dan die verkregen uit de multivariate modellen.

Zoals aangekondigd worden deze twee methodes vergeleken met een derde, nieuwe optie: poolen over de voorspellingen van meerdere multivariate modellen. Voor een eerlijke vergelijking mogen slechts de zes geselecteerde leading indicators worden gebruikt: enig verschil in prestatie kan dan niet worden verklaard uit het feit dat meer informatie is benut. Om toch meerdere sets van leading indicators te creëren, worden combinaties van de zes reeksen gemaakt. Deze combinaties kunnen tussen de twee en zes reeksen bevatten. In totaal zijn er 57 van zulke combinaties. Omdat het verkrijgen van voorspellingen voor al deze 57 combinaties nogal bewerkelijk is, wordt er hier voor gekozen om twee reeksen vast te zetten: alleen combinaties waarin deze leading indicators voorkomen worden meegenomen. Dit brengt het aantal combinaties terug tot zestien. De twee vastgezette variabelen zijn beurskoers en bedrijvigheid.

De resultaten van de vergelijking staan weergegeven in tabel 6.3. De tabel maakt onderscheid tussen de drie aanpakken van figuur 6.1. Kolom twee en drie geven de resultaten weer als één leading indicator per keer wordt gebruikt. De vierde en vijfde kolom doet hetzelfde als er niet wordt gepoold, maar alle geselecteerde leading indicators worden opgenomen in een enkel model. Tot slot staan de resultaten van poolen over de voorspelresultaten van verschillende combinaties van meerdere leading indicators in de zesde en zevende kolom.

Tabel 6.3 Vergelijking van verschillende aanpakken om informatie te combineren[a]

	Één leading indicator		Alle leading indicators		Meerdere leading indicators	
	Gem.	Gepoold	Gem.	Gepoold	Gem.	Gepoold
$h = 1$	1,03	0,97	0,57	0,42	0,52	0,34
$h = 2$	0,90	0,87	0,72	0,52	0,71	0,61
$h = 3$	0,98	0,97	0,96	1,15	0,94	0,83

[a] Deze tabel vergelijkt drie aanpakken: poolen over modellen die één leading indicator per keer gebruiken, een enkel model waarin alle leading indicators zijn opgenomen en het poolen over de voorspelresultaten voortkomend uit het gebruik van verschillende sets van meerdere leading indicators. Per aanpak worden twee statistieken weergegeven. De linker is de prestatie van de aanpak gemiddeld genomen over alle combinaties van model, vertragingsfunctie en selectiemethode. Rechts staat de prestatie als wordt gepoold over al deze combinaties. De prestatie wordt steeds weergegeven in termen van de relatieve RMSE.

Per aanpak staan twee getallen weergegeven. Het linker getal is de gemiddelde relatieve RMSE over alle 24 combinaties van model, vertragingsfunctie en selectiemethode. Het rechter getal is de relatieve RMSE als gepoold wordt over alle 24 combinaties. Om de verschillende benaderingswijzen beter duidelijk te maken, staan deze schematisch toegelicht in de bijlage. Van de tweede tot zevende kolom zijn de uitkomsten het gevolg van de aanpak die staat weergegeven in respectievelijk figuur C.2, C.3, C.4, C.5, C.7 en C.8 van bijlage C. De schematische weergave in bijlage C verschilt van figuur 6.1 in dat in de bijlage ook de verschillende combinaties van model, vertragingsfunctie en selectiemethode zijn meegenomen. Een grafische weergaven van de resultaten kan worden teruggevonden in bijlage D.

Uit tabel 6.3 kunnen twee conclusies worden getrokken. Allereerst is poolen over de voorspellingen verkregen door het gebruik van meerdere combinaties van meerdere leading indicators beter dan de twee alternatieven zijnde poolen over voorspellingen van modellen die één leading indicator gebruiken en het gebruik van één model waarin alle geselecteerde leading indicators zijn opgenomen. Ten tweede kan poolen over de verschillende combinaties van model, vertragingsfunctie en selectiemethode de voorspelnauwkeurigheid nog verder verbeteren. Deze twee conclusies impliceren dat de methode die schematisch wordt weergegeven in figuur C.8 het beste perspectief biedt op goede voorspelresultaten.

In deze paragraaf is een aantal methoden om tot een puntvoorspelling te komen met elkaar vergeleken. Allereerst is er gekeken naar het poolen over meerdere voorspellingen die voortkomen uit het gebruik van modellen waarin telkens één leading indicator is gebruikt. De tweede methode behelsde het gebruik van een enkel model waarin alle leading indicators zijn opgenomen. Voor de derde methode wordt gepoold over voorspellingen die voortkomen uit het gebruik van verschillende sets van leading indicators. Als vierde wordt gekeken of de prestaties verbeteren wanneer er wordt gepoold over de voorspellingen gemaakt door verschillende combinaties van model, vertragingsfunctie en selectiemethode.

De resultaten laten zien dat het gebruik van multivariate modellen de voorspelresultaten kunnen verbeteren. Hetzelfde geldt voor het poolen over meerdere combinaties van model, vertragingsfunctie en selectiemethode. Hieruit volgt dat de methode waarbij wordt gepoold over de voorspellingen van verschillende combinaties van sets van leading indicators, modellen, vertragingsfuncties en selectiemethoden de beste resultaten oplevert. Deze methode behelst een aantal stappen. Allereerst moeten meerdere combinaties van leading indicators worden samengesteld. Vervolgens moet elk van deze sets worden gebruikt om de 24 verschillende combinaties van model te schatten, waarmee voorspellingen kunnen worden gemaakt. Tot slot levert poolen over alle voorspellingen de gewenste puntvoorspeller op.

7 Conclusie

Dit onderzoek richt zich op het voorspellen van BBP-groei met behulp van leading indicators. Deze aanpak is complementair aan de methode om economische ontwikkeling te voorspellen door middel van een grootschalig macro-economisch model als SAFFIER, aangezien zowel de onderbouwing als de gebruikte informatie verschilt tussen de twee benaderingswijzen. De twee methoden zijn derhalve goed geschikt om naast elkaar te worden gebruikt.

Het voorspellen van BBP-groei met behulp van leading indicators kent een aantal complicaties. Een daarvan is de gemengde beschikbaarheid van data: BBP wordt op kwartaalbasis gepubliceerd, terwijl leading indicators meestal op maandbasis of frequenter beschikbaar zijn. Het onderzoek concentreert zich op dit probleem. Een voorbeeld van een model dat wel om kan gaan met gemengde beschikbaarheid van data is het MIDAS-model. Hoewel dit model pas recent is geïntroduceerd in de literatuur, geniet het al enige populariteit. In dit onderzoek is de conventionele aanpak met behulp van het MIDAS-model zoals toegepast in de bestaande literatuur onder de loep genomen.

De bestaande methodologie is geëvalueerd langs drie dimensies. Het eerste is het model zelf: als alternatief voor het MIDAS-model wordt het *autoregressive distributed lag model* voorgesteld. Hoewel dit model bekend is van de situatie met gelijke beschikbaarheid van data, is het in de hier voorgestelde vorm nog niet eerder voor een toepassing met gemengde data gebruikt. De tweede dimensie betreft de vertragingsfunctie: als alternatief voor de conventionele *exponential* Almon functie wordt gekeken naar twee andere functies. Als derde dimensie wordt gekeken naar de methode om een geschikte specificatie te selecteren. De bestaande literatuur doet dit ad hoc; hier wordt het gebruik van modelselectie criteria voorgesteld.

Een theoretische vergelijking van de alternatieven laat zien dat er in alle gevallen zowel voor- als nadelen zijn aan te wijzen. Het is daarom niet mogelijk om a priori een rangschikking te maken uitsluitend gebaseerd op theoretische argumenten. Om meer zicht te verschaffen op welke van de alternatieven het beste geschikt zijn om BBP-groei te voorspellen met leading indicators, is een *out-of-sample* voorspelexercitie uitgevoerd. Hieruit kan een aantal conclusies worden getrokken. Allereerst leveren de modellen zonder autoregressief component beter voorspellingen op dan de modellen die wel vertraagde waardes van BBP-groei opnemen als verklarende variabele. Een tweede conclusie is dat de vertragingsfuncties die al enige structuur meegeven aan de gewichten van de vertragingen beter presteren dan de functie die de gewichten helemaal vrij laat. Ten derde blijkt het gebruik van modelselectie criteria de voorspelprestatie te verbeteren.

Het gebruik van meerdere (sets van) leading indicators, modellen, vertragingsfuncties en selectiemethoden resulteert in een hele verzameling van voorspellingen. Echter, uiteindelijk

moet dit terug worden gebracht tot een enkele voorspelling. Een instrument dat hierbij gebruikt kan worden is poolen, e.g. het nemen van het gemiddelde over een hele verzameling van voorspellingen als unieke voorspelling. In dit onderzoek zijn meerdere manieren om zo'n verzameling van voorspellingen te construeren met elkaar vergeleken. Het eerste is het gebruik van modellen die steeds één leading indicator gebruiken om BBP-groei te voorspellen. De tweede aanpak is het gebruik van een enkel model waarin alle leading indicators zijn opgenomen. Als derde is gekeken naar het gebruik van verschillende sets van meerdere leading indicators. In het eerste en derde geval is er sprake van een verzameling van voorspellingen en wordt poolen toegepast om tot een enkele voorspelling te komen. Dit wordt gedaan voor alle 24 combinaties van model, vertragingsfunctie en modelselectie criterium die hier zijn beschouwd. Tot slot is gekeken of de voorspelprestatie nog verder verbeterd als er wordt gepoold over de voorspellingen van deze 24 combinaties.

De resultaten laten zien dat voorspellingen gemaakt door modellen die één leading indicator gebruiken, minder goed presteren dan modellen die meerdere leading indicators gebruiken. Ook blijkt poolen een goede methode om tot één voorspelling te komen: zowel poolen over de voorspellingen van meerdere modellen met meerdere leading indicators als poolen over alle combinaties van model, vertragingsfunctie en modelselectie criterium presteert beter dan het gemiddelde alternatief.

Er zijn meerdere mogelijkheden voor vervolgonderzoek. Allereerst is hier tot maximaal drie kwartalen in de toekomst voorspeld. Het is interessant om te kijken hoe een voorspelling over het hele komende jaar zich verhoud tot andere methodes. Een jaarvoorspelling kan worden gemaakt door de puntvoorspellingen van de vier kwartalen bij elkaar op te tellen of door jaardata over BBP-groei te gebruiken. Ten tweede wordt hier steeds gebruik gemaakt van een volledige dataset, terwijl in *real-time* vaak een onregelmatige dataset beschikbaar is, i.e. sommige reeksen zijn volledig beschikbaar, terwijl van anderen de laatste maand(en) missen. Het is interessant om te kijken wat er gebeurt met de voorspelprestatie als de werkelijk beschikbare dataset wordt gebruikt. Tot slot is hier abstractie gemaakt van het probleem van de dataherzieningen. Op voorwaarde dat eerste versie data beschikbaar is, kan onderzocht worden in hoeverre de voorspelprestatie verslechtert als er eerste versie data wordt gebruikt.

Literatuurlijst

Almon, S., 1965, The distributed lag between capital appropriations and expenditures, *Econometrica* 33(1), pag. 178-196.

Baffigi, A., R. Golinelli, en G. Parigi, 2004, Bridge models to forecast the euro area GDP, *International Journal of Forecasting* 20, pag. 447-460.

Bai, J., E. Ghysels en J.H. Wright, 2010, State space models and MIDAS regressions, Working paper.

Clements, M. en A. Galvào, 2006, Combining predictors and combining information in modelling: forecasting US recession probabilities and output growth, in C. Milas, P. Rothman en D. van Dijk (Eds.), *Nonlinear time series analysis of business cycles*, hoofdstuk 2, Elsevier.

Clements, M. en A. Galvào, 2008, Macroeconomic forecasting with mixed-frequency data, *Journal of Business and Economic Statistics*, 26, pag. 546-554.

Clements, M. en A. Galvào, 2009, Forecasting US output growth using leading indicators: an appraisal using MIDAS models, *Journal of* Applied Econometrics 57, pag. 1187-1206.

CPB, 2010, SAFFIER II, 1 model , in 2 hoedanigheden, voor 3 toepassingen, CPB document 217.

Croushore, D., 2006, *Forecasting with real-time macroeconomic data*, in G. Elliot, C.W. Granger en A. Timmermann (Eds.), *Handbook of Economic Forecasting*, hoofdstuk 17, Elsevier Noordholland.

Dek, L., 2010, Forecasting with mixed frequency data using a single equation model, Universiteit van Amsterdam.

Diebold, F. X. en G.D. Rudebusch, 1991, Forecasting output with the composite leading index: a real time analysis, Journal of the American Statistical Association 86, pag. 603-610.

Forni, M., M. Hallin, M. Lippi, en L. Reichlin, 2001, Coincident and leading indicator for the euro area, *The Econometric Journal* 111, pag. C62-C85.

Ghysels, E., P. Santa-Clara en R. Valkanov, 2002, The MIDAS touch: Mixed data sampling regression models, UCLA and UNC Discussion Paper.

Ghysels, E., P. Santa-Clara, en R. Valkanov, 2005, There is a risk-return tradeoff after all, *Journal of Financial Economics*, 76(3), pag. 509-548.

Ghysels, E., A. Sinko, en R. Valkanov, 2007, MIDAS regressions: Further results and new directions, *Econometric Reviews*, 26, pag. 53-90.

Greene, M. N., E.P. Howrey en S.H. Hymans, 1985, The use of outside information in econometric forecasting, in D. A. Belsey en E. Kuh (Eds.), *Model reliability*, hoofdstuk 4, The MIT press.

Howrey, E. P., 1996, Forecasting GBP with noisy data: a case study, *Journal of Economic and Social Measurement*, 22, pag. 181-200.

Klein, L.R. en E. Sojo, 1989, Combinations of high and low frequency data in macroeconometric models, in L. Klein en J. Marquez (Eds.), *Economic in Theory and Practice: An Eclectic Approach*, hoofdstuk 1, Kluwer Academic Publishers.

Klein, P. A., 2001, The leading indicators in historical perspective, in *Business Cycle Indicators Handbook*, BCI-Handbook., The Conference Board.

Kranendonk, H. en J. Bonenkamp en J. Verbruggen, 2003, De CPB-conjunctuurindicator geactualiseerd en gereviseerd, CPB document 40.

Kranendonk, H. en J.P. Verbruggen, 2006, SAFFIER: een 'multi purpose'-model van de Nederlandse economie voor analyses op korte en middellange termijn, CPB document 123.

Kuzin, V., Marcellino M. en Schumacher, C., 2009, MIDAS versus mixed-frequency VAR: nowcasting GDP in the euro area, Deutsche Bundesbank Discussion paper No. 07/2009.

Marcellino, M.G., J.H. Stock en M.W. Watson, 2006, A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series, *Journal of Econometrics*, 16, pag. 451-476.

Marcellino, M.G. en A. Musso, 2010, Real time estimates of the euro area output gap: Reliability and forecasting performance, ECB working paper No. 1157.

Mariano, R.S. en Y. Murasawa, 2003, A new coincident index of business cycles based on monthly and quarterly series, *Journal of Applied Econometrics*, 18, pag. 427-443.

Mariano, R.S. en Y. Murasawa, 2004, Constructing a coincident index of business cycles without assuming a one-factor model, SMU Economics and Statistics Working Paper No. 22-2004.

Mitchell, W. en A.F. Burns, 1938, Statistical indicators of cyclical revivals, NBER bulletin, 69, pg. 1-12.

Muns, S. en H, Kranendonk, 2010, De CPB-conjunctuurindicator gereviseerd; over endtime versus realtime, filters en puntschattingen, CPB document 219.

Parigi, G. en G. Schlitzer, 1995, Quarterly forecasts of the Italian business cycle by means of monthly economic indicators, *Journal of Forecasting*, 14(2), pag. 117-141.

Stark, T. en D. Croushore, 2002, Forecasting with a real-time data set for macroeconomists, *Journal of Macroeconomics*, 24, pag. 507-531.

Stock, J. H. en M.M. Watson, 1989, New indexes of coincident and leading economic indicators, NBER Macroeconomics Annual, 4, pag. 351-394.

Teräsvirta, T. en I. Mellin, 1983, Estimation of polynomial distributed lag models, Research Report No. 41, Department of Statistics, University of Helsinki.

Timmermann, A., 2006, Forecast combinations, in G. Elliot, G. C. W Granger en A. Timmermann (Eds.), *Handbook of economic forecasting*, hoofdstuk 4, Elsevier.

Zadrozny, P., 1988, Gaussian likelihood of continuous-time ARMAX models when data are stocks and flows at different frequencies, *Econometric Theory* 4, pag. 108-124.

Zadrozny, P.A., 1990, Forecasting U.S. GNP at monthly intervals with an estimated bivariate time series model, *Econometric Review* 75, pag. 2-15.

Bijlage A Onderzochte reeksen

Tabel A.1 Overzicht gebruikte leading indicatoren van dit onderzoek

Code	Betekenis	Herkomst	Transformatie
bdrkmc	Bedrijvigheid komende maand industrie	CBS[a]	
bdrvmc	Bedrijvigheid vorige maand industrie	CBS	
bedrutc	Bedrijvigheid utiliteitsbouw	EIB[b]	
bedrwoc	Bedrijvigheid woningbouw	EIB	
beursr	Beurs (reëel)	Yahoo Finance	1st verschil van logs
bfan	Uitvoer goederen exclusief diensten	CPB	1st verschil
borderc	Orderboek bouw	EC[c]	1st verschil
can	Particuliere consumptie	CPB	1st verschil
convrt	Consumentenvertrouwen	CBS	1st verschil
dollarwd	Mutatie dollarkoers	Yahoo Finance	1st verschil
ek	Economisch klimaat	CBS	1st verschil
faillc	Faillissementen excl. eenmanszaken	CBS	1st verschil
ifo	Producentenvertrouwen Duitsland	CESifo Group	
ifov	Verwachtingscomponent van IFO	CESifo Group	
koop	Koopbereidheid	CBS	1st verschil
lieur	OECD Leading indicator (euro area)	OECD[d]	
livs	OECD Leading indicator (VS)	OECD	
m1r	Geldhoeveelheid (M1)	DNB[e]	1st verschil van logs
ortotc	Totale orderontvangst	CBS	1st verschil
prod_ind	Producentenvertrouwen	DNB	1st verschil
rkrn	Korte termijn rente	DNB	1st verschil
rlrn	Lange termijn rente	CBS	1st verschil
vbbgvn	Bouwvergunningen (bedrijfsbouw)	CBS	
vbwovn	Bouwvergunningen (woningen)	CBS	1st verschil
vrendc	Voorraden	CBS	1st verschil
yield	Lange rente min korte rente	DNB	
yvihin	Industriële productie	CPB	1st verschil

[a] Centraal Bureau voor de Statistiek.

[b] Economisch Instituut voor de Bouwnijverheid.

[c] Europese Commissie.

[d] Organisation for Economic Co-operation and Development.

[e] De Nederlandsche Bank.

Bijlage B Reeksen voor multivariate analyse

Tabel B.1 Leading indicators gebruikt in multivariate analyse

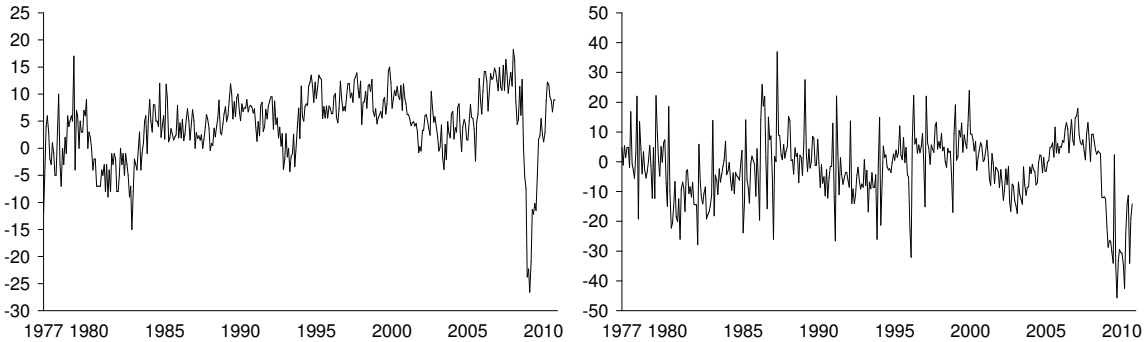
Code	Soort bron	Voorspelprestatie		
		$h = 1$	$h = 2$	$h = 3$
bdrkmc	Bedrijfsenquête	0,77	1,03	1,19
bedrwoc	Bedrijfsenquête	1,17	1,09	1,11
beursr	Monetaire variabele	0,66	0,73	0,96
koop	Consumentenonderzoek	1,20	1,10	1,06
m1r	Monetaire variabele	1,07	0,83	0,88
vbwovn	Overige	1,31	1,11	1,07

Overzicht van de reeksen die zijn gebruikt in de multivariate analyse van paragraaf 6. Per reeks is het soort bron en de voorspelprestatie voor drie horizonnen weergegeven. Deze voorspelprestatie is uitgedrukt in termen van de relatieve gemiddelde gekwadrateerde voorspelfout over de gehele evaluatielooptijd, zoals besproken in paragraaf 5.

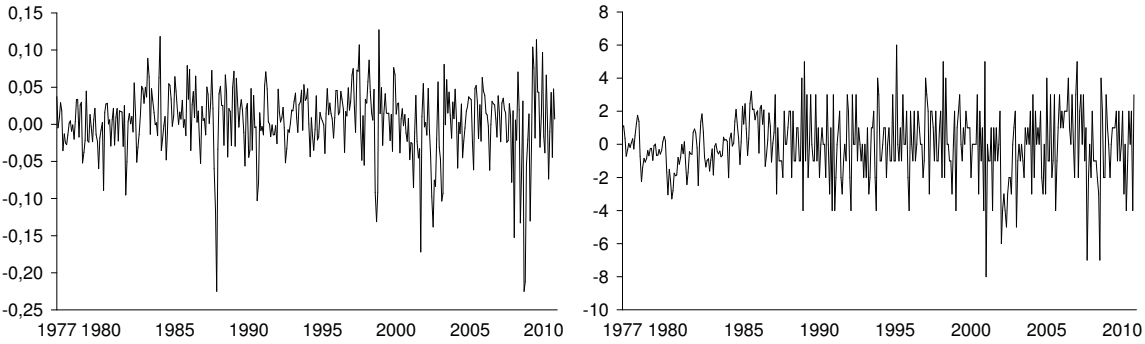
Tabel B.2 De correlatiematrix van de reeksen die zijn gebruikt voor de multivariate analyse

	bdrkmc	bedrwoc	beursr	koop	m1r	vbwovn
bdrkmc	1	0,49	0,02	0,11	– 0,01	0,02
bedrwoc		1	0,05	0,11	0,03	0,01
beursr			1	0,11	0,03	– 0,08
koop				1	0,00	– 0,01
m1r					1	0,00
vbwovn						1

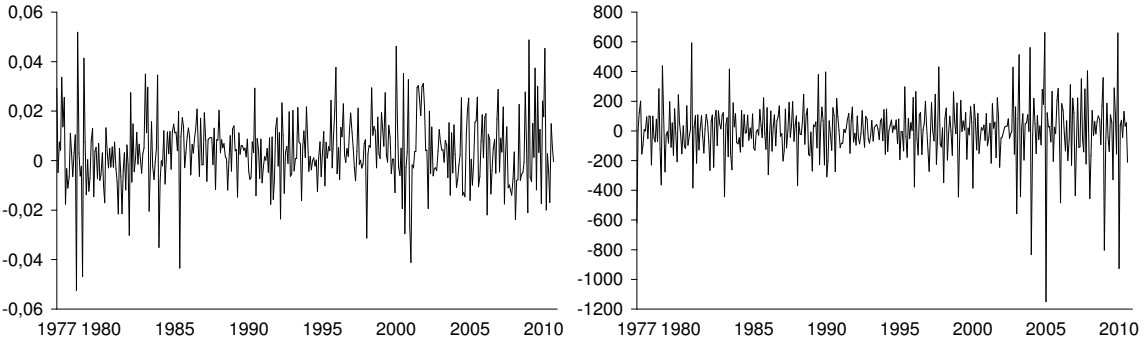
Figuur B.1 Verwachte bedrijvigheid industrie en verwachte bedrijvigheid woningbouw



Figuur B.2 AEX index (logverschil) en koopbereidheid (eerste verschil)

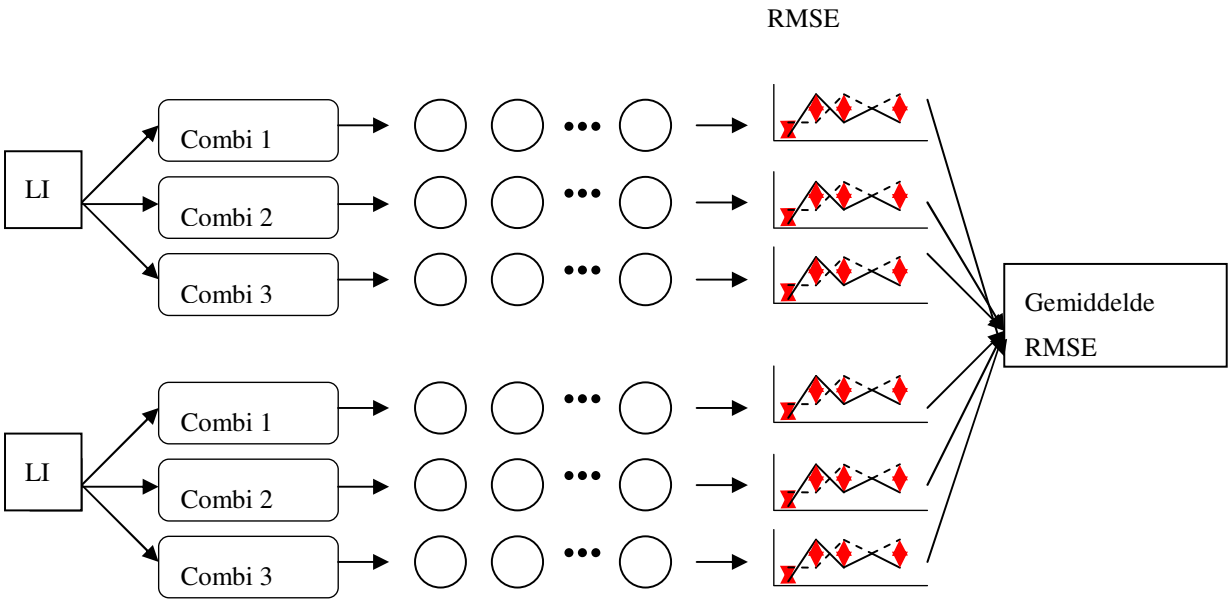


Figuur B.3 Geldhoeveelheid (logverschil) en bouwvergunningen woningbouw (logverschil)



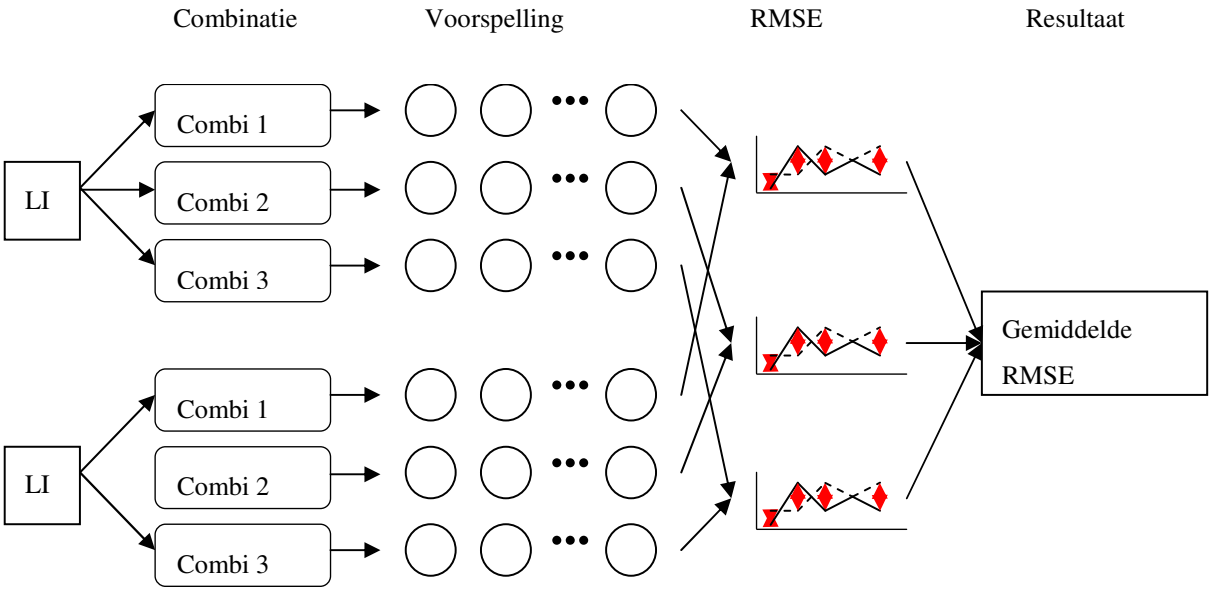
Bijlage C Schematische weergave poolen

Figuur C.1 Individuele voorspellingen met één leading indicator



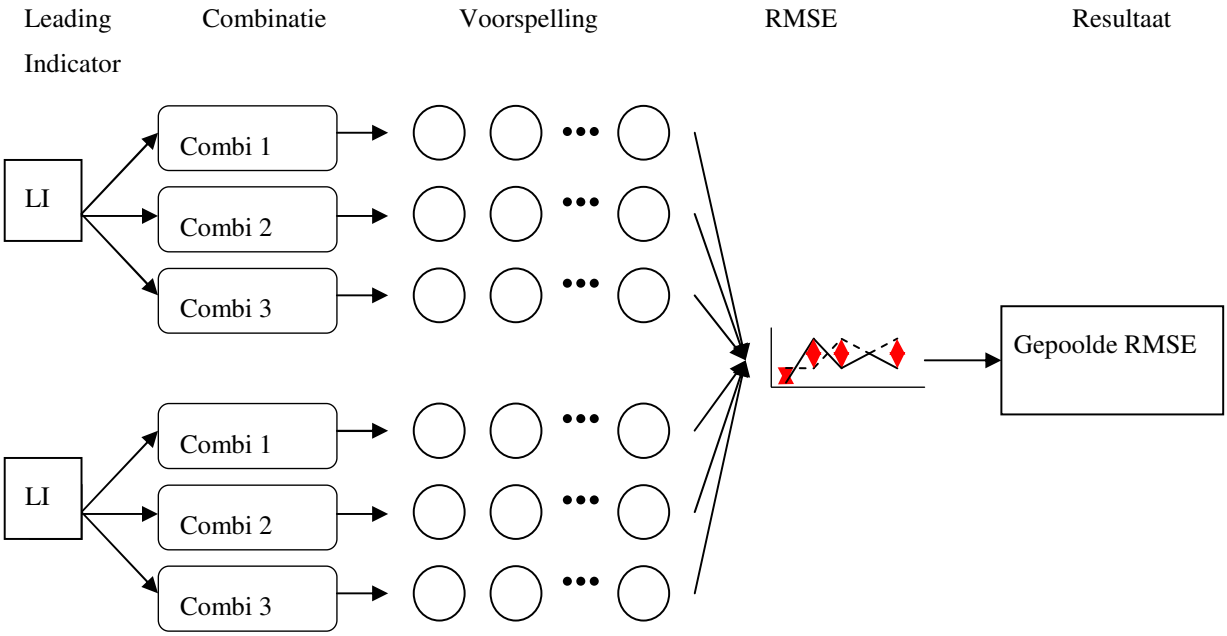
In deze grafiek wordt steeds één leading indicator per model gebruikt. Per leading indicator worden alle combinaties van model, vertragingsfunctie en selectiemethode geschat en gebruikt om een reeks voorspellingen te produceren. Deze voorspellingen worden op individuele basis vergeleken met de realisaties waarmee de voorspelfout en de gemiddelde gekwadrateerde voorspelfout kan worden berekend.

Figuur C.2 Poolen per combinatie over voorspellingen gemaakt met één leading indicator



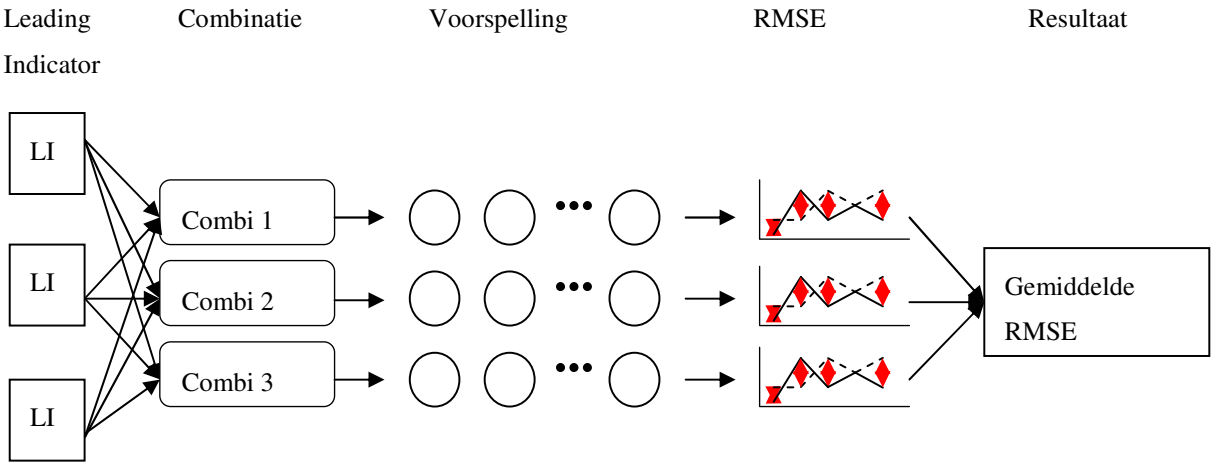
De figuur kent dezelfde opzet als figuur C.1. Het verschil zit hem er in dat de voorspellingen eerst worden gepoold over alle voorspellingen gemaakt met dezelfde combinatie van model, vertragingsfunctie en selectiemethode alvorens de gepoolde voorspellingen worden vergeleken met de realisatie. De in de eerste kolom van tabel 6 afgebeelde waarden zijn vervolgens het gemiddelde van deze RMSEen.

Figuur C.3 Poolen over alle combinaties en alle voorspellingen gemaakt met één leading indicator



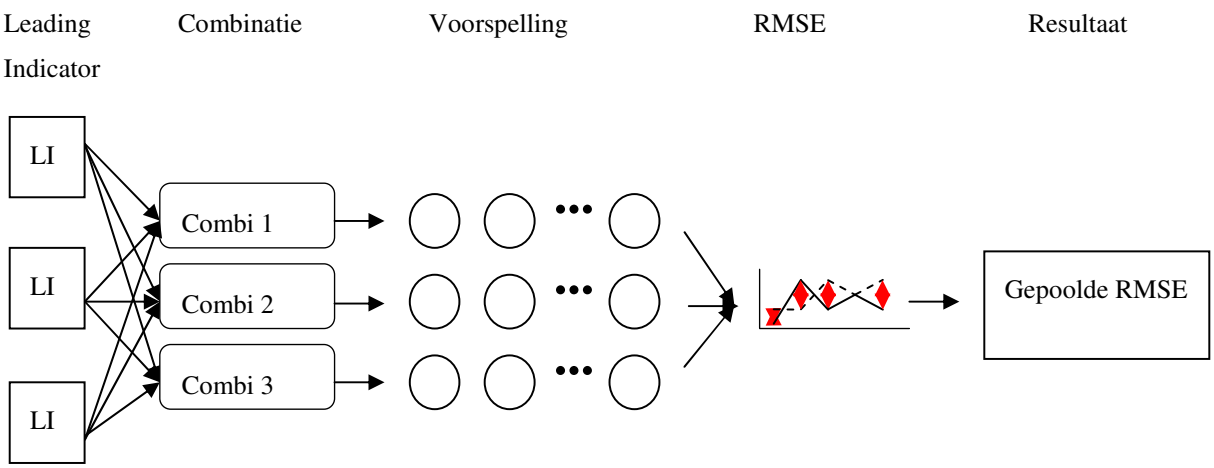
De voorspellingen in deze figuur zijn dezelfde als gemaakt in figuur C.1 en C.2. Hier wordt echter gepoold over al deze voorspellingen, dus over alle leading indicators én combinaties van model, vertragingsfunctie en selectiemethode.

Figuur C.4 Individuele voorspellingen met alle leading indicators



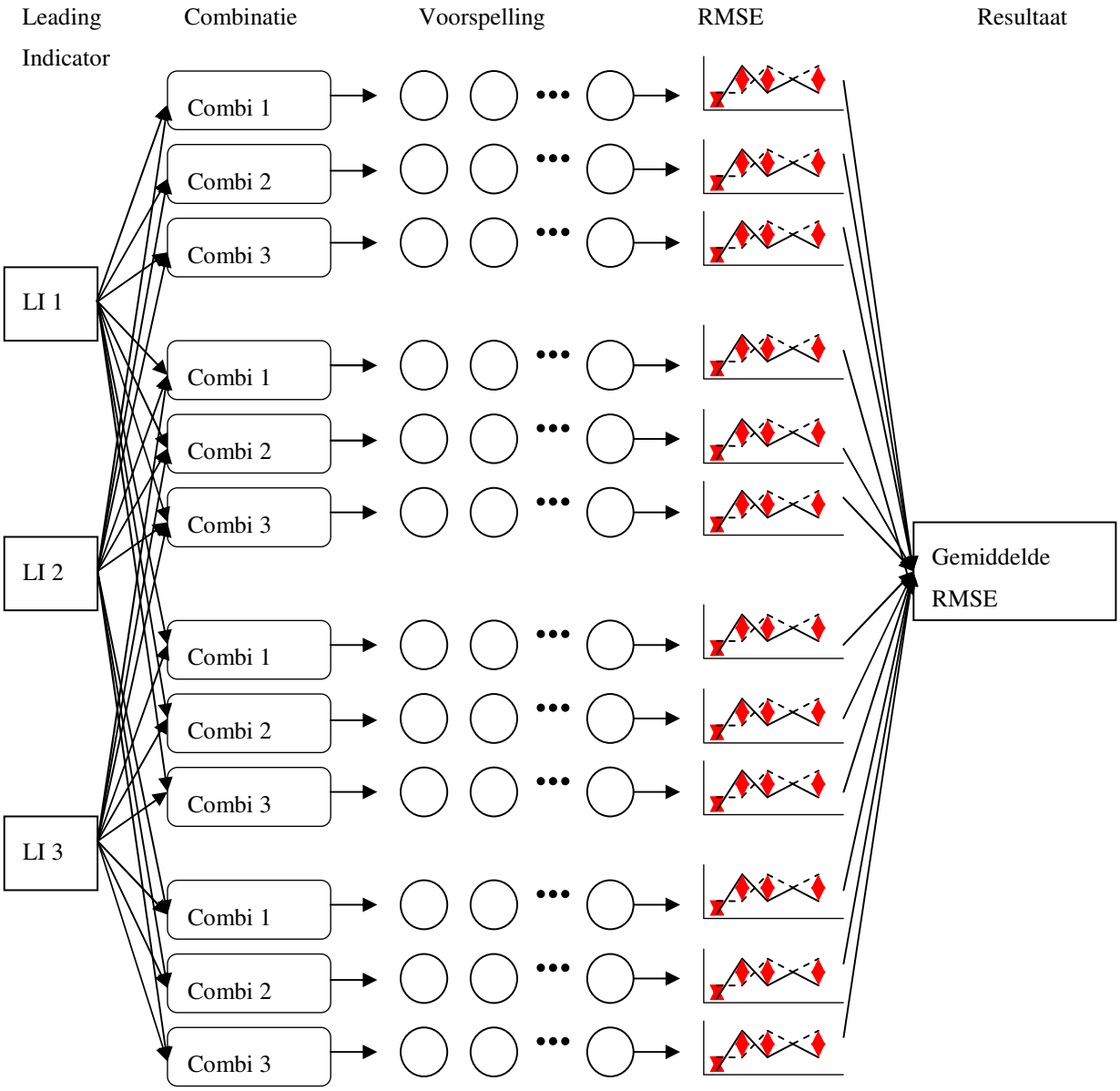
In deze figuur worden alle leading indicators gebruikt om te voorspellen. Dit wordt wederom gedaan door alle combinaties van model, vertragingsfunctie en selectiemethode te schatten, hier uitsluitend met alle leading indicators. Dit levert voorspellingen op die voor de verschillende combinaties worden vergeleken met de gerealiseerde BBP-groei om de voorspelfouten te berekenen.

Figuur C.5 Gepoolde voorspellingen met alle leading indicators



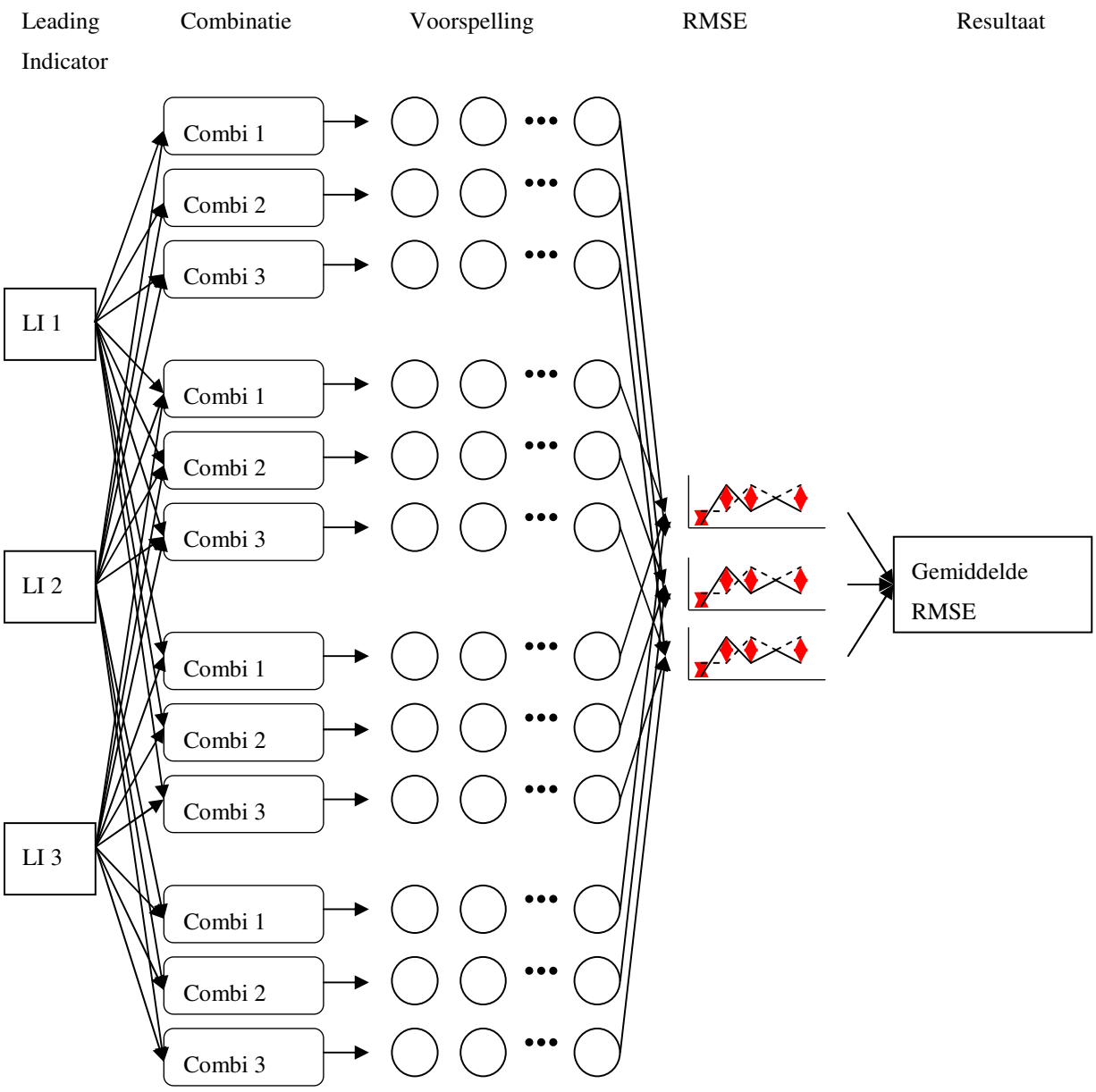
In deze figuur worden wederom alleen voorspellingen gedaan door combinaties die alle leading indicators gebruiken. Vervolgens wordt gepoold over de voorspellingen van alle combinaties. De resulterende voorspelling wordt vergeleken met gerealiseerde BBP-groei.

Figuur C.6 Combineren met meerdere leading indicators



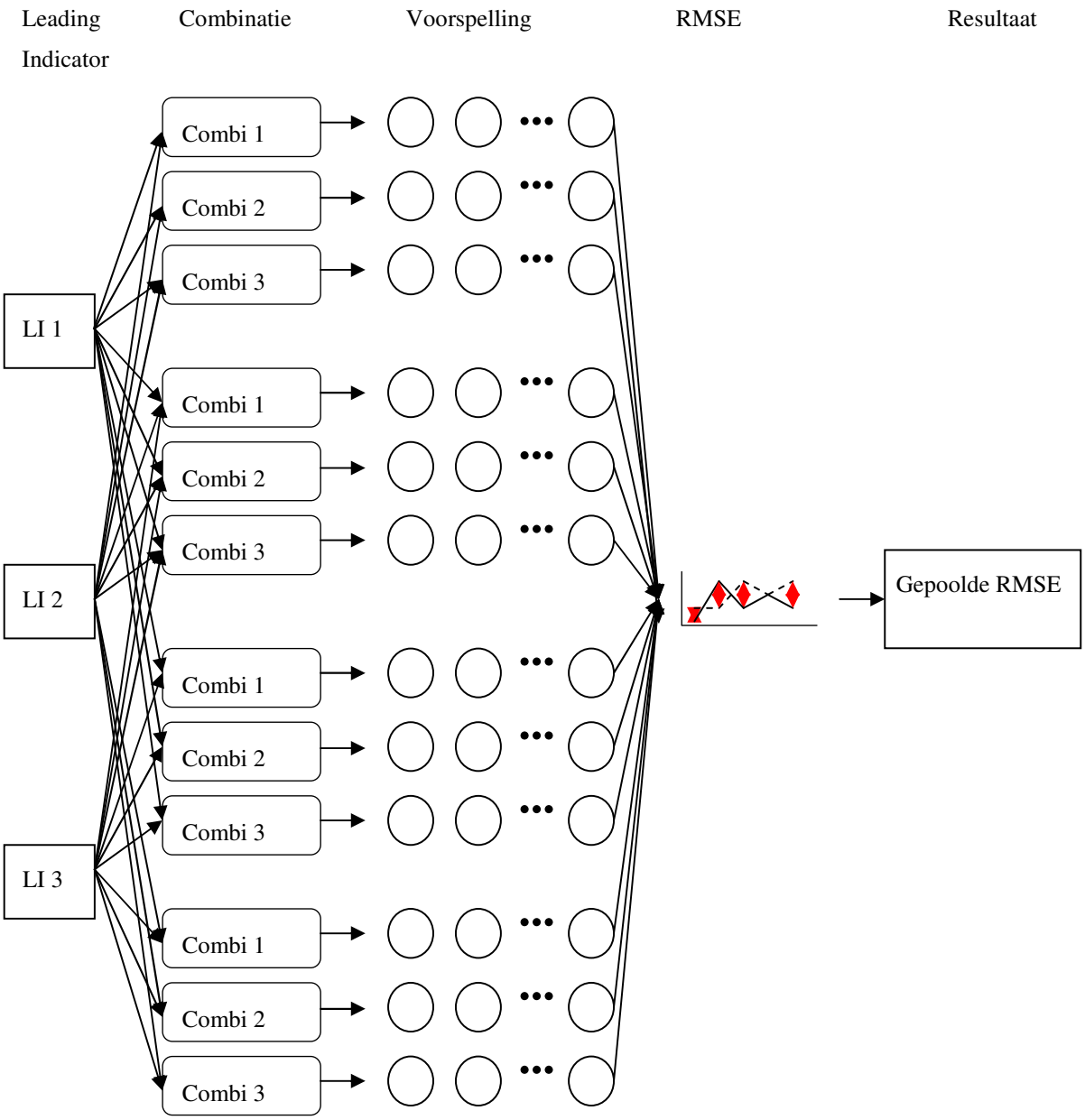
In deze figuur worden voorspellingen gedaan door modellen waarin meerdere combinaties van leading indicators worden gebruikt. Met drie leading indicators zijn er vier van deze combinaties: drie met twee leading indicators en één met alle drie. Er vindt geen pooling plaats, zodat de voorspellingen per set van leading indicators en combinatie van model, vertragingsfunctie en selectiemethode worden vergeleken met gerealiseerde BBP-groei.

Figuur C.7 Poolen over voorspellingen met meerdere leading indicators.



Deze figuur geeft dezelfde werkwijze weer als figuur C.6. Er wordt echter afgeweken doordat hier wordt gepoold over alle voorspellingen die met een bepaalde combinatie van model, vertragingsfunctie en selectiemethode zijn gemaakt.

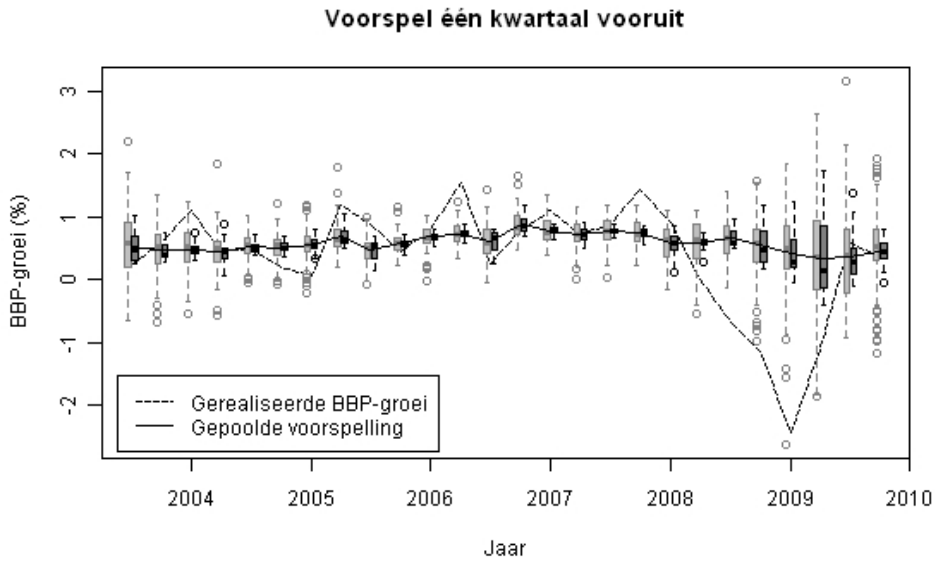
Figuur C.8 Poolen over alle voorspellingen gedaan met combinaties van leading indicators



In deze figuur wordt niet alleen over alle combinaties van leading indicators gepoold, maar ook alle combinaties van model, vertragingsfunctie en selectiemethode.

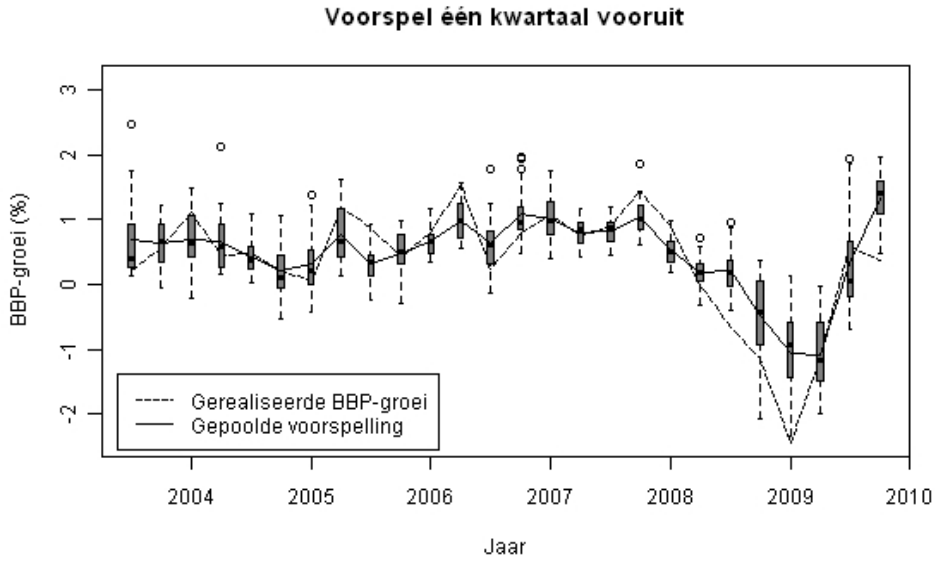
Bijlage D Grafische weergaven poolen

Figuur D.1 De resultaten van het gebruik van één leading indicator per keer



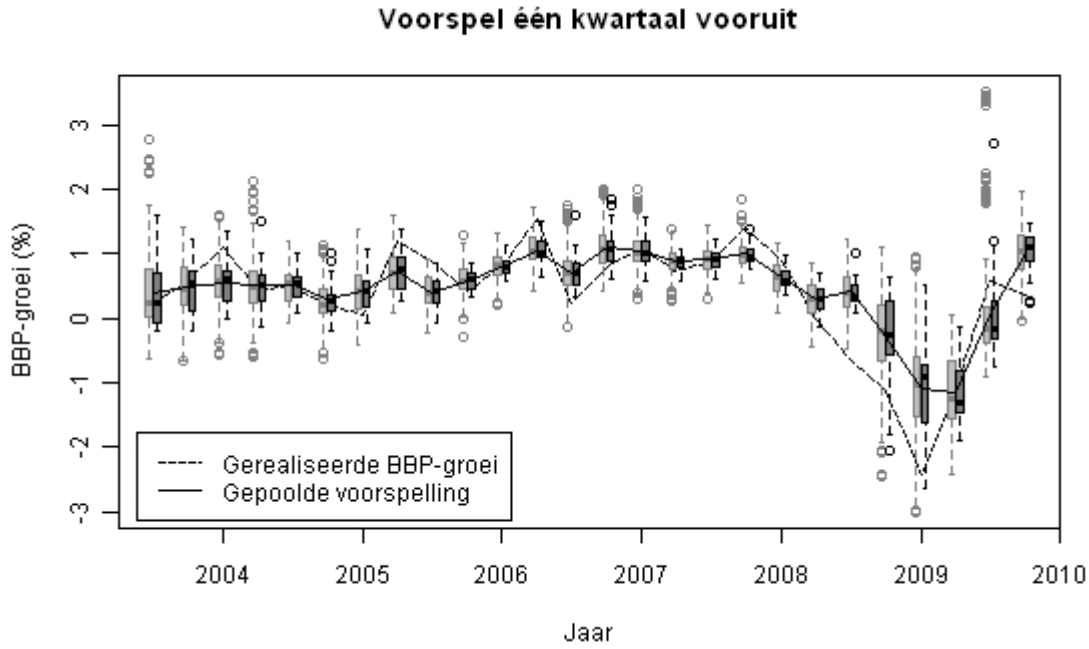
In de grafiek wordt gerealiseerde BBP-groei aangegeven door de stippellijn. Per kwartaal staan er twee boxplotten afgebeeld. De linker laat de spreiding zien van alle voorspellingen, gemaakt door de verschillende combinaties van leading indicator, model, vertragingsfunctie en selectiemethode. Deze aanpak staat schematisch weergegeven in figuur C.1. De rechter boxplot laat de spreiding zien als er wordt gepoold over alle leading indicators, wat correspondeert met figuur C.2. De doorgetrokken lijn geeft ten slotte het resultaat weer als er wordt gepoold over alle leading indicators, modellen, vertragingsfuncties en selectiemethoden, zoals schematische weergegeven in figuur C.3.

Figuur D.2 De resultaten van het gebruik van alle leading indicators in één model



Wederom geeft de stippellijn gerealiseerde BBP-groei aan. De boxplotten laten de spreiding zien van de voorspellingen gegenereerd door de verschillende combinaties van model, vertragingsfunctie en selectiemethode. Deze aanpak wordt schematisch weergegeven in figuur C.4. Poolen over al deze combinaties resulteert in de doorgetrokken lijn, zie figuur C.5.

Figuur D.3 De voorspellingen als meerdere combinaties van leading indicators worden gebruikt



Het format is dezelfde als in figuur D.1. Nu correspondeert de linker boxplot met figuur C.6, de rechter met figuur C.7, terwijl de doorgetrokken lijn het gevolg is van de procedure die in figuur C.8 wordt geschetst.