



Paper

# Versnelling indicatoren Monitor Brede Welvaart met behulp van nowcasting

Thema's R&D, Overheidsuitgaven, Armoede en  
Inkomen, Criminaliteit en Gezondheid

Pim Ouwehand  
Daan Zult  
Bob Lodder

Mei 2020

**CBS Den Haag**  
Henri Faasdreef 312  
2492 JP Den Haag  
Postbus 24500  
2490 HA Den Haag  
+31 70 337 38 00  
[www.cbs.nl](http://www.cbs.nl)

projectnummer

DRD  
6 maart 2020

# Inhoudsopgave

<b>1.</b>	<b>Inleiding</b>	<b>4</b>
<b>2.</b>	<b>Resultaten Thema 2: Research &amp; Development</b>	<b>8</b>
2.1	Onderzochte indicatoren	8
2.2	ENR13: Gewerkte uren speur- en ontwikkelingswerk	8
<b>3.</b>	<b>Resultaten Thema 3: Overheidsuitgaven</b>	<b>15</b>
3.1	Onderzochte indicatoren	15
<b>4.</b>	<b>Resultaten Thema 4: Armoede en inkomen</b>	<b>16</b>
4.1	Onderzochte indicatoren	16
4.2	EBD01: Prijsindex uitgaven aanschaf en bezit koopwoningen	17
4.3	SES01: Beloningsverschil mannen en vrouwen	20
<b>5.</b>	<b>Samenvatting en conclusies</b>	<b>23</b>

# 1. Inleiding

In dit rapport behandelen we de versnelling van een aantal indicatoren dat gepubliceerd wordt in de Monitor Brede Welvaart & Sustainable Development Goals. Er is gekeken naar indicatoren waarbij nog geen cijfer voor jaar T-1 (dus 2019 voor de Monitor 2020) opgenomen is. Aangezien het daarbij om een groot aantal indicatoren gaat, zijn deze ingedeeld in de volgende hoofdthema's:

1. Milieu, duurzaamheid en landbouw
2. Research & Development (R&D)
3. Overheidsuitgaven
4. Armoede en inkomen
5. Sterfte/criminaliteit/veiligheid en gezondheid

In dit rapport richten we ons op indicatoren uit thema's 2, 3 en 4. Thema 1 is in ons eerdere rapport<sup>1</sup> behandeld, thema 5 zal in een vervolgrapport aan bod komen.

In ons eerdere rapport is een uitgebreide beschrijving gegeven van de methodologie, de beoordeling van de uitkomsten, en de acceptatiecriteria om te besluiten of een indicator versneld opgenomen kan worden in de Monitor. Om de leesbaarheid van het huidige rapport te vergroten, herhalen we hieronder de belangrijkste elementen uit dat rapport.

## 1.1 Korte beschrijving methodologie

Indicatoren die minder actueel zijn dan T-1 proberen we te versnellen middels *nowcasting*-methoden. Dit zijn modelmatige technieken waarbij met toepassing van tijdreeksmodellen een zo recent mogelijk cijfer van de indicatoren geschat wordt. Bij *nowcasting* wordt aan de hand van de historische ontwikkeling van een indicator in combinatie met actuele hulpinformatie een schatting gemaakt van de meest recente ontwikkeling van de indicator. De historische reeks van de indicator bevat veel informatie, zoals trends en seizoenpatronen. Door deze dynamiek te modelleren wordt het mogelijk de doelvariabele te extrapoleren naar de meest actuele perioden. Om deze schattingen verder te verbeteren zal ook hulpinformatie in de modellen gebruikt worden. De hulpinformatie bestaat in de praktijk uit gerelateerde indicatoren, ook wel *proxies* genoemd, die een sterke samenhang vertonen met de doelvariabele.

## 1.2 Kwaliteitsmaten

Voor elke *nowcast* worden zes historische jaren *genowcast*, waardoor de werkelijke waarden ( $y_t$ ) kunnen worden vergeleken met de *genowcaste* waarden ( $\hat{y}_t$ ). In dit geval de jaren 2013 t/m 2018. Het verschil tussen deze twee waarden kan worden geëvalueerd aan de hand van de volgende maten:

Mean error:

$$ME = \frac{1}{N} \sum_{t=T-N}^T (\hat{y}_t - y_t)$$

Mean Absolute Error:

---

<sup>1</sup> Ouwehand, P., D. Zult, B. Lodder (2019), Versnelling indicatoren Monitor Brede Welvaart met behulp van *nowcasting*; Thema Milieu, duurzaamheid en landbouw, CBS-rapport.

$$\text{MAE} = \frac{1}{N} \sum_{t=T-N}^T |\hat{y}_t - y_t|$$

Root Mean Squared Error:

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{t=T-N}^T (\hat{y}_t - y_t)^2}$$

Mean Absolute Percentage Error:

$$\text{MAPE} = \frac{1}{N} \sum_{t=T-N}^T \left| \frac{\hat{y}_t}{y_t} - 1 \right|$$

Als de nowcasts perfect zijn (ofwel  $\hat{y}_t = y_t$ ) zal voor elk van deze maten het resultaat 0 zijn. Omdat dit in de praktijk nooit zo zal zijn, geldt voor al deze maten dat ze idealiter dicht bij nul liggen. Het kan echter zijn dat de ene maat, bijvoorbeeld ME het ene model prefereert terwijl de andere maat, bijvoorbeeld MSE, een ander model prefereert. Hieronder beschrijven we hoe we op basis van o.a. de kwaliteitsmaten een geschikt model kiezen.

### 1.3 Acceptatiecriteria

Wanneer vinden we de nowcasts van voldoende kwaliteit om deze te kunnen publiceren in de volgende editie van de Monitor? Op het CBS is een richtsnoer ‘modelmatig schatten bij het CBS’ (Buelens et al., 2013) opgesteld. De volgende aanbevelingen hieruit vormen de basis voor onze acceptatiecriteria:

*In de officiële statistiek worden modellen ingezet in situaties waar niet alle data die nodig zijn ook beschikbaar zijn. Met behulp van een model worden ontbrekende data geschat. Als richtsnoer voor dit gebruik van modellen geldt:*

1. *De gepubliceerde statistische gegevens moeten objectief en betrouwbaar blijven. De data waarmee het model geschat wordt, moeten daarom betrekking hebben op het verschijnsel dat beschreven wordt, dat wil zeggen dat de objecten en populaties van het model overeenkomen met het statistische verschijnsel dat beschreven wordt. Het model wordt alleen gebruikt voor tijdsperioden waarvoor gegevens beschikbaar zijn. Dit sluit het maken van voorspellingen, prognoses en analyses van beleidsvoornemens uit. Verder moeten de gepubliceerde statistische gegevens robuust zijn tegen falen van het model.*
2. *Of voor niet-gepubliceerde onderdelen, de “tussenproducten,” van statistische gegevens het gebruik van modellen wel aanvaardbaar is, zal afhangen van de relatieve omvang van die onderdelen en van het antwoord op de vraag of de risico’s die het gebruik van modellen voor die onderdelen met zich mee brengen ook voor het te publiceren gegeven gelden.*
3. *De gebruikte modellen moeten voldoen aan eisen rond de functie van het model, de voor schatting gebruikte data, standaardisatie van de methode, de modelselectie, de modelfit, de robuustheid en de stabiliteit van het model, de omvang van de gemiddelde kwadratische fout, en de verantwoording in de documentatie (zie paragraaf 3.3).*
4. *Het gebruik van modellen is in eerste instantie de verantwoordelijkheid van de statisticus en het statistisch management, waarbij DMS en PIM advies en ondersteuning*

*kunnen geven. Een assessment van het modelgebruik conform dit richtsnoer behoort tot de verantwoording van het statistische proces en behoort op de gebruikelijke wijze in de procesdocumentatie opgenomen te worden. Betreft het gebruik van modellen een wezenlijk element van de desbetreffende statistiek, dan wordt dit ook in de externe documentatie vermeld.*

Mede op basis hiervan komen we tot de volgende criteria:

#### **Procesversnelling of modelmatige versnelling**

Er is een voorkeur voor het versnellen door de inhoudelijke afdeling, door bijvoorbeeld het verwerkingsproces te versnellen of door een schatting te maken op basis van gedeeltelijke gegevens. Pas als dit niet mogelijk is, wordt een modelmatige aanpak via nowcasting onderzocht. In dat geval gelden de volgende criteria:

#### **Lengte tijdreeksen**

Om een goede nowcast te kunnen maken zijn er voldoende historische gegevens nodig. Dit is nodig om een model goed te schatten en om een real time analyse (crossvalidatie) te doen over voldoende perioden. Voor jaardata (wat alle indicatoren in de Monitor zijn) willen we minimaal 12 waarnemingen, waarbij een crossvalidatie gedaan wordt over minimaal de laatste 4 jaar. Dit is uiteraard een ondergrens, langere tijdreeksen verdienen de voorkeur. Ook is dit mede afhankelijk van de complexiteit van de dynamiek in de reeks. In de praktijk zijn voor alle indicatoren in de Monitor voldoende lange reeksen beschikbaar, en kunnen we een crossvalidatie doen over de laatste 6 jaar.

#### **Aanwezigheid hulpinformatie**

Om een goede nowcast te maken is informatie van een hulpbron nodig over de verslagperiode. Dit kan gedeeltelijke informatie over de doelvariabele zijn, of gerelateerde (zowel inhoudelijk samenhangend als gecorreleerd) informatie van een hulpvariabele. Als deze er niet is, zou een extrapolatie gemaakt worden op basis van allen de historie van de doelreeks. Conform punt 1 van het richtsnoer doen we dat niet.

Daarbij is het verstandig om terughoudend te zijn bij het gebruik van bbp als hulpvariabele, omdat anders teveel (nowcasts van) brede-welvaartsindicatoren afhankelijk worden van de ontwikkeling van het bbp, wat onwenselijk is. Het kan echter wel voorkomen, met name bij economische variabelen, dat het bbp als hulpreeks het model dermate kan verbeteren dat er toch voor gekozen wordt deze in te zetten.

#### **Type model**

Het beste model is niet altijd goed genoeg om te gebruiken voor publicatie. Het doel is immers om het gekozen model één of meerdere jaren te gaan gebruiken voor toekomstige nowcasts. Een Naïef model of lineaire regressie (zonder hulpvariabelen) zien we als benchmark, maar willen we niet gebruiken omdat deze de complexere dynamiek in de tijdreeks niet kunnen modelleren, met het risico dat toekomstige nowcasts onnauwkeurig zijn. Als deze modellen er als beste uitkomen is dit een teken dat de reeks complex is. Ook als een model zonder hulpvariabelen als beste uit de vergelijking komt, kiezen we ervoor deze niet te gebruiken.

#### **Kwaliteitsmaten**

Om een keuze te kunnen maken tussen modellen is aan de onderzoekers die de Monitor samenstellen gevraagd wat ze belangrijk vinden bij de kwaliteit van een nowcast. Hieruit bleek dat

een aantal constante voorspelfouten minder erg wordt gevonden dan afwisselend een aantal kleine en hele grote voorspelfouten. We kijken daarom eerst naar de gemiddelde kwadratische fout (RMSE) en daarna naar de absolute fout (MAE en MAPE). Vervolgens kijken we nog naar de structurele vertekening (ME). Kortom, in principe heeft een model met de laagste RMSE onze voorkeur, maar als de RMSE's van de verschillende modellen dicht bij elkaar liggen, kan het zijn dat op basis van één van de andere kwaliteitsmaten een andere model de voorkeur krijgt.

#### **Onzekerheidsmarge**

De historische analyse (crossvalidatie) levert een kwaliteitsmaat op aan de hand van de verschillen tussen de nowcasts en de werkelijke waarden. Daarnaast heeft elke nowcast ook een voorspelinterval of -marge, waarin de werkelijke waarde met 95% zekerheid zal liggen. We eisen dat dat de werkelijke waarde van de indicator in alle zes de perioden van de crossvalidatie binnen het 95%-voorspelinterval ligt. Eventueel mag één van de zes afwijken als het resultaat verder plausibel is.

#### **Plausibiliteit**

De historische nowcasts en bijbehorende marges worden ook door een expert bekeken om te beoordelen of ze nauwkeurig en plausibel genoeg zijn om de indicator te nowcasten. Daarbij wordt ook gelet op het voldoende signaleren van omslagpunten.

## 2. Resultaten Thema 2: Research & Development

### 2.1 Onderzochte indicatoren

In onderstaande tabellen staat een overzicht van de indicatoren binnen dit thema waarvoor nog geen cijfer voor T-1 was opgenomen, en de wijze waarop deze al dan niet versneld kunnen worden op basis van dit onderzoek.

Code in dit rapport	Naam indicator	Versneld door
-	Uitgaven aan R&D	Niet
-	Technologisch innoverende bedrijven	Niet
-	Publieke uitgaven aan R&D	Niet
-	Private uitgaven aan R&D	Niet
-	Bedrijven met productinnovaties	Niet

Tabel 2.1: Indicatoren die **niet** versneld zullen worden

Code in dit rapport	Naam indicator	Versneld door
-	Toegevoegde waarde MKB	Expert*

Tabel 2.2: Indicatoren die **wel** versneld kunnen worden (zonder nowcastmodellen)

\*deze indicator kan versneld worden van T-3 naar T-2.

Code in dit rapport	Naam indicator	Versneld door
ENR13	Gewerkte uren in speur- en ontwikkelingswerk	Model
-	Fysieke kapitaalgoederenvoorraad	Model
-	Kenniskapitaalgoederenvoorraad	Model

Tabel 2.3: Indicatoren die **wel** versneld kunnen worden **middels nowcastmodellen**

De gekozen modellering voor ENR13 wordt in de volgende paragraaf toegelicht.

### 2.2 ENR13: Gewerkte uren speur- en ontwikkelingswerk

De indicator *Gewerkte uren in speur- en ontwikkelingswerk* is afkomstig uit de Arbeidsrekeningen. In de Monitor wordt deze reeks weergegeven als gewerkte uren per inwoner. In dit onderzoek zullen we de reeks gewerkte uren zelf analyseren.

De reeks is zowel op kwartaalniveau als jaarniveau beschikbaar, waarbij de jaarcijfers telkens de som zijn van de kwartalen in een jaar. Er is historische data vanaf 1995.

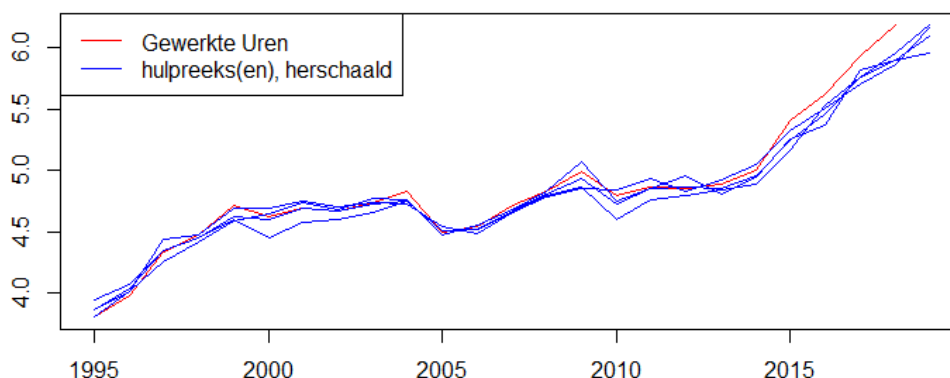
Bij het maken van de nowcast voor de Monitor zijn de reguliere schattingen van de eerste drie kwartalen van het afgelopen jaar beschikbaar, plus een flashraming voor het vierde kwartaal. De flashraming is een eerste voorlopige schatting, de reguliere zijn de opvolgende, bijgestelde schattingen. Deze laatste liggen dus in principe iets dichterbij de uiteindelijke schattingen. Uit het verleden zijn deze cijfers niet meer voorhanden. Om toch hulpreeksen te maken voor onze nowcast



gebruiken we de definitieve kwartaalcijfers sinds 1995. Vanaf 2016 hebben deze een minder definitieve status. We definiëren de volgende reeksen:

- Q1: Een reeks bestaande uit alle eerste kwartalen per jaar
- Q2: Een reeks bestaande uit alle tweede kwartalen per jaar
- Q3: Een reeks bestaande uit alle derde kwartalen per jaar
- Q123: Een reeks bestaande uit de som van kwartalen 1 t/m 3 per jaar

In onderstaande grafiek is de goede samenhang met de doelvariabele duidelijk te zien:



Bij het doorrekenen van de modellen staan we niet alle combinaties van bovenstaande regressoren toe. Q1, Q2 en Q3 kunnen wel samen in een model worden opgenomen, maar Q123 in combinatie met een of meerdere van Q1, Q2 of Q3 staan we niet toe, aangezien deze laatste al opgenomen zijn in Q123.

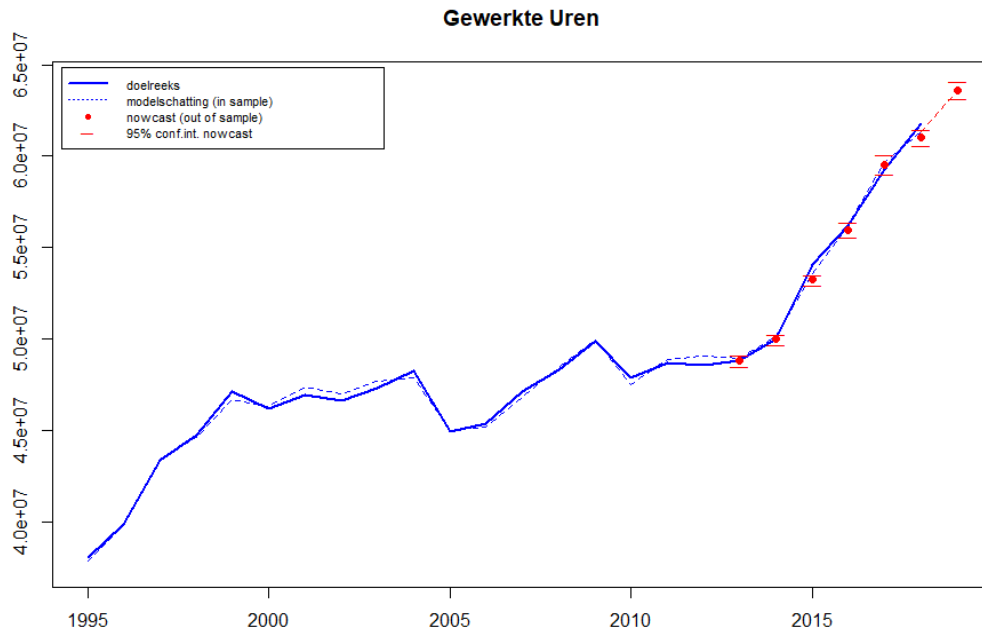
We komen zo tot de volgende 10 beste modellen:

RMSE	ME	MAPE	MAE	Model	Hulpreeksen
0,035036	-0,01397	0,474783	0,027037	3 = Lineaire regressie	Q1 + Q2 + Q3
0,035966	-0,00559	0,482798	0,027479	3 = Lineaire regressie	Q123
0,035974	-0,00558	0,482985	0,02749	5 = Smooth trend model	Q123
0,037057	-0,01818	0,502329	0,028509	5 = Smooth trend model	Q1 + Q2 + Q3
0,039567	0,021704	0,654226	0,03578	1 = Deterministic level model	Q123
0,040399	-0,00248	0,614431	0,034816	2 = Local level model	Q123
0,041198	-0,01919	0,614468	0,034232	4 = Local level with deterministic slope	Q1 + Q2 + Q3
0,041198	-0,01922	0,614353	0,034225	6 = Local linear trend model	Q1 + Q2 + Q3
0,041286	-0,00659	0,65403	0,036461	4 = Local level with deterministic slope	Q123
0,041293	-0,00657	0,65408	0,036465	6 = Local linear trend model	Q123

Om de resultaten te kunnen berekenen zijn de reeksen met een factor 10 000 000 verkleind, om zo numerieke problemen te voorkomen. In bovenstaande tabel staat een RMSE van 0,035 dus gelijk aan 350 000 op de oorspronkelijke schaal.

De modellen schelen niet heel veel qua prestaties, maar het derde model is na visuele beoordeling het beste. Deze levert een nowcast voor 2019 van 63 618 940 (met een 95%-interval van 63 151 670 – 64 086 200).

In onderstaande grafiek zijn de uitkomsten van dit model weergegeven.



### Conclusie

Voor de hulpvariabelen wordt gebruik gemaakt van data die over de eerste drie kwartalen van het verslagjaar beschikbaar zijn, en bovendien van de doelvariabele (en onderdelen daarvan) zelf. Daarmee lukt het een nauwkeurige nowcast te maken. Het beste model heeft de afgelopen jaren dan ook goede nowcasts geleverd. Gemiddeld zat deze er 0,47 % naast (MAPE). De inhoudelijk expert vindt de nowcast ook plausibel. We adviseren dus de nowcast van deze reeks op te nemen in de Monitor 2020.

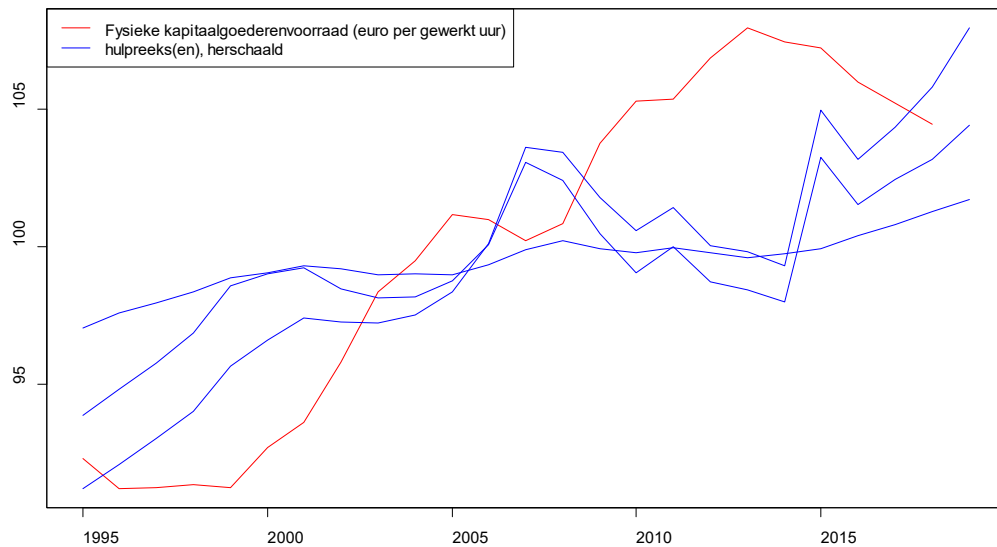
### 2.3 Fysieke kapitaalgoederenvoorraad.

De *fysieke kapitaalgoederenvoorraad* wordt in de Monitor weergegeven in euro's aan kapitaalgoederenvoorraad per gewerkt uur, maar de nowcast wordt gedaan op het niveau van de totale fysieke kapitaalgoederenvoorraad en daarna terug getransformeerd. Er is historische data beschikbaar vanaf 1995.

Bij het maken van de nowcast zijn de volgende hulpreeksen beschikbaar:

- Aantal gewerkte uren
- Bruto investeringen in prijzen van 2015
- Bruto investeringen in werkelijke prijzen.

De fysieke kapitaalgoederenvoorraad (in euro per gewerkt uur) en deze hulpreeksen (herschaald) zijn hieronder te zien.



Figuur 2.3.1: Fysieke kapitaalgoederenvoorraad (euro per gewerkt uur) en hulpreeksen (herschaald)

### Nowcasts

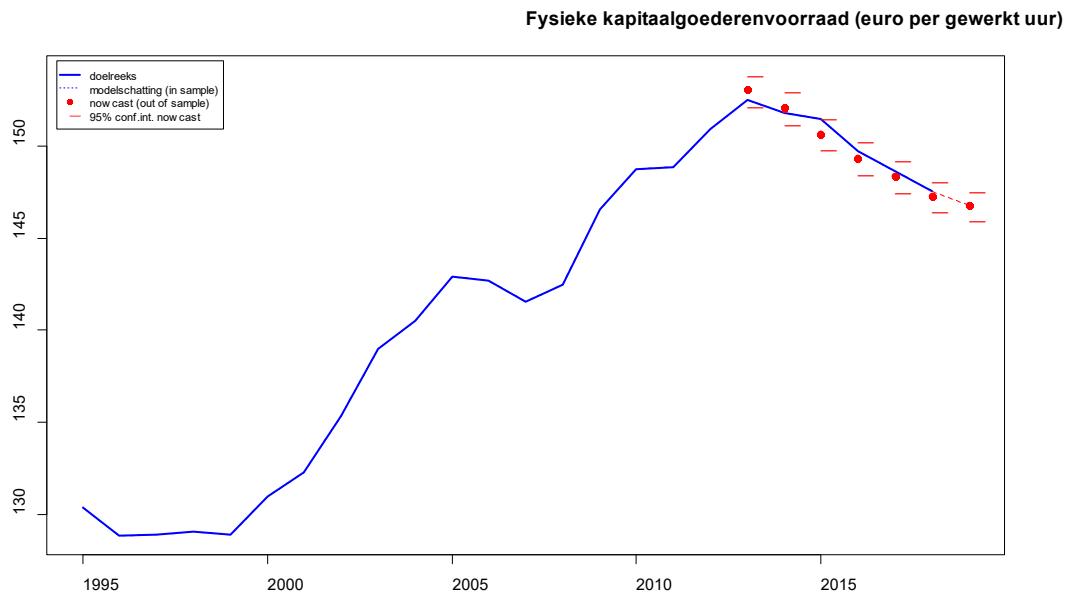
Omdat de doelreeks per gewerkt uur is weergegeven, vermenigvuldigen we eerst met het aantal gewerkte uren, zodat we de originele reeks krijgen en het aantal gewerkte uren zelf ook als hulpvariabele kan worden gebruikt. Alle modellen voorspellen zes keer één periode vooruit, d.w.z. 2013 tot en met 2018. Op basis van deze modellen zijn de evaluatiecriteria berekend. Voor de toegevoegde waarde ziet de tabel met de modelevaluatie van de beste tien modellen er zo uit:

RMSE	ME	MAPE	MAE	model	hulpreeksen
0,32	-0,10	0,0027	0,29	5 = Smooth trend model	Gewerkte uren
0,32	-0,10	0,0027	0,29	6 = Local linear trend model	Gewerkte uren
0,34	-0,10	0,0021	0,23	2 = Local level model	Gewerkte uren
0,35	-0,18	0,0028	0,30	5 = Smooth trend model	Gewerkte uren
0,35	-0,18	0,0028	0,30	6 = Local linear trend model	Gewerkte uren
0,36	-0,10	0,0026	0,28	2 = Local level model	Totale bruto investeringen in 2015 prijzen + Gewerkte uren
0,41	-0,09	0,0032	0,34	6 = Local linear trend model	Totale bruto investeringen in werkelijke prijzen + Gewerkte uren
0,41	-0,09	0,0032	0,34	5 = Smooth trend model	Totale bruto investeringen in 2015 prijzen + Totale bruto investeringen in werkelijke prijzen + Gewerkte uren
0,45	-0,19	0,0036	0,39	6 = Local linear trend model	Totale bruto investeringen in werkelijke prijzen + Gewerkte uren
0,45	-0,19	0,0036	0,39	5 = Smooth trend model	Totale bruto investeringen in 2015 prijzen + Gewerkte uren

Tabel 2.3.1: Modelvergelijking beste 10 modellen fysieke kapitaalgoederenvoorraad.

Op basis van tabel 2.3.1 verdient voor de fysieke kapitaalgoederenvoorraad model 5 (= Smooth trend model) met de hulpreeks gewerkte uren de voorkeur, omdat deze zowel de kleinste RMSE als kleinste ME kent.

Met dit model is de nowcast voor de fysieke kapitaalgoederenvoorraad per gewerkt uur in 2019 gelijk aan 146,76 euro met 95%-betrouwbaarheidsintervallen van (145,96; 147,55). De grafiek met historische nowcasts en een nowcast voor 2019 ziet er als volgt uit:



Figuur 2.3.2: Nowcast fysieke kapitaalgoederenvoorraad (euro per gewerkt uur)

### Conclusie

De nowcasts van de reeks fysieke kapitaalgoederenvoorraad (euro per gewerkt uur) ziet er goed uit. De nowcasts liggen allemaal dicht bij de werkelijke waarden en binnen de genowcaste betrouwbaarheidsintervallen. Ook zijn de nowcasts gebaseerd op een inhoudelijk hulpreeks en kan de inhoudelijk specialist zich er prima in vinden. Er lijkt methodologisch en inhoudelijk gezien dus geen bezwaar tot het nowcasten van de fysieke kapitaalgoederenvoorraad.

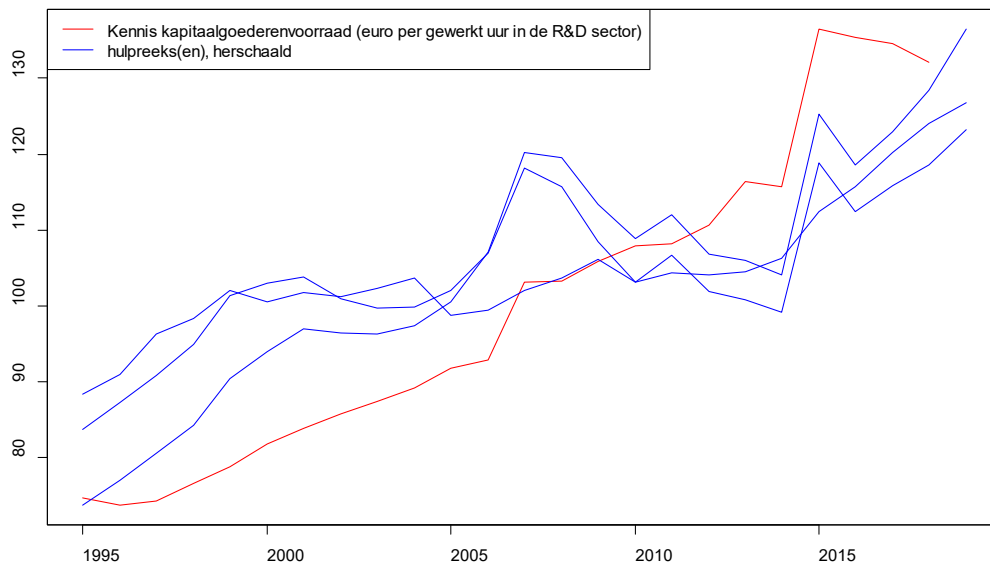
## 2.4 Kenniskapitaalgoederenvoorraad.

De *kenniskapitaalgoederenvoorraad* wordt in de Monitor weergegeven in euro's aan kapitaalgoederenvoorraad per gewerkt uur in de R&D sector, maar de nowcast wordt gedaan op het niveau van de totale kenniskapitaalgoederenvoorraad en daarna teruggetransformeerd. Er is historische data beschikbaar vanaf 1995.

Bij het maken van de nowcast zijn de volgende hulpreeksen beschikbaar:

- Aantal gewerkte uren in R&D.
- Bruto investeringen in prijzen van 2015
- Bruto investeringen in werkelijke prijzen.

De fysieke kapitaalgoederenvoorraad (in euro per gewerkt uur) en deze hulpreeksen (herschaald) zijn hieronder te zien.



Figuur 2.4.1: Kenniskapitaalgoederenvoorraad (euro per gewerkt uur in de R&D sector) en hulpreeksen (herschaald)

### Nowcasts

Omdat de doelreeks per gewerkt uur is weergegeven, vermenigvuldigen we eerst met het aantal gewerkte uren, zodat we de originele reeks krijgen en het aantal gewerkte uren zelf ook als hulpvariabele kan worden gebruikt. Alle modellen voorspellen zes keer één periode vooruit, d.w.z. 2013 tot en met 2018. Op basis van deze modellen zijn de evaluatiecriteria berekend. Voor de toegevoegde waarde ziet de tabel met de modelevaluatie van de beste tien modellen er zo uit:

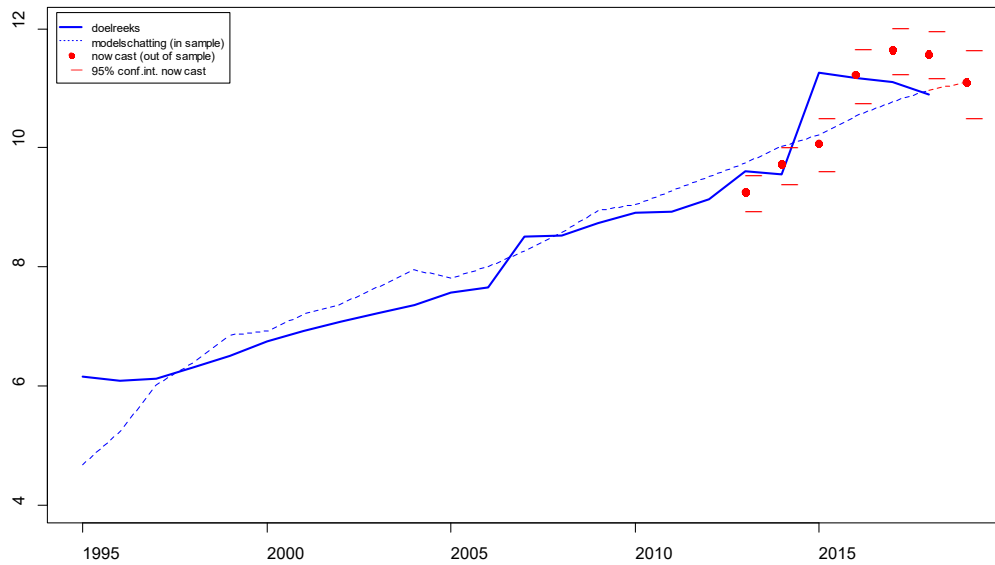
RMSE	ME	MAPE	MAE	model	reg
6,65	-1,31	0,037	4,858	6 = Local linear trend model	bruto investeringen totaal: werkelijke prijzen + bruto investeringen totaal; prijsniveau 2015 + Uren RenD
7,56	-0,17	0,046	6,032	6 = Local linear trend model	bruto investeringen totaal: werkelijke prijzen + Uren RenD
7,61	-2,14	0,037	4,851	7 = ARIMA	bruto investeringen totaal: werkelijke prijzen + Uren RenD
7,74	-3,56	0,043	5,703	2 = Local level model	bruto investeringen totaal: werkelijke prijzen + Uren RenD
7,80	0,47	0,049	6,419	5 = Smooth trend model	bruto investeringen totaal: werkelijke prijzen + bruto investeringen totaal; prijsniveau 2015 + Uren RenD
8,04	-2,95	0,043	5,696	4 = Local level with deterministic slope	bruto investeringen totaal: werkelijke prijzen + Uren RenD
8,06	-2,01	0,044	5,785	7 = ARIMA	bruto investeringen totaal; prijsniveau 2015
8,15	-0,08	0,049	6,427	7 = ARIMA	bruto investeringen totaal; prijsniveau 2015 + Uren RenD
8,17	-2,86	0,045	5,985	4 = Local level with deterministic slope	bruto investeringen totaal: werkelijke prijzen + bruto investeringen totaal; prijsniveau 2015 + Uren RenD
8,23	-3,93	0,042	5,571	2 = Local level model	bruto investeringen totaal; prijsniveau 2015 + Uren RenD

Tabel 2.4.1: Modelvergelijking beste 10 modellen kenniskapitaalgoederenvoorraad.

Op basis van tabel 2.4.1 verdient voor de kenniskapitaalgoederenvoorraad model 6 (= Local linear trend model) met de hulpreeks gewerkte uren de voorkeur en bruto investeringen in werkelijke prijzen de voorkeur, omdat deze bijna de kleinste RMSE en de kleinste ME kent.

Met dit model is de nowcast voor de fysieke kapitaalgoederenvoorraad per gewerkt uur in 2019 gelijk aan 11,10 euro met 95%-betrouwbaarheidsintervallen van (10,54; 11,67). De grafiek met historische nowcasts en een nowcast voor 2019 ziet er als volgt uit:

### Kennis kapitaalgoederenvoorraad (euro per gewerkt uur)



Figuur 2.4.2: Nowcast kenniskapitaalgoederenvoorraad (euro per gewerkt uur in de R&D sector)

#### Conclusie

De nowcasts van de reeks kenniskapitaalgoederenvoorraad (euro per gewerkt uur) zien er niet zo goed uit als eerdere nowcasts. Volgens de inhoudelijke expert komt dit met name omdat er in 2015 een grote investering gedaan is die in de jaren daarna voor extra afschrijvingen zorgde die niet worden opgepakt door het model. De expert verwacht zelf voor 2019 een waarde van 10,82, welke lager ligt dan de nowcast maar nog wel binnen het betrouwbaarheidsinterval. Na overleg heeft de expert aangegeven nog wel met de nowcast te kunnen leven, omdat het beter is dan niks. Er kan echter ook voor gekozen worden om de geschatte waarde van de expert in te zetten.

### 3. Resultaten Thema 3: Overheidsuitgaven

#### 3.1 Onderzochte indicatoren

In onderstaande tabellen staat een overzicht van de indicatoren binnen dit thema waarvoor nog geen cijfer voor T-1 was opgenomen, en de wijze waarop ze al dan niet versneld kunnen worden op basis van dit onderzoek.

Code in dit rapport	Naam indicator	Versneld door
-	Overheidsuitgaven aan algemeen overheidsbestuur	Niet
-	Overheidsuitgaven aan landsverdediging	Niet
-	Overheidsuitgaven aan sociale bescherming	Niet
-	Overheidsuitgaven aan volksgezondheid	Niet
-	Overheidsuitgaven aan milieubescherming	Niet
-	Overheidsuitgaven aan openbare orde en veiligheid	Niet
-	Overheidsuitgaven aan onderwijs	Niet

Tabel 3.1: Indicatoren die **niet** versneld zullen worden.

Code in dit rapport	Naam indicator	Versneld door
-	Schuld van de overheid	CPB / expert

Tabel 3.1: Indicatoren die **wel** versneld zullen worden.

## 4. Resultaten Thema 4: Armoede en inkomen

### 4.1 Onderzochte indicatoren

In onderstaande tabellen staat een overzicht van de indicatoren binnen dit thema waarvoor nog geen cijfer voor T-1 was opgenomen, en de wijze waarop ze al dan niet versneld kunnen worden op basis van dit onderzoek.

Code in dit rapport	Naam indicator	Versneld door
-	Verwachte kans op kindermishandeling**	Niet
-	Pensioenaanspraken	Niet
-	Chartaal geld en deposito's per huishouden	Niet
-	Gemiddelde schuld per huishouden	Niet
-	Goedkope scheefhuur	Niet
-	Gemiddelde hypotheekschuld huishoudens	Niet
-	Mediaan vermogen van huishoudens	Niet
-	Huishoudens met een langdurig laag inkomen	Niet
-	Mediane koopkracht	Niet
-	Gemiddeld besteedbaar inkomen	Niet
-	Mediaan besteedbaar inkomen	Niet
-	Economische zelfstandigheid vrouwen	Niet
-	Economische zelfstandigheid mannen	Niet
-	Ratio 80/20	Niet
-	Gini-coëfficiënt	Niet
-	Slachtofferschap van misdaad**	Niet
-	Slachtofferschap van cybercrime**	Niet
-	Vaak onveilig voelen in de buurt**	Niet
-	Fysiek en/of seksueel geweld door (ex-)partner*	Niet
-	Reistijd woon-werkverkeer	Niet

Tabel 4.1: Indicatoren die **niet** versneld zullen worden

\*\* deze indicatoren worden om het jaar gepubliceerd. In de Monitor 2020 zullen deze over T-1 rapporteren.

Code in dit rapport	Naam indicator	Versneld door
-	Mediane ratio verkoopprijs/vraagprijs	Expert
-	Arbeidsinkomensquote	Expert
-	Loon per gewerkt uur	Expert
-	Gezonde levensverwachting mannen	Expert
-	Gezonde levensverwachting vrouwen	Expert

Tabel 4.1: Indicatoren die **wel** versneld kunnen worden (zonder nowcastmodellen)



Code in dit rapport	Naam indicator	Versneld door
EBD01	Prijsindex uitgaven aanschaf en bezit koopwoningen	Model
SES01	Beloningsverschil mannen en vrouwen	Model

Tabel 4.1: Indicatoren die **wel** versneld kunnen worden **middels nowcastmodellen**

## 4.2 EBD01: Prijsindex uitgaven aanschaf en bezit koopwoningen

De indicator *Prijsindex uitgaven aanschaf en bezit koopwoningen* (met 2015=100) beschrijft de prijsontwikkeling van de uitgaven van woningeigenaren voor het bezitten van een eigen woning. De uitgaven betreft aanschaf/aankoop nieuwboukoopwoning, zelfbouwwoning en voormalige huurwoningen. Daarnaast zijn ook de verwervingskosten en de doorlopende kosten voor (groot) onderhoud en opstalverzekeringen opgenomen.

De reeks is zowel op kwartaalniveau als jaarniveau beschikbaar, waarbij de jaarcijfers telkens het gemiddelde zijn van de kwartalen in een jaar.

De gepubliceerde prijsindex is een gewogen gemiddelde van een aantal deelreeksen, allen prijsindices. De opbouw van het totaal is in onderstaand schema weergegeven, inclusief het gewicht wat elk van de deelreeksen krijgt.

Reekscode	gewichten 2019	beschikbaar (weken na kwartaal)
<b>R1 - Totaal</b>	1000	13
<b>R11 - Aggregaat R111, R112 en R113</b>	736,0580	13
<b>R111 - Aggregaat R1111 en R1112</b>	491,0805	13
R1111 - PNK	276,6679	13
R1112 - Zelfbouw+renovatie	214,4126	12
R112 - PBKfr	98,4745	4
R113 - PAK	146,5031	10
<b>R12 - Aggregaat R121 en R122</b>	263,9420	10
R121 - PGO	235,1459	5
R122 - POV	28,7960	10

Van deze deelreeksen zijn er twee op tijd (voor 1 maart) beschikbaar om het cijfer voor de Monitor te maken, dit zijn de aggregaten R112 en R121, met een aandeel van 33% in het totaal.

We onderzoeken de volgende aanpakken om een jaarcijfer te nowcasten voor de totaalreeks R1:

1. We kijken alleen naar de totaalreeks, en gebruiken als hulpvariabele de reeks bestaande uit het gemiddelde van de eerste drie kwartalen van elk jaar van de totaalreeks. We negeren dus dat twee deelreeksen reeds gepubliceerd zijn.
2. Idem, maar we onderzoeken alle deelreeksen als mogelijke hulpreeks. Voor de nog niet gepubliceerde nemen we het gemiddelde van kwartaal 1-3, voor R112 en R121 nemen we het gemiddelde van kwartaal 1-4. We negeren hier dus het wegingsschema, de geschatte regressiecoëfficiënten voor de hulpvariabelen zullen de weging bepalen. Op die manier

geeft het model aan welke deelreeksen voorspellend zijn voor het totaalcijfer. Let wel: aanpak 1 is één van de vele gevallen die in aanpak 2 wordt doorgerekend.

Niet onderzocht maar eventueel ook mogelijk:

3. We passen dezelfde aanpak als onder 1. toe op alle deelreeksen op het laagste aggregatieniveau, behalve op R112 en R121. Daarna berekenen we het aggregaat via het wegingsschema. Dit betekent dat we vier nowcasts moeten maken (dus R1111, R1112, R113, R122) om het aggregaat R1 te kunnen maken.
4. Een oplossing tussen 1 en 2 in. We nemen alle deelreeksen tezamen (volgens het wegingsschema) die nog niet tijdig beschikbaar zijn en nowcasten deze met als hulpreeks een reeks o.b.v. de eerste drie kwartalen van dit aggregaat, net als bij 1. Voor reeks R112 en R121 nemen we het gepubliceerde cijfer, vervolgens wegen we met de nowcast.

### Resultaten

In onderstaande tabel zijn de resultaten weergegeven voor aanpak 1, waarbij R1\_Q123 staat voor de hulpreeks gebaseerd op de eerste drie kwartalen per jaar. De tabel is gesorteerd op de kolom RMSE. Zoals blijkt is een model mét hulpvariabele altijd beter dan een pure extrapolatie zonder hulpvariabele.

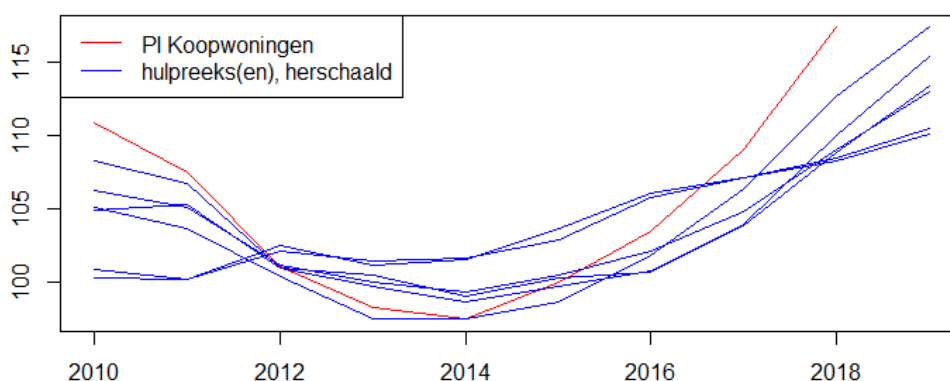
RMSE	ME	MAPE	MAE	Model	Hulpreksen
0,589603	-0,33425	0,538008	0,557729	2 = Local level model	R1_Q123
0,629477	-0,53761	0,533136	0,55813	7 = ARIMA	R1_Q123
0,829446	-0,56976	0,710342	0,757401	1 = Deterministic level model	R1_Q123
0,922033	-0,8103	0,791436	0,810298	5 = Smooth trend model	R1_Q123
0,938246	-0,648	0,708579	0,708266	4 = Local level with deterministic slope	R1_Q123
0,952799	-0,8283	0,809877	0,828298	6 = Local linear trend model	R1_Q123
1,162755	-0,75818	0,821901	0,818523	3 = Lineaire regressie	R1_Q123
2,838492	-2,61193	2,533062	2,611927	5 = Smooth trend model	-
3,034208	-2,75448	2,675878	2,754475	6 = Local linear trend model	-
4,610504	-2,71901	3,628995	3,919732	0 = Naive	-
4,610978	-2,71907	3,629444	3,920206	2 = Local level model	-
5,864669	-5,44966	5,134104	5,449657	4 = Local level with deterministic slope	-
7,855589	-1,85208	6,533671	6,93122	7 = ARIMA	-
8,691342	0,227862	6,954419	7,324472	1 = Deterministic level model	-
9,805959	-8,55122	7,956444	8,55122	3 = Lineaire regressie	-

Bij aanpak 2 wordt een groter aantal hulpreeks onderzocht, inclusief alle resultaten uit bovenstaande tabel. Aangezien het qua computergeheugen niet lukt om alle mogelijke combinaties door te rekenen hebben we de reeksen die een zeer afwijkend patroon t.o.v. de doelreeks vertonen niet meegenomen. De volgende doelreeksen blijven dan over:

- R1\_Q123
- R11\_Q123
- R111\_Q123
- R12\_Q123
- R112\_Q1234
- R121\_Q1234

Waarbij de eerste 4 gebaseerd zijn op kwartaal 1-3 en de laatste twee de reeksen zijn zoals gepubliceerd, dus o.b.v. kwartaal 1-4. In onderstaande figuur is de samenhang tussen deze reeksen

en de doelvariabelen weergegeven:



In onderstaande tabel worden de beste modellen weergegeven:

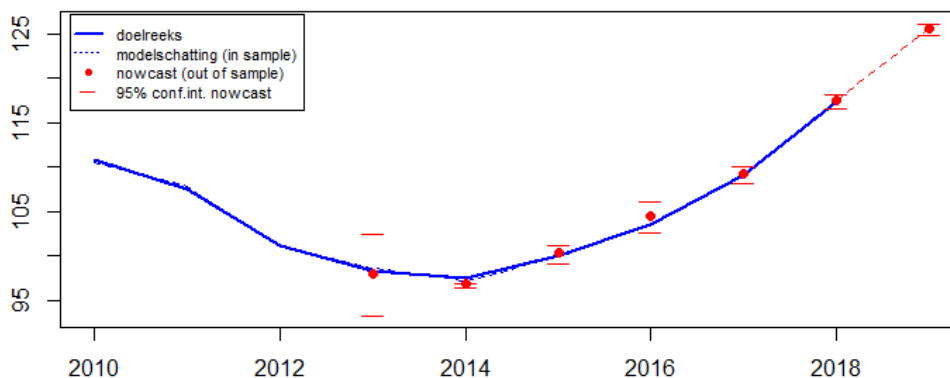
RMSE	ME	MAPE	MAE	Model	Hulpreeksen
0,48919	-0,02012	0,300621	0,327816	1 = Deterministic level model	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R121_Q1234
0,526265	-0,31824	0,461034	0,489101	7 = ARIMA	R1_Q123 + R11_Q123
0,535035	0,101363	0,425516	0,432083	1 = Deterministic level model	R11_Q123 + R12_Q123
0,549937	-0,36903	0,358822	0,392667	5 = Smooth trend model	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R121_Q1234
0,553244	-0,37118	0,360652	0,394816	3 = Lineaire regressie	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R121_Q1234
0,553939	-0,37163	0,361034	0,395265	6 = Local linear trend model	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R121_Q1234
0,558264	-0,3744	0,363398	0,39804	4 = Local level with deterministic slope	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R121_Q1234
0,571497	-0,27768	0,437374	0,453527	4 = Local level with deterministic slope	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R112_Q1234
0,572177	-0,27836	0,437948	0,454201	3 = Lineaire regressie	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R112_Q1234
0,572707	-0,27888	0,438393	0,454724	6 = Local linear trend model	R1_Q123 + R11_Q123 + R111_Q123 + R12_Q123 + R112_Q1234

We zien dat deze modellen qua RMSE, ME en MAPE beter scoren dan de beste modellen van aanpak 1. Kennelijk loont het dus om op een lager niveau te modelleren. Van deze modellen vinden we het eerste model minder geschikt omdat zowel R1 als R11 als R111 als hulpbron gebruikt wordt. Het derde model is daarom het beste model. Aangezien het een 'deterministic level model' is, levert dit model feitelijk een gewogen som van de hulpreeksen plus een lineaire transformatie om deze som aan te laten sluiten op de doelreeks. De geschatte coëfficiënten komen daarom ook ongeveer overeen met het eerder weergegeven gewichtenschema:

	Estimate	Std. Error
R11_Q123	0.72081	0.01208
R12_Q123	0.42530	0.01751
level	-14.21073	1.93884

Dit model levert een nowcast voor 2019 op van 125.5454 (met 95%-betrouwbaarheidsinterval 124.9283 - 126.1625). Onderstaande figuur toont hoe dit model de afgelopen perioden presteerde ten opzichte van de daadwerkelijk gepubliceerde reeks:

## PI Koopwoningen



De grotere marge in 2013 is overigens te verklaren door het feit dat er op dat moment nog erg weinig historische data was om een goede schatting te maken.

### Conclusie

Voor de hulpvariabelen wordt gebruik gemaakt van data die over de eerste drie kwartalen van het verslagjaar beschikbaar zijn, en bovendien van de doelvariabele (en onderdelen daarvan) zelf. Daarmee lukt het een nauwkeurige nowcast te maken. Het beste model heeft de afgelopen jaren dan ook goede nowcasts geleverd. Gemiddeld zat deze er 0,3% naast (MAPE). De inhoudelijk expert vindt de nowcast ook plausibel. We adviseren dus de nowcast van deze reeks op te nemen in de Monitor 2020.

### 4.3 SES01: Beloningsverschil mannen en vrouwen

Deze indicator beschrijft hoeveel procent het gemiddelde uurloon van vrouwen verschilt van dat van mannen. Deze uurlonen op jaarbasis zijn te vinden in de volgende Statline-tabel:

<https://opendata.cbs.nl/statline/#/CBS/nl/dataset/81431ned/table?dl=FE71>

Het procentuele beloningsverschil is vervolgens te berekenen als

$$\left(1 - \frac{\text{uurloon vrouwen}}{\text{uurloon mannen}}\right) \cdot 100\%$$

In de Monitor 2019 loopt deze indicator tot en met 2017, ofwel T-2. De data is beschikbaar vanaf 2006.

Bij het begin van het jaar zijn er cijfers over de eerste acht maanden van het voorgaande jaar beschikbaar. Rond de publicatiedeadline (1 maart 2020) voor de Monitor 2020 zijn er dus cijfers tot en met augustus 2019 beschikbaar. Om een jaarcijfer voor T-1 te kunnen nowcasten construeren we daarom een hulpreeks op jaarniveau die bestaat uit de beloningsverschillen berekend op basis van telkens de eerste acht maanden per jaar. Op deze manier kan ons nowcastmodel een correlatie vaststellen tussen de eerste acht maanden en het jaarcijfer.

Het kost echter enige moeite dit voor de eerste acht maanden uit het productiesysteem te halen. Daarom gebruiken we hiervoor als proxy direct beschikbare Statline-gegevens. Het uurloon wordt

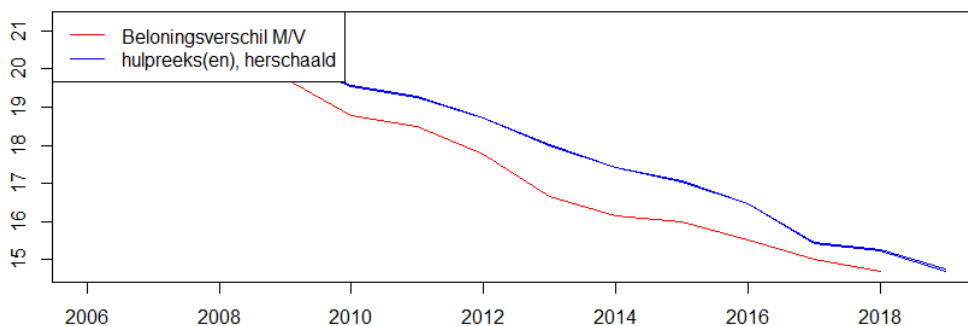
berekend door de som van het brutoloon van alle werknemersbanen in de betreffende populatie te delen door de som van de bijbehorende arbeidsduur. Als proxy gebruiken we:

$$\text{loon per vte} = \text{maandloon} * \text{aantal banen} / \text{arbeidsvolume}$$

Dit zijn allen cijfers die op maandniveau beschikbaar zijn t/m augustus, en te vinden op:

<http://opendata.cbs.nl/statline/#/CBS/nl/dataset/83451NED/table?dl=30741>

Met deze gegevens berekenen we een beloningsverschil per maand, en vervolgens middelen we over de eerste acht maanden per jaar. We berekenen hierbij een variant op basis van seizoengecorrigeerde reeksen, en één op basis van de ongecorrigeerde reeksen. Dit blijkt echter weinig verschil te maken, wat ook te zien is in onderstaande figuur, waar beide hulpreeksen nagenoeg gelijk lopen.

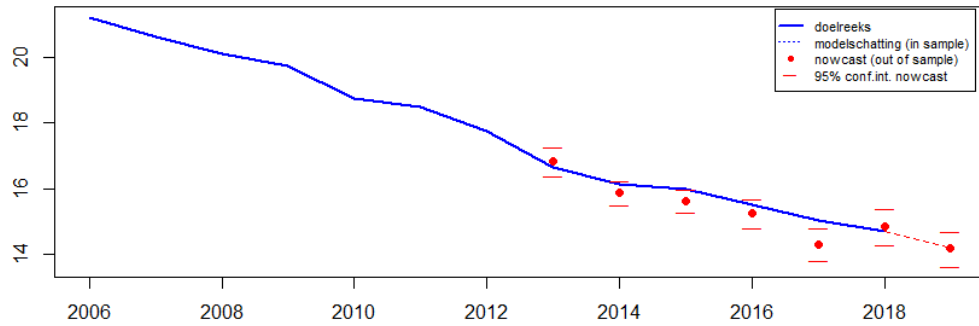


De beste tien modellen hebben de volgende kwaliteitsmaten:

RMSE	ME	MAPE	MAE	Model	Hulpreeksen
0,361148	-0,09456	1,891429	0,300044	4 = Local level with deterministic slope	-
0,371018	-0,21283	2,064904	0,320369	2 = Local level model	BelVersch_tmsept_SC
0,378149	-0,20778	2,108271	0,327086	2 = Local level model	BelVersch_tmsept
0,378856	-0,21674	2,106766	0,330473	4 = Local level with deterministic slope	BelVersch_tmsept_SC
0,382367	-0,08174	2,265315	0,351625	5 = Smooth trend model	BelVersch_tmsept_SC
0,383826	-0,21754	2,125627	0,333471	4 = Local level with deterministic slope	BelVersch_tmsept
0,389267	-0,08209	2,289571	0,355589	5 = Smooth trend model	BelVersch_tmsept
0,390881	-0,13117	2,049628	0,325403	6 = Local linear trend model	-
0,398143	-0,09937	2,378583	0,36904	6 = Local linear trend model	BelVersch_tmsept_SC
0,403609	-0,09999	2,404563	0,373105	6 = Local linear trend model	BelVersch_tmsept

De modellen verschillen niet veel qua kwaliteitsmaten, ook qua nowcast zijn er slechts kleine verschillen. Het beste model gebruikt geen hulpinformatie en valt daarom af. Het tweede model presteert vrijwel even goed en levert een nowcast van 14,16 (met een 95%-betrouwbaarheidsinterval van 13,61- 14,71) voor 2019. De uitkomsten van dit model zijn weergegeven in onderstaande figuur.

### Beloningsverschil M/V



### Conclusie

Voor de hulpvariabelen wordt gebruik gemaakt van data voor de eerste acht maanden van het verslagjaar, en bovendien van een proxy van de doelvariabele. Daarmee lukt het een nauwkeurige nowcast te maken. De afgelopen jaren zat deze gemiddeld 1,9% (MAPE) naast de werkelijke waarde. De inhoudelijk expert vindt de nowcast ook plausibel. We adviseren dus de nowcast van deze reeks op te nemen in de Monitor 2020.

## 5. Samenvatting en conclusies

In dit rapport zijn 44 indicatoren onderzocht, waarvoor in de Monitor Brede Welvaart & Sustainable Development Goals nog geen cijfer voor T-1 was opgenomen. De indicatoren zijn afkomstig uit de thema's R&D, Overheidsuitgaven, Armoede en Inkomen, Criminaliteit en Gezondheid.

Voor een aantal van de indicatoren is gebleken dat deze op dit moment nog niet versneld kunnen worden. Een deel van de indicatoren kan versneld worden door bijvoorbeeld het interne verwerkingsproces te versnellen door de inhoudelijk expert. Daarnaast is gekeken naar versnelling middels nowcastmodellen.

Voor de indicatoren waar de toepassing van nowcastmodellen onderzocht is, lukt het om goede schattingen te maken met kleine betrouwbaarheidsintervallen. In een analyse van historische data zouden de schattingen van deze modellen slechts relatief weinig afwijken van de later gepubliceerde waarden.

In onderstaand overzicht staan de indicatoren binnen de onderzochte thema's die versneld kunnen worden, en de wijze waarop ze al dan niet versneld kunnen worden op basis van dit onderzoek.

Code in rapport	Naam indicator	Versneld door
-	Fysieke kapitaalgoederenvoorraad	Model
-	Kenniskapitaalgoederenvoorraad	Model
-	Toegevoegde waarde MKB	Expert*
ENR13	Gewerkte uren in speur- en ontwikkelingswerk	Model
-	Schuld van de overheid	CPB / expert
-	Mediane ratio verkoopprijs/vraagprijs	Expert
-	Arbeidsinkomensquote	Expert
-	Loon per gewerkt uur	Expert
-	Gezonde levensverwachting mannen	Expert
-	Gezonde levensverwachting vrouwen	Expert
EBD01	Prijsindex uitgaven aanschaf en bezit koopwoningen	Model
SES01	Beloningsverschil mannen en vrouwen	Model

\*deze indicator kan versneld worden van T-3 naar T-2.

## Verklaring van tekens

Niets (blanco)	Een cijfer kan op logische gronden niet voorkomen
.	Het cijfer is onbekend, onvoldoende betrouwbaar of geheim
*	Voorlopige cijfers
**	Nader voorlopige cijfers
2019-2020	2019 tot en met 2020
2019/2020	Het gemiddelde over de jaren 2019 tot en met 2020
2019/'20	Oogstjaar, boekjaar, schooljaar enz., beginnend in 2019 en eindigend in 2020
2017/'18-2019/'20	Oogstjaar, boekjaar, enz., 2017/'18 tot en met 2019/'20

In geval van afronding kan het voorkomen dat het weergegeven totaal niet overeenstemt met de som van de getallen.

## Colofon

### *Uitgever*

Centraal Bureau voor de Statistiek  
Henri Faasdreef 312, 2492 JP Den Haag  
[www.cbs.nl](http://www.cbs.nl)

### *Prepress*

Centraal Bureau voor de Statistiek

### *Ontwerp*

Edenspiekermann

### *Inlichtingen*

Tel. 088 570 70 70  
Via contactformulier: [www.cbs.nl/infoservice](http://www.cbs.nl/infoservice)

© Centraal Bureau voor de Statistiek, Den Haag/Heerlen/Bonaire, 2020.  
Verveelvoudigen is toegestaan, mits het CBS als bron wordt vermeld.