



Centraal Bureau
voor de Statistiek

Rapport

Modelleren van methodebreuken in maandcijfers over de beroepsbevolking – resultaten op nationaal niveau

Jan van den Brakel en Sabine Krieg

Samenvatting: De maandelijkse schattingen over de beroepsbevolking zijn gebaseerd op de Enquête Beroepsbevolking. Deze enquête is gebaseerd op een roterend paneldesign. Hierbij worden de responderende huishoudens vijf keer met tussenpozen van drie maanden ondervraagd over hun arbeidssituatie. Sinds juni 2010 wordt een structureel tijdreeksmodel gebruikt om maandcijfers over de beroepsbevolking te schatten.

De waarnemingstrategie van de EBB is sindsdien twee keer gewijzigd, zowel in 2010 als ook in 2012. Om het effect van deze wijzigingen op de uitkomsten in kaart te brengen, zijn de nieuwe en de oude opzet zes maanden lang parallel aan elkaar uitgevoerd. Een dergelijk paralleltraject heeft voor de eerste peiling in 2010 en voor de eerste en tweede peiling in 2012 plaatsgevonden. De wijzigingen leiden tot hogere directe schattingen voor de werkloosheid. De schatting voor de totale werkloze beroepsbevolking neemt ten gevolge van de veranderingen in de waarneming toe met circa 50.000 personen. Deze systematische effecten maken het noodzakelijk om het tijdreeksmodel aan te passen. Dit gebeurt door het opnemen van interventievariabelen waarbij tevens gebruik wordt gemaakt van de informatie uit beide paralleltrajecten.

In dit document wordt de schattingsmethodiek voor de maandcijfers over de beroepsbevolking volgens de internationale (ILO) definitie beschreven en hoe deze is aangepast om met de wijzigingen ten gevolge van de veranderingen in de waarneming rekening te houden.

Trefwoorden: Herontwerp Persoonsenquêtes, beroepsbevolking, structurele tijdreeksmodellen, Kalman-filter, interventieanalyse

CBS Heerlen
CBS-weg 11
6412 EX Heerlen
Postbus 4481
6401 CZ Heerlen
+31 45 570 60 00
www.cbs.nl

projectnummer

PPM
20 oktober 2014

kennisgeving De in dit rapport weergegeven opvattingen zijn die van de auteur(s) en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek.

1. Inleiding

De Enquête Beroepsbevolking (EBB) is gebaseerd op een roterend panelontwerp. Iedere maand wordt een gestratificeerde tweetrapssteekproef van huishoudens geselecteerd. Deze worden vijf keer benaderd, steeds met tussenpozen van drie maanden. Vanaf juni 2010 worden maandcijfers over de beroepsbevolking met een multivariaat structureel tijdreeksmodel geschat (Van den Brakel en Krieg, 2009, 2010). Deze schattingen worden gemaakt voor heel Nederland en voor zes deelpopulaties (mannen en vrouwen uitgesplitst naar drie leeftijdscategorieën).

De waarneemstrategie van de EBB is gewijzigd, de eerste keer in 2010 en de tweede keer in 2012. De belangrijkste wijziging houdt in dat nu meer gebruik gemaakt wordt van telefonische waarneming en internetwaarneming en minder van de (duurdere) face-to-face waarneming. Deze wijzigingen zijn ingevoerd in het kader van het project Herontwerp Persoonsenquêtes (HPE). Ten gevolge van deze veranderingen was het ook noodzakelijk om de vragenlijst aan te passen. De veranderingen in de onderzoeksopzet hebben een systematisch effect op de uitkomsten tot gevolg. Dit verschijnsel wordt kortweg aangeduid met de term methodebreuk. De waarneemstrategie voor 2010 wordt in dit rapport het oude design genoemd. De strategie tussen juli 2010 en 2012 wordt het tussendesign genoemd. De waarneemstrategie die in 2012 is ingevoerd, wordt in dit rapport het nieuwe design genoemd.

Om de effecten van de wijzigingen in de waarneemstrategie te kwantificeren is in het eerste en tweede kwartaal van 2010 de eerste peiling van de EBB onder het tussendesign en het oude design parallel aan elkaar uitgevoerd. De tweede peiling is alleen in het tweede kwartaal van 2010 onder het tussendesign en het oude design parallel aan elkaar uitgevoerd. In 2012 heeft weer een dergelijk paralleltraject plaatsgevonden. In het tweede en derde kwartaal van 2012 is de eerste peiling van de EBB onder het tussendesign en het nieuwe design parallel aan elkaar uitgevoerd. In het derde en vierde kwartaal is ook de tweede peiling onder het tussendesign en onder het nieuwe design parallel aan elkaar uitgevoerd. De derde peiling is alleen in het vierde kwartaal zowel onder het tussendesign als ook onder het nieuwe design parallel aan elkaar uitgevoerd. De vierde en vijfde peiling zijn respectievelijk in januari en april 2013 overgegaan op het nieuwe design. Deze overgangen zijn zonder paralleltraject uitgevoerd. De wijzigingen in de waarneemstrategie hebben systematische effecten op de uitkomsten. Om modelmisspecificatie te voorkomen moet het tijdreeksmodel worden aangepast bij de invoering van zowel het tussendesign in 2010 en het nieuwe design in 2012. Deze aanpassingen worden in dit rapport beschreven. Vanaf januari 2015 staan de cijfers over de beroepsbevolking gebaseerd op de internationale definitie van Eurostat (ILO definitie) centraal in de maandelijkse publicaties. De cijfers in dit rapport zijn gebaseerd op deze definitie.

Het rapport is als volgt opgebouwd. In paragraaf 2 worden de veranderingen in de waarneemstrategie van de EBB kort samengevat. Daarnaast wordt het paralleltraject beschreven, dat uitgevoerd wordt om de effecten van deze veranderingen te kwantificeren. In paragraaf 3 wordt beschreven hoe het tijdreeksmodel aangepast wordt om voor de breuk te corrigeren. De resultaten van het paralleltraject komen in paragraaf 4 aan de orde. Daarna volgen in paragraaf 5 de resultaten van de breukcorrectie. Het rapport wordt afgerond met een conclusie in paragraaf 6.

2. Veranderingen in de waarneemstrategie van de EBB

De EBB is sinds 2001 gebaseerd op een roterend panelontwerp. Huishoudens worden in totaal vijf keer benaderd. Voor het HPE van 2010 werden de huishoudens in de steekproef de eerste keer door een interviewer thuis benaderd, die de elektronische vragenlijst via een persoonlijk vraaggesprek afneemt. Deze manier van data verzamelen wordt aangeduid met Computer Assisted Personal Interviewing (CAPI). Vervolgens werden de huishoudens nog vier keer telefonisch herbenaderd, steeds met tussenpozen van een kwartaal, om een verkorte versie van de vragenlijst af te nemen. Deze manier van data verzamelen wordt aangeduid met Computer Assisted Telephone Interviewing (CATI).

De waarneemstrategie is recentelijk twee keer aangepast. In 2010 is de waarneming van de eerste peiling overgegaan van CAPI naar een mixed-mode strategie van CAPI en CATI. De steekproefadressen met bekend telefoonnummer worden telefonisch benaderd, de overige adressen aan huis. De waarnemingsmode voor de vervolgpeilingen blijft onveranderd CATI.

In 2012 is overgegaan naar een sequentieel mixed-mode design voor de eerste peiling. Hierbij worden de adressen zo veel mogelijk via internet waargenomen. Dit wordt aangeduid met Computer Assisted Web Interviewing (CAWI). Een deel van de adressen die niet via internet responderen, worden één of twee maanden later via een andere mode benaderd. Als een telefoonnummer bekend is, gebeurt dat één maand later via CATI. Als geen telefoonnummer bekend is, gebeurt dit twee maanden later via CAPI. Deze vertraging heeft te maken met de planning van het veldwerk. De eerste peiling onder de nieuwe opzet voor een zekere maand t bestaat uit:

- de CAWI respons van de steekproef die getrokken is voor maand t ,
- de CATI respons van de herbenadering van de steekproef die voor de eerste peiling getrokken is voor maand $t-1$,
- de CAPI respons van de herbenadering van de steekproef die voor de eerste peiling getrokken is voor maand $t-2$.

Bij een sequentieel mixed-mode design zou normaliter iedere adres dat niet via internet respondeert vervolgens via een andere mode benaderd worden. Bij de EBB is ervoor gekozen deze herbenadering maar bij een deel van de adressen te doen. Hierdoor is het werkaanbod voor de interviewers van de afdeling dataverzameling stabiel. Bovendien wordt hierdoor een relatief groot deel van de huishoudens via de goedkoopste mode (internet) waargenomen. In 2012 is de netto-steekproefgrootte voor de eerste peiling verhoogd. Dit is gedaan om voor de grotere paneluitval onder het nieuwe design te compenseren: huishoudens die in de eerste peiling via internet gerespondeerd hebben, zijn vaak niet bereid om aan de vervolgpeilingen deel te nemen. Het is de bedoeling dat de totale netto-steekproefgrootte per maand over alle peilingen samen ongeveer constant blijft. De waarnemingsmode voor de vervolgpeilingen blijft onveranderd CATI.

Ten slotte is zowel in 2010 als ook in 2012 de vragenlijst aangepast om deze geschikt te maken voor CATI-waarneming en uiteindelijk ook internet-waarneming. Om de vragenlijst voor de eerste peiling geschikt te maken voor CATI-waarneming zijn in 2010 een aantal vragenblokken

uit de eerste peiling doorgeschoven naar de tweede en derde peiling. Tegelijk zijn ook andere aanpassingen in de vragenlijst geweest, de belangrijkste aanpassing is de introductie van de basisvragenlijst¹ in 2010.

Deze veranderingen hebben tot gevolg dat methodebreuken optreden in de eerste peiling maar ook in de vervolgpeilingen. De methodebreuken in 2010 zijn al eerder in Van den Brakel en Krieg (2012) beschreven. Dit rapport geeft een overzicht over beide breuken en beschrijft daarnaast ook de totale breuk tussen het oude en het nieuwe design. Omdat het CBS vanaf januari 2015 de cijfers over de beroepsbevolking baseert op de internationale definitie (ILO) zijn de cijfers in dit rapport gebaseerd op deze definitie.

Zoals vermeld in de inleiding is de eerste peiling gedurende 6 maanden in 2010 onder het oude design en onder het tussendesign parallel aan elkaar uitgevoerd. In 2012 was er een paralleltraject van vergelijkbare omvang voor het tussendesign en het nieuwe design voor zowel de eerste als ook de tweede peiling. Het paralleltraject in 2010 liep van januari tot en met juni. Tot en met juni zijn de maandpublicaties gebaseerd op de oude waarneming. Vanaf juli 2010 is gebruik gemaakt van de cijfers die gebaseerd zijn op het tussendesign waarbij in de input voor het tijdreeksmodel voor de eerste peiling de cijfers onder het oude design vanaf januari 2010 zijn vervangen door de cijfers onder het tussendesign. De tweede, derde, vierde en vijfde peiling zijn overgegaan in respectievelijk april 2010, juli 2010, oktober 2010 en januari 2011.

Het paralleltraject voor de overgang van het tussendesign naar het nieuwe design liep in het tweede en derde kwartaal van 2012 voor de eerste peiling. Het doorlopen van het sequentiële mixed-mode design voor één maandsteekproef duurt in totaal drie maanden. Daarom is het paralleltraject in februari 2012 gestart, zodat vanaf april 2012 de waarneming voor de eerste peiling is gebaseerd op de volledige mix van internet, CAPI en CATI. De tweede peiling wordt vanaf juli onder het nieuwe design waargenomen, de derde peiling vanaf oktober. De waarneming voor het tussendesign voor de eerste peiling liep tot en met september, voor de tweede en derde peiling tot en met december. Voor de publicaties van de maandcijfers werd tot en met september volledig gebruik gemaakt van de reeksen gebaseerd op het tussendesign, vanaf oktober worden de schattingen zoveel mogelijk gebaseerd op het nieuwe design. Vanaf dat moment zijn de cijfers in de tijdreeks van de eerste peiling vanaf april 2012 gebaseerd op het nieuwe design. De reeksen in de tweede, derde, vierde en vijfde peiling gaan respectievelijk in juli 2012, oktober 2012, januari 2013 en april 2013 over op het nieuwe design. Het verloop van beide paralleltraject is schematisch weergegeven in figuur 1.

Via het eerste paralleltraject kan een betrouwbare directe schatting voor de methodebreuken voor de eerste peiling worden verkregen. Via het tweede paralleltraject kan een betrouwbare directe schatting voor de methodebreuken voor zowel de eerste als ook de tweede peiling worden verkregen. Voor de tweede peiling (in 2010) en voor de derde peiling (in 2012) hebben paralleltrajecten van drie maanden plaatsgevonden. Dit is niet voldoende om betrouwbare directe schattingen voor de methodebreuk te berekenen. Deze data zijn wel gebruikt voor de kwartaal- en jaarcijfers. Voor de overige peilingen kunnen geen directe schattingen voor de breuken berekend worden via het paralleltraject.

¹ De zogenaamde basisvragenlijst is een set van vragen die in alle persoons- en huishoudensenquêtes van het CBS gesteld wordt. Hierdoor worden de analysemogelijkheden vergroot.

Voor een gedetailleerdere beschrijving van de veranderingen die in de onderzoeksofzet van de EBB zijn doorgevoerd en de ofzet van het paralleltraject om methodebreuken te kwantificeren wordt verwezen naar Van den Brakel en Van Berkel (2012).

Uitgangspunt bij het HPE is om de maandcijfers die in 2010 tot en met 2012 gepubliceerd worden, te corrigeren voor de waargenomen methodebreuken. Daardoor blijven de cijfers in deze periode op het niveau van de oude waarnemingstrategie. Er is uiteindelijk voor gekozen om de overgang naar het niveau van het nieuwe design samen te laten vallen met de overgang naar de internationale definities. Vanaf januari 2015 zijn de maandcijfers over de beroepsbevolking gebaseerd op de internationale definitie op het niveau van het nieuwe design. Tot en met eind 2014 zijn de cijfers gecorrigeerd naar het niveau van het oude design.

Figuur 1: Waarneming EBB naar ofzet en periode

Opzet	Periode																	
	2009		2010				2011				2012				2013			
	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	
A1	■	■	■	■														
A2	■	■	■	■														
A3	■	■	■	■														
A4	■	■	■	■	■													
A5	■	■	■	■	■	■												
B1			■	■	■	■	■	■	■	■	■	■	■	■				
B2				■	■	■	■	■	■	■	■	■	■	■	■			
B3					■	■	■	■	■	■	■	■	■	■	■			
B4						■	■	■	■	■	■	■	■	■	■			
B5							■	■	■	■	■	■	■	■	■	■	■	■
C1													■	■	■	■	■	■
C2														■	■	■	■	■
C3															■	■	■	■
C4																■	■	■
C5																	■	■
#p	5	5	6	7	5	5	5	5	5	5	5	6	7	7	5	5	5	

Het cijfer achter de letter van de ofzet geeft het peilingnummer weer.

In de laatste rij staat het aantal peilingen dat in het betreffende kwartaal wordt waargenomen.

3. Aanpassingen in het structurele tijdreeksmodel

De schattingsmethodiek bij de EBB is grotendeels gebaseerd op de gegeneraliseerde regressie (GREG) schatter (Särndal e.a., 1992). De steekproefomvang van de EBB is echter te klein om met deze schattingsmethodiek op maandbasis voldoende nauwkeurige schattingen te maken. Om voldoende nauwkeurige schattingen over de beroepsbevolking op maandbasis te maken is een structureel tijdreeksmodel ontwikkeld, dat in deze paragraaf beschreven wordt. Het model leent informatie uit het verleden en verbetert zo de precisie van de GREG-schattingen, het houdt rekening met de autocorrelatie tussen de peilingen die ontstaat door het roterend panel ontwerp, corrigeert voor systematische verschillen tussen de peilingen en houdt rekening met de breuken in 2010 en in 2012 ten gevolge van het herontwerp. Het hier beschreven model is een uitbreiding van het in Van den Brakel en Krieg (2010) ontwikkelde model, dat geen rekening houdt met de breuken. In Van den Brakel en Krieg (2012) is een versie van het model beschreven die alleen voor de methodebreuk in 2010 corrigeert. Voor de methodiek van methodebreukschatting en –correctie via een structureel tijdreeksmodel zie Van den Brakel en Roels (2010).

Laat \hat{Y}_t^j de GREG-schatting zijn voor maand t voor de onbekende populatieparameter θ_t , gebaseerd op de steekproef waargenomen in peiling j . Iedere maand worden in vijf peilingen data verzameld zodat in elke maand t een vector $\hat{\mathbf{Y}}_t = (\hat{Y}_t^1 \hat{Y}_t^2 \hat{Y}_t^3 \hat{Y}_t^4 \hat{Y}_t^5)^T$ van vijf GREG-schattingen wordt waargenomen. De tijdreeks $\hat{\mathbf{Y}}_t$, $t = 1, \dots, T$, voor de verschillende doelvariabelen wordt als volgt gemodelleerd:

$$\hat{\mathbf{Y}}_t = \mathbf{1}_5 \theta_t + \boldsymbol{\lambda}_t + \boldsymbol{\Delta}_{1,t} \mathbf{b}_{1,t} + \boldsymbol{\Delta}_{2,t} \mathbf{b}_{2,t} + \mathbf{e}_t, \quad (1)$$

met $\mathbf{1}_5$ een vijf-dimensionale vector met ieder element gelijk aan één, $\boldsymbol{\lambda}_t = (\lambda_t^1 \lambda_t^2 \lambda_t^3 \lambda_t^4 \lambda_t^5)^T$ een vector met tijdsafhankelijke componenten voor de panelvertekening, $\boldsymbol{\Delta}_{1,t} = \text{Diag}(\delta_{1,t}^1 \delta_{1,t}^2 \delta_{1,t}^3 \delta_{1,t}^4 \delta_{1,t}^5)$ een diagonaalmatrix met dummy variabelen die aangeven op welk moment de vijf peilingen overgaan van het oude design naar het tussendesign, $\mathbf{b}_{1,t} = (b_{1,t}^1 b_{1,t}^2 b_{1,t}^3 b_{1,t}^4 b_{1,t}^5)^T$ een vector met regressiecoëfficiënten die de methodebreuken van deze overgang beschrijven, en $\boldsymbol{\Delta}_{2,t} = (\delta_{2,t}^1 \delta_{2,t}^2 \delta_{2,t}^3 \delta_{2,t}^4 \delta_{2,t}^5)$ een diagonaalmatrix met dummy variabelen die aangeven op welk moment de vijf peilingen overgaan van het tussendesign naar het nieuwe design en $\mathbf{b}_{2,t} = (b_{2,t}^1 b_{2,t}^2 b_{2,t}^3 b_{2,t}^4 b_{2,t}^5)^T$ de vector met regressiecoëfficiënten die de bijbehorende methodebreuken beschrijven. Ten slotte is $\mathbf{e}_t = (e_t^1 e_t^2 e_t^3 e_t^4 e_t^5)^T$ een term voor de steekproeffouten voor de GREG-schattingen. Model (1) is een uitbreiding van Pfeffermann (1991) en wordt sinds juni 2010 gebruikt om iedere maand schattingen te maken voor de werkloze, werkzame en totale beroepsbevolking. Deze schattingen worden gemaakt op nationaal niveau en voor een uitsplitsing naar leeftijd en geslacht in zes categorieën.

De populatieparameter θ_t in (1) wordt gemodelleerd als de som van een trend, een seizoenseffect en een ruisterm:

$$\theta_t = L_t + S_t + \varepsilon_t. \quad (2)$$

Hier is L_t de stochastische trend die met het zogenaamde smooth trend model gemodelleerd wordt als:

$$\begin{aligned}
L_t &= L_{t-1} + R_{t-1}, \\
R_t &= R_{t-1} + \eta_t, \\
E(\eta_t) &= 0, \quad \text{Cov}(\eta_t, \eta_{t'}) = \begin{cases} \sigma_\eta^2 & \text{als } t = t' \\ 0 & \text{als } t \neq t'. \end{cases}
\end{aligned} \tag{3}$$

Het seizoenseffect S_t wordt met een trigonometrisch model gemodelleerd als:

$$S_t = \sum_{l=1}^6 S_{l,t}, \tag{4}$$

met

$$\begin{aligned}
S_{l,t} &= S_{l,t-1} \cos(h_l) + S_{l,t-1}^* \sin(h_l) + \omega_{l,t} \\
S_{l,t}^* &= S_{l,t-1}^* \cos(h_l) - S_{l,t-1} \sin(h_l) + \omega_{l,t}^*, \quad l = 1, \dots, 6, \\
h_l &= \frac{\pi l}{6}, \quad l = 1, \dots, 6, \\
E(\omega_{l,t}) &= E(\omega_{l,t}^*) = 0, \\
\text{Cov}(\omega_{l,t}, \omega_{l',t'}) &= \text{Cov}(\omega_{l,t}^*, \omega_{l',t'}^*) = \begin{cases} \sigma_\omega^2 & \text{als } l = l' \text{ en } t = t' \\ 0 & \text{als } l \neq l' \text{ of } t \neq t' \end{cases}, \\
\text{Cov}(\omega_{l,t}, \omega_{l',t}^*) &= 0 \text{ voor alle } l \text{ en } t.
\end{aligned} \tag{5}$$

Tenslotte is ε_t een storingsterm voor de niet verklaarbare variatie in de populatieparameter.

Deze wordt gemodelleerd als witte ruis:

$$E(\varepsilon_t) = 0, \quad \text{Cov}(\varepsilon_t, \varepsilon_{t'}) = \begin{cases} \sigma_\varepsilon^2 & \text{als } t = t' \\ 0 & \text{als } t \neq t'. \end{cases} \tag{6}$$

Roterende panel ontwerpen hebben last van panelvertekening. Hiermee wordt bedoeld dat er systematische verschillen zitten tussen de uitkomsten van de opeenvolgende peilingen. Dit fenomeen staat in de literatuur bekend als rotation group bias (Bailair, 1975). Panelvertekening wordt veroorzaakt door verschil in meetfouten tussen de peilingen ten gevolge van verschillen in vragenlijst en dataverzamelingsmethoden en paneluitval. Zie Van den Brakel en Krieg (2009, 2010) voor meer informatie over panelvertekening bij de EBB. De panelvertekening wordt in (1) met λ_t gemodelleerd. Om het model te identificeren, wordt verondersteld dat de eerste peiling niet vertekend is. Dit betekent dat de eerste component van λ_t gelijk is aan nul (zie Van den Brakel en Krieg (2009) voor een motivatie). De andere componenten beschrijven de systematische vertekening ten opzichte van de eerste peiling en worden gemodelleerd als random walks. Dit levert de volgende uitdrukkingen voor de elementen van λ_t :

$$\begin{aligned}
\lambda_t^1 &= 0, \quad \lambda_t^j = \lambda_{t-1}^j + \eta_{\lambda,j,t}, \quad j = 2, 3, 4, 5, \\
E(\eta_{\lambda,j,t}) &= 0, \quad \text{Cov}(\eta_{\lambda,j,t}, \eta_{\lambda,j',t'}) = \begin{cases} \sigma_\lambda^2 & \text{als } t = t' \text{ en } j = j' \\ 0 & \text{als } t \neq t' \text{ of } j \neq j'. \end{cases}
\end{aligned} \tag{7}$$

De breuken worden in (1) met de interventietermen $\Delta_{1,t}$ en $\Delta_{2,t}$ gemodelleerd. Dit zijn diagonaalmatrices waarbij de diagonaalelementen $\delta_{m,t}^j, j=1, 2, 3, 4, 5$, dummyvariabelen zijn die aangeven wanneer peiling j overgaat van het oude naar het tussendesign (de breuk in 2010,

$m=1$) en van het tussendesign naar het nieuwe design (de breuk in 2012, $m=2$). Dat wil zeggen, $\delta_{1,t}^1$ heeft de waarde nul zolang de eerste peiling onder het oude design wordt waargenomen en gelijk wordt aan één vanaf januari 2010, het moment dat de waarneming overgaat op het tussendesign. Drie maanden later gaat de tweede peiling over op het tussendesign zodat $\delta_{1,t}^2$ in april 2010 verandert van nul naar één. Op overeenkomstige wijze wisselen $\delta_{1,t}^3$, $\delta_{1,t}^4$ en $\delta_{1,t}^5$ van nul naar één in respectievelijk juli 2010, oktober 2010 en januari 2011. De termen $\delta_{2,t}^1$, $\delta_{2,t}^2$, $\delta_{2,t}^3$, $\delta_{2,t}^4$ en $\delta_{2,t}^5$ voor de breuk in 2012 wisselen van waarde in april 2012, juli 2012, oktober 2012, januari 2013 en april 2013.

De regressiecoëfficiënten die horen bij de dummyvariabelen $\delta_{m,t}^j$, dat wil zeggen $\mathbf{b}_{1,t}$ en $\mathbf{b}_{2,t}$ beschrijven het systematische effect op de uitkomsten in de tijdreeksen van de verschillende peilingen en kunnen worden geïnterpreteerd als de omvang van de breuken. De omvang van de breuk in 2010 voor de eerste peiling en de omvang van de breuken in de eerste en tweede peiling in 2012 worden geschat op basis van het paralleltraject. Naar aanleiding van de resultaten van het paralleltraject in 2010 is de regressiecoëfficiënt voor de eerste peiling, $b_{1,t}^1$, tijdsafhankelijk gekozen. Gedurende het paralleltraject is de regressiecoëfficiënt gelijk aan het verschil tussen het oude design en het tussendesign in iedere afzonderlijke maand. Na afloop van het paralleltraject is de regressiecoëfficiënt gelijk aan het gemiddelde verschil over de zes maanden:

$$b_{1,t}^1 = \begin{cases} \hat{Y}_t^{1,tussen} - \hat{Y}_t^{1,oud} & \text{als } t \in [\tilde{t}, \dots, \tilde{t} + 5] \\ \frac{1}{6} \sum_{t=\tilde{t}}^{\tilde{t}+5} (\hat{Y}_t^{1,tussen} - \hat{Y}_t^{1,oud}) & \text{als } t > \tilde{t} + 5 \end{cases}$$

waarbij \tilde{t} overeenkomt met de maand waarin het paralleltraject gestart is (januari 2010 voor de eerste peiling van het eerste paralleltraject). Zie Van den Brakel en Krieg (2012) voor een motivatie voor deze keuze. Dezelfde keuze is ook voor de regressiecoëfficiënten van de eerste en tweede peiling in het tweede paralleltraject in 2012 gemaakt:

$$b_{2,t}^j = \begin{cases} \hat{Y}_t^{j,nieuw} - \hat{Y}_t^{j,tussen} & \text{if } t \in [\tilde{t}, \dots, \tilde{t} + 5] \\ \frac{1}{6} \sum_{t=\tilde{t}}^{\tilde{t}+5} (\hat{Y}_t^{j,nieuw} - \hat{Y}_t^{j,tussen}) & \text{if } t > \tilde{t} + 5 \end{cases}, \quad j = 1, 2.$$

Voor de overige regressiecoëfficiënten $b_{1,t}^j$ $j=2, 3, 4, 5$ en $b_{2,t}^j$ $j=3, 4, 5$, is geen of onvoldoende informatie uit het paralleltraject beschikbaar om een schatting voor de omvang van de breuk te maken. Deze worden daarom geschat op basis van het tijdreeksmodel en worden om deze reden tijdsafhankelijk verondersteld.

Het tijdreeksmodel voor de steekproeffouten \mathbf{e}_t in (1) houdt rekening met veranderingen in de variantie van de steekproeffout, waarbij deze veranderingen in variantie veroorzaakt worden door veranderingen in de steekproefomvang en in het steekproefontwerp. Verder wordt rekening gehouden met de autocorrelatie in de steekproeffouten ten gevolge van het roterende panelontwerp. De varianties van de steekproeffouten worden als prior informatie aan het tijdreeksmodel meegegeven via het volgende meetfoutmodel (Binder en Dick, 1990):

$$e_t^j = k_t^j \tilde{e}_t^j, \quad j = 1, 2, 3, 4, 5, \quad (8)$$

met k_t^j de standaardfout van de GREG-schattingen van de j -de peiling in periode t . De steekproeffout in de eerste peiling is niet gecorreleerd met steekproeffouten waargenomen in

het verleden. Hieruit volgt dat de term voor de eerste peiling \tilde{e}_t^1 wordt gemodelleerd als witte ruis met $E(\tilde{e}_t^1) = 0$ en $Var(\tilde{e}_t^1) = \sigma_{e_1}^2$. Merk op dat $Var(e_t^1) = (k_t^1)^2 \sigma_{e_1}^2$. Indien de modelschatting voor $\sigma_{e_1}^2$ gelijk is aan één, is $Var(e_t^1)$ gelijk aan de variantie van de GREG-schattingen.

De autocorrelatie voor de overige peilingen kan worden beschreven met het volgende AR(1)-model (Van den Brakel en Krieg, 2009):

$$\tilde{e}_t^j = \rho \tilde{e}_{t-3}^{j-1} + v_t^j$$

$$E(v_t^j) = 0, \quad Cov(v_t^j, v_{t'}^j) = \begin{cases} \sigma_{e_j}^2 & \text{als } t = t', \quad j = 2,3,4,5, \\ 0 & \text{als } t \neq t'. \end{cases} \quad (9)$$

Omdat voor \tilde{e}_t^j een AR(1)-proces wordt aangenomen, geldt dat $Var(\tilde{e}_t^j) = \sigma_{e_j}^2 / (1 - \rho^2)$. Indien de schatting voor $\sigma_{e_j}^2$ gelijk is aan $(1 - \rho^2)$, dan geldt voor de vervolgpelingen dat $Var(e_t^j)$ gelijk is aan de variantie van de GREG-schattingen. Omdat in de programmatuur $\sigma_{e_j}^2$ wordt geschaald met een factor $(1 - \rho^2)$, zeg $\tilde{\sigma}_{e_j}^2$, is $Var(e_t^j)$ gelijk aan de variantie van de GREG-schattingen, indien de modelschattingen voor $\tilde{\sigma}_{e_j}^2$ gelijk zijn aan één. De autocorrelatiecoëfficiënt ρ wordt geschat uit de GREG-schattingen per peiling per maand via de procedure van Pfeffermann et al. (1998).

Een AR(1)-proces voor de steekproeffouten houdt rekening met de autocorrelatie die ontstaat doordat de opeenvolgende peilingen voor een groot deel gebaseerd zijn op dezelfde steekproefeenheden. Een grote positieve waarde voor ρ wordt gevonden als de steekproefoverlap groot is en de arbeidspositie tussen de peilingen niet snel verandert. Vergelijking (9) veronderstelt dezelfde autocorrelatiecoëfficiënt voor alle peilingen. Daarnaast wordt de autocorrelatie constant in de tijd verondersteld, ook als het design verandert. De eerste veronderstelling wordt geschonden doordat de paneluitval tussen de eerste en de tweede peiling groter is dan tussen de overige vervolgpelingen. Hierdoor is de steekproefoverlap tussen eerste en tweede peiling kleiner dan bijvoorbeeld tussen de tweede en derde peiling. Met de introductie van het nieuwe design lijkt de paneluitval tussen de eerste en de tweede peiling groter te worden. Dit betekent dat ook de tweede veronderstelling (autocorrelatie constant in de tijd) geschonden wordt. De steekproefoverlap tussen de vervolgpelingen blijkt nauwelijks te veranderen onder het nieuwe design.

Vanwege deze veranderingen zou het model aangepast kunnen worden. Hiervoor zou een methode ontwikkeld moeten worden die verschillende autocorrelatiecoëfficiënten voor de verschillende peilingen schat en ook verschillende autocorrelatiecoëfficiënten voor de periodes voor en na de breuk van 2012. Omdat de autocorrelatiecoëfficiënten in de praktijk een vrij kleine invloed hebben op de modelschattingen, is het niet urgent om het model op dit onderdeel aan te passen.

Om het tijdreeksmodel te schatten wordt het als toestands-ruimte model geschreven. Dan kunnen met behulp van het Kalman-filter optimale schattingen voor de modelvariabelen berekend worden. De zogenaamde gefilterde schattingen zijn de schattingen voor een bepaalde periode t die gebaseerd zijn op de tijdreeks tot en met periode t . Deze schattingen kunnen verbeterd worden door gebruik te maken van de informatie die na periode t beschikbaar komt. De schattingen die gebaseerd zijn op de volledige reeks worden de gesmoothte schattingen genoemd. In dit paper wordt met gefilterde schattingen gewerkt omdat dit de schattingen zijn

gebaseerd op de informatie die in het productieproces beschikbaar is om maandcijfers over de beroepsbevolking te publiceren. Via het Kalman-filter worden ook standaardfouten voor de modelschattingen berekend. Voor details over toestands-ruimtemodellen en het Kalman-filter zie Harvey (1989) of Durbin en Koopman (2001). De schattingen worden berekend met het softwarepakket Ox in combinatie met Ssfpack 3.0, zie Doornik (2007) en Koopman e.a. (1999, 2008).

4. Resultaten paralleltraject

4.1 Respons tijdens paralleltrajecten in 2010 en 2012, eerste peiling

In de periode januari 2010 tot en met juni 2010 is de eerste peiling van de EBB zowel volgens het oude als ook volgens het tussendesign waargenomen. Huishoudens waarvoor geen telefoonnummer bekend is, zijn onder beide designs via CAPI waargenomen; de overige huishoudens zijn volgens het oude design via CAPI en volgens het tussendesign via CATI waargenomen. Tabel 4.1 laat de respons en non-respons voor dit paralleltraject zien. In totaal is het responspercentage onder het tussendesign lager dan onder het oude design. Vooral in de groep met bekend telefoonnummer gaat de respons omlaag. Dit wordt vooral veroorzaakt door een toename van het aantal kaderfouten (kaderfouten bij CATI hangen vaak samen met niet-correcte telefoonnummers), en het aantal huishoudens waarmee geen contact gelegd kon worden. Het aantal niet benaderde huishoudens wordt echter nihil. Zoals verwacht verandert in de groep zonder bekend telefoonnummer het responspercentage nauwelijks, daar is immers ook weinig in de waarneemstrategie veranderd.

Tabel 4.1: respons eerste peiling, oud design en tussendesign, paralleltraject 2010

Resultaat	Oud design				totaal	
	CAPI - telefoonnummer bekend		CAPI - telefoonnummer onbekend		Aantal huishoudens	%
	Aantal huishoudens	%	Aantal huishoudens	%		
Uitgezette steekproef	20813	100,0%	14469	100,0%	35282	100,0%
Kaderfouten	769	3,7%	1039	7,2%	1808	5,1%
Niet benaderd	618	3,0%	463	3,2%	1081	3,1%
Taalbarrière	390	1,9%	878	6,1%	1268	3,6%
Weigering	4909	23,6%	3112	21,5%	8021	22,7%
Geen contact	889	4,3%	1455	10,1%	2344	6,6%
Ander non-respons	921	4,4%	689	4,8%	1610	4,6%
Volledige respons	12317	59,2%	6833	47,2%	19150	54,3%

Resultaat	Tussendesign				totaal	
	CATI - telefoonnummer bekend		CAPI - telefoonnummer onbekend		Aantal huishoudens	%
	Aantal huishoudens	%	Aantal huishoudens	%		
Uitgezette steekproef	20234	100,0%	13345	100,0%	33579	100,0%
Kaderfouten	1539	7,6%	982	7,4%	2521	7,5%
Niet benaderd	1	0,0%	428	3,2%	429	1,3%
Taalbarrière	317	1,6%	788	5,9%	1105	3,3%
Weigering	4545	22,5%	2903	21,8%	7448	22,2%
Geen contact	2233	11,0%	1333	10,0%	3566	10,6%
Ander non-respons	963	4,8%	641	4,8%	1604	4,8%
Volledige respons	10636	52,6%	6270	47,0%	16906	50,3%

De resultaten voor de eerste peiling uit het tweede paralleltraject worden in tabel 4.2 getoond. De respons onder het tussendesign is ten opzichte van de respons in het paralleltraject van 2010 licht gestegen. In het tweede deel van tabel 4.2 wordt de respons onder het nieuwe design weergegeven. In de eerste stap worden alle huishoudens via CAWI benaderd. Een klein deel van deze huishoudens (2420 van in totaal 66750) kan vrij snel als non-respons afgehandeld worden. Hierbij gaat het om kaderfouten, huishoudens die weigeren mee te doen of niet kunnen responderen en om onvolledig ingevulde vragenlijsten.

Tabel 4.2: respons eerste peiling, tussendesign en nieuw design, paralleltraject 2012

Resultaat	Tussendesign					
	CATI - telefoonnummer bekend		CAPI - telefoonnummer onbekend		totaal	
	Aantal huishoudens	%	Aantal huishoudens	%	Aantal huishoudens	%
Uitgezette steekproef	17130	100,0%	18157	100,0%	35287	100,0%
Kaderfouten	1728	10,1%	1101	6,1%	2829	8,0%
Niet benaderd	1	0,0%	511	2,8%	512	1,5%
Taalbarrière	247	1,4%	1042	5,7%	1289	3,7%
Weigering	2848	16,6%	3876	21,3%	6724	19,1%
Geen contact	1868	10,9%	1760	9,7%	3628	10,3%
Ander non-respons	855	5,0%	859	4,7%	1714	4,9%
Volledige respons	9583	55,9%	9008	49,6%	18591	52,7%

Nieuw design								
	CATI		CAPI		Totaal			
	Aantal huishoudens	% van uitzet CATI	Aantal huishoudens	% van uitzet CAPI	Aantal huishoudens	% van uitzet CATI/CAPI	% van uitzet CAWI/ CATI/CAPI (*)	
Uitzet CAWI	66750							
Non-respons CAWI	2420							
Respons CAWI	16686							
Niet herbenaderd	22072							
Uitzet CATI/CAPI	11806	100,0%	13791	100,0%	25597	100,0%	71,4%	
Kaderfouten	1461	12,4%	874	6,3%	2335	9,1%	6,5%	
Niet benaderd	-	0,0%	404	2,9%	404	1,6%	1,1%	
Taalbarrière	203	1,7%	979	7,1%	1182	4,6%	3,3%	
Weigering	3020	25,6%	4138	30,0%	7158	28,0%	20,0%	
Geen contact	1564	13,2%	1368	9,9%	2932	11,5%	8,2%	
CAWI-non-respons	-	-	-	-	-	-	3,6%	
Ander non-respons	823	7,0%	738	5,4%	1561	6,1%	4,4%	
Volledige respons	4735	40,1%	5290	38,4%	10025	39,2%	53,0%	

(*) fictieve percentages als er niet uitgedund zou worden en de feitelijk uitgedunde huishoudens op dezelfde manier verdeeld zijn over de respons- en non-responscategorieën als de feitelijk uitgezette huishoudens

Een aanzienlijk deel respondeert via CAWI (16686 huishoudens). Van de overige huishoudens wordt slechts een deel herbenaderd via CATI of via CAPI. In het onderste deel van tabel 4.2 worden de resultaten van de CATI- en CAPI-benadering getoond. In kolom 3, 5 en 7 worden de percentages berekend van respectievelijk CATI, CAPI en CATI/CAPI ten opzichte van de CAWI-non-respons die voor herbenadering via CATI en CAPI is uitgezet (respectievelijk CATI, CAPI en CATI + CAPI). Het responspercentage is lager dan onder het oude design en het tussendesign, vooral het percentage weigeringen neemt sterk toe. Dit ligt in de lijn der verwachtingen, het gaat immers om een speciale groep huishoudens die al via CAWI benaderd zijn maar nog niet gerespondeerd hebben. In deze responspercentages zijn de respondenten uit de CAWI-benadering niet meegenomen. Het totale responspercentage is 40%. Ruim de helft van de 60% niet responderende huishoudens is alleen via internet benaderd, het responspercentage is dus vooral zo laag omdat niet alle huishoudens via CATI of CAPI herbenaderd worden. Een totaal

responspercentage dat vergelijkbaar is met de responspercentages van het tussendesign en het oude design kan berekend worden als aangenomen wordt dat de gerealiseerde respons op de steekproef die voor CATI/CAPI-herbenadering is uitgezet representatief is voor het deel dat na CAWI benadering niet meer voor herbenadering uitgezet is. Dit is gedaan in de laatste kolom van tabel 4.2. Op deze manier wordt een responspercentage van 53% gevonden. Dit is bijna gelijk aan het responspercentage van het tussendesign in 2012 gedurende het paralleltraject. Ook de verdeling van de non-respons over de belangrijkste categorieën onder het nieuwe design lijkt op de verdeling onder het tussendesign.

4.2 Respons in de vervolgpeilingen

Ook voor de tweede peiling heeft een paralleltraject plaatsgevonden en kan de respons onder het oude design en het tussendesign (in 2010) en onder het tussendesign en het nieuwe design (in 2012) met elkaar vergeleken worden. Huishoudens die in de eerste peiling responderen worden tijdens dit interview gevraagd of ze bereid zijn ook aan het panel deel te nemen. Een deel van de huishoudens weigert verdere deelname aan de EBB. De overige huishoudens worden voor de tweede peiling benaderd, waarvan een deel niet respondeert. Tabel 4.3 laat zien hoeveel huishoudens in de eerste peiling responderen, hoeveel aangeven bereid te zijn aan de tweede peiling deel te nemen en hoeveel uiteindelijk responderen in de tweede peiling. De percentages worden berekend met betrekking tot het aantal respondenten in de eerste peiling. De tabel laat zien dat de bereidheid om aan het panel deel te nemen en de respons voor de tweede peiling onder het tussendesign en het oude design vergelijkbaar is. Ruim 80% van de respondenten van de eerste peiling neemt ook deel aan de tweede peiling.

Tabel 4.3: respons tweede peiling tijdens paralleltraject

respons peiling 1 (*)	Oud design April – juni 2010		Tussendesign April – juni 2010		Tussendesign Juli – december 2012		Nieuw design Juli – december 2012	
	Aantal huis- houdens	%	Aantal huis- houdens	%	Aantal huis- houdens	%	Aantal huis- houdens	%
CAWI							16686	
CATI			4991		9583		4735	
CAPI	9220		3210		9008		5290	
totaal	9220		8201		18591		26711	
bereid tot panel								
CAWI							11623	69,7%
CATI			4829	96,8%	9122	95,2%	4353	92,2%
CAPI	8612	93,4%	2949	91,9%	8163	90,6%	4587	86,7%
totaal	8612	93,4%	7778	94,8%	17285	93,0%	20563	77,0%
respons peiling 2	7609	82,5%	6697	81,7%	15136	81,4%	16301	61,1%

()voor het oude design en het tussendesign tijdens het paralleltraject 2010 is de respons van de eerste peiling exclusief huishoudens met alleen personen van 65 jaar of ouder, omdat deze huishoudens niet doorstroomden naar het panel. Voor het tussendesign(na afloop van het paralleltraject) en het nieuwe design worden deze huishoudens wel voor het panel geworven.*

De bereidheid tot paneldeelname onder het nieuwe design is echter veel lager dan die onder de voorgaande designs. Vooral CAWI-respondenten zijn minder bereid hiertoe. Het resultaat is dat het percentage van de respondenten van de eerste peiling die ook in de tweede peiling responderen is gedaald naar 61%.

Voor de overige peilingen heeft er geen (of alleen een kort) paralleltraject plaats gevonden en is een vergelijking van de responspercentages zoals in tabel 4.3 niet mogelijk. De responspercentages voor de derde, vierde en vijfde peiling zijn bij de overgang van het oude naar het tussendesign en ook naar het nieuwe design niet sterk veranderd. Deze zijn altijd rond de 90% (in de vijfde peiling zelfs 94%), waarbij gerekend wordt ten opzichte van de respons van de voorgaande peiling.

Meer informatie over de respons bij de EBB is te vinden in Leden taskforce EBB (2010), Dataverzameling – monitoren en analyse (2010, 2011, 2013).

4.3 Vergelijking schattingen gebaseerd op oude design, tussendesign en nieuwe design in de eerste peiling

De tabellen 4.4 en 4.5 laten voor de eerste peiling de verschillen zien tussen de GREG-schattingen gebaseerd op het tussendesign en het oude design en tussen de GREG-schattingen gebaseerd op het nieuwe design en het tussendesign voor Nederland. In de eerste zes rijen zijn de verschillen per maand berekend, daarna volgt het gemiddelde van de waargenomen verschillen over de zes maanden. In de achtste rij is de gemiddelde standaardfout voor het verschil per maand berekend, en in de laatste rij de standaardfout voor het gemiddelde verschil over zes maanden.

De verschillen tussen het tussendesign en het oude design zijn vooral voor de werkloze beroepsbevolking vrij groot. Voor Nederland is het verschil gemiddeld ruim 17% van de doelvariabele, waarbij is gepercenteerd op het niveau van het oude design. Hierbij is de schatting gebaseerd op het tussendesign (mixed mode CATI en CAPI) groter dan die gebaseerd op het oude design (alleen CAPI).

Voor de werkzame en totale beroepsbevolking zijn de absolute verschillen tussen het tussendesign en het oude design in dezelfde orde van grootte als voor de werkloze beroepsbevolking, relatief ten opzichte van de doelvariabele zijn de verschillen duidelijk kleiner. De verschillen tussen het nieuwe design en het tussendesign zijn voor de werkloze beroepsbevolking duidelijk kleiner dan die tussen het oude design en het tussendesign. Voor de werkzame en totale beroepsbevolking zijn de verschillen tussen het nieuwe design en het tussendesign in dezelfde orde van grootte als de verschillen tussen het tussendesign en het oude design. De schattingen voor de werkzame en totale beroepsbevolking op basis van het tussendesign, zijn lager ten opzichte van het oude design. De schattingen voor de werkzame en totale beroepsbevolking op basis van het nieuwe design zijn doorgaans hoger ten opzichte van het tussendesign.

Het gemiddelde verschil tussen het nieuwe design en het tussendesign over 6 maanden is voor geen enkele variabele significant (5%-significantieniveau); de standaardfout is echter vrij groot, de verschillen kunnen dus alleen met vrij grote betrouwbaarheidsmarges geschat worden. Zoals al in Van den Brakel en Krieg (2012) is opgemerkt, schommelt het verschil door de maanden

heen sterk. Het kan hierbij gaan om gewone steekproeffluctuaties, om een verschil in seizoenspatroon tussen de designs, om onregelmatigheden in het veldwerk gedurende de uitvoering van beide paralleltrajecten in de verschillende maanden of een combinatie van deze oorzaken.

Tabel 4.6 toont de schattingen voor de methodebreuken tussen het nieuwe en het oude design, berekend als som van het verschil tussen het tussendesign en het oude design en het verschil tussen het nieuwe design en het tussendesign. Alleen voor de werkloze beroepsbevolking voor heel Nederland is een significante methodebreuk gevonden.

Tabel 4.4: Verschillen tussen de GREG-schattingen gebaseerd op het tussendesign en het oude design voor werkloze, werkzame en totale beroepsbevolking (ILO definitie), eerste peiling (x 1000), vet: significant verschil

		Werkloze BV	Werkzame BV	Totale BV
Vershil per maand	Januari	26	-54	-28
	Februari	36	-83	-47
	Maart	72	-103	-31
	April	24	-148	-125
	Mei	141	-145	-4
	Juni	108	-47	61
Gemiddeld verschil		68	-97	-29
Gemiddelde standaardfout per maand		41	86	84
Standaardfout gemiddeld verschil		17	35	34

Tabel 4.5: Verschillen tussen de GREG-schattingen gebaseerd op het nieuwe design en het tussendesign voor werkloze, werkzame en totale beroepsbevolking (ILO definitie), eerste peiling (x 1000), vet: significant verschil

		Werkloze BV	Werkzame BV	Totale BV
Vershil per maand	Januari	3	11	14
	Februari	-3	-24	-27
	Maart	-15	89	74
	April	-96	162	66
	Mei	14	-141	-126
	Juni	-13	-3	-16
Gemiddeld verschil		-19	16	-3
Gemiddelde standaardfout per maand		42	81	79
Standaardfout gemiddeld verschil		17	33	32

Tabel 4.6: Breukschattingen eerste peiling: verschil tussen nieuwe design en oude design, berekend op basis van paralleltrajecten 2010 en 2012 (x 1000), vet: significant verschil

	<i>Vershil</i>	<i>Standaardfout</i>
Werkloze beroepsbevolking	49	24
Werkzame beroepsbevolking	-81	48
Totale beroepsbevolking	-32	47

4.4 Vergelijking schattingen gebaseerd op tussendesign en nieuwe design, tweede peiling

Tabel 4.7 laat de verschillen zien tussen de GREG-schattingen gebaseerd op het tussendesign en het nieuwe design voor de tweede peiling. De verschillen voor de werkloze beroepsbevolking in de tweede peiling zijn vergelijkbaar met de waargenomen verschillen in de eerste peiling. Voor de werkzame beroepsbevolking is het verschil in de tweede peiling negatief in plaats van positief. De verschillen zijn echter niet significant. Voor de totale beroepsbevolking is het verschil in de tweede peiling toegenomen, maar is nog steeds niet significant ongelijk aan nul. De waarnemingsmethodiek voor de tweede peiling is niet gewijzigd. Mogelijke verklaringen voor waargenomen breuken in de tweede peiling zijn aanpassingen in de vragenlijst en selectie-effecten vanuit de eerste peiling die een rol blijven spelen in de vervolgpeilingen. De meeste verschillen zijn echter niet significant.

Tabel 4.7: Verschillen tussen de GREG-schattingen gebaseerd op het nieuwe design en het tussendesign voor werkloze, werkzame en totale beroepsbevolking (ILO definitie), tweede peiling (x 1000), vet: significant verschil

		Werkloze BV	Werkzame BV	Totale BV
Vershil per maand	Januari	-91	20	-71
	Februari	27	-42	-15
	Maart	-70	-4	-74
	April	15	-3	12
	Mei	10	-70	-60
	Juni	-3	-36	-38
Gemiddeld verschil		-19	-22	-41
Gemiddelde standaard fout per maand		51	97	95
Standaardfout gemiddeld verschil		21	40	39

5. Resultaten breukcorrectie

Maandelijks worden vanuit de EBB schattingen over de werkloze, werkzame en totale beroepsbevolking gepubliceerd voor heel Nederland en voor de 6 domeinen. Maandschattingen, die gecorrigeerd zijn voor de methodebreuken, worden berekend als som van de trend en het seizoenseffect, $L_t + S_t$. Deze schatting wordt ook wel het signaal genoemd. Als alternatief voor seizoensgecorrigeerde cijfers wordt de trend L_t gepubliceerd. Deze schattingen liggen op het niveau van de eerste peiling van het oude design. Schattingen op het niveau van de eerste peiling van het nieuwe design kunnen berekend worden als respectievelijk $L_t + S_t + b_1^1 + b_2^1$ en $L_t + b_1^1 + b_2^1$. De schattingen voor de twee breuken in de eerste peiling b_1^1 en b_2^1 zijn in dit geval de gemiddelden over het paralleltraject en daarom tijdsafhankelijk.

In plaats van een additive correctie, waarbij de breuken bij de trend of het signaal worden opgeteld, zou ook kunnen worden gekozen voor een zogenaamde multiplicatieve correctie. In dat geval wordt de trend of signaal vermenigvuldigd met de verhouding tussen de GREG schatting onder het nieuw en het oude design, gedurende het paralleltraject. Deze correctie impliceert dat de omvang van de methodebreuk evenredig is aan de grootte van de doelvariabele. Meer concreet:

$$\tilde{L}_t = L_t \times \frac{\bar{Y}_{GREG}^{Tussen .1}}{\bar{Y}_{GREG}^{Oud}} \times \frac{\bar{Y}_{GREG}^{Nieuw}}{\bar{Y}_{GREG}^{Tussen .2}} .$$

Hierbij is \bar{Y}_{GREG}^{Oud} de GREG-schatting voor de doelvariabele onder het oude design op basis van het paralleltraject in de eerste peiling in 2010, $\bar{Y}_{GREG}^{Tussen .1}$ de GREG-schatting voor de doelvariabele onder het tussendesign op basis van het paralleltraject in de eerste peiling in 2010, $\bar{Y}_{GREG}^{Tussen .2}$ de GREG-schatting voor de doelvariabele onder het tussendesign op basis van het paralleltraject in de eerste peiling in 2012 en \bar{Y}_{GREG}^{Nieuw} de GREG-schatting voor de doelvariabele onder het nieuwe design op basis van het paralleltraject in de eerste peiling in 2012. Een soortgelijke correctie kan worden uitgevoerd voor het signaal. Een dergelijke correctie heeft als voordeel dat gecorrigeerde schattingen nooit negatieve waarden kunnen aannemen. Er is om de volgende redenen toch voor een additive correctie gekozen. Ten eerste past een additive correctie beter bij het veronderstelde tijdreeksmodel. Bij het ontwikkelen van het tijdreeksmodel is onderzocht of multiplicatieve modellen beter passen bij de waargenomen tijdreeksen door de data logaritmisch te transformeren. Dit resulteert niet in een verbetering van het model. Ten tweede leert een vergelijking van de schattingen van de panelvertekening met de trend, dat de panelvertekening niet evenredig is met de waarde van de trend (zie Van den Brakel en Krieg (2012), figuur 4.2 en 4.3). Omdat methodebreuken en panelvertekening veroorzaakt worden door dezelfde soorten foutenbronnen ligt het voor de hand om beide vormen van vertekening op een vergelijkbare manier te modelleren en voor te corrigeren. De vergelijking in Van den Brakel en Krieg (2012) geeft geen onderbouwing voor de veronderstelling dat de omvang van methodebreuken evenredig is aan de hoogte van de doelvariabelen.

In deze paragraaf worden resultaten getoond voor de schattingen voor de werkloze, werkzame en totale beroepsbevolking voor Nederland. Figuur 5.1 laat de gefilterde trend voor de werkloze

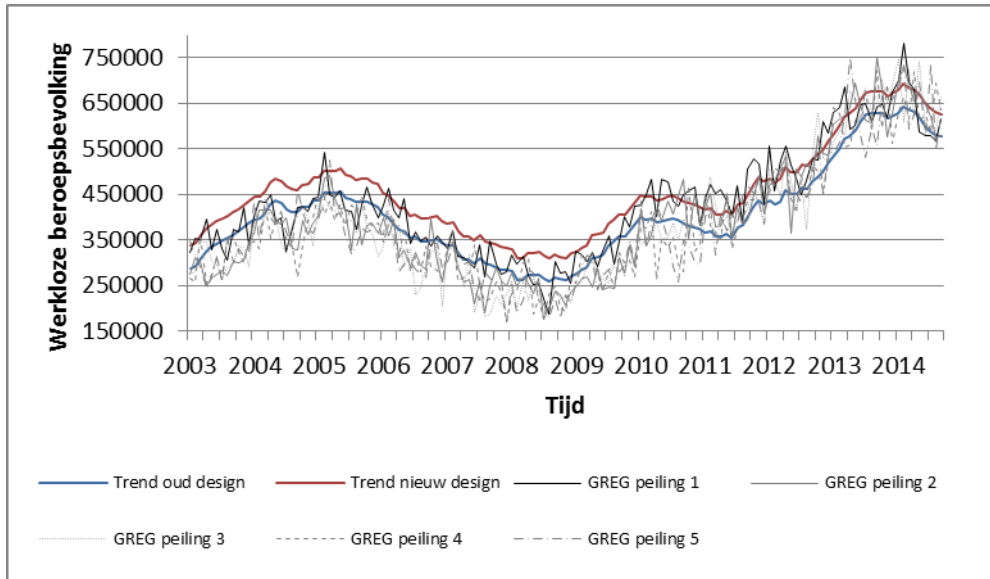
beroepsbevolking zien op het niveau van het oude design (blauwe lijn) en het niveau van het nieuwe design (rode lijn). Aan de figuur zijn ook de reeksen van de GREG-schattingen toegevoegd (zwarte lijnen). De figuur laat zien dat in de periode voor de eerste HPE-breuk de trendschattingen op het niveau van de GREG-schatting voor de eerste peiling liggen (ononderbroken zwarte lijn). Dat verandert in 2010 doordat door de breuken de inputreeksen ten opzichte van het oude design op een hoger niveau komen te liggen. De trendschattingen op het niveau van het nieuwe design liggen vanaf 2012 wel op het niveau van de GREG-schattingen voor de eerste peiling, die in 2012 op het nieuwe design gebaseerd zijn.

Figuur 5.2 laat het gefilterde signaal voor de werkloze beroepsbevolking voor Nederland zien op het niveau van het oude design (blauwe lijn) en op het niveau van het nieuwe design (rode lijn). Ook aan deze figuur zijn de reeksen van de GREG-schattingen toegevoegd (zwarte lijnen). Net als de trend ligt ook het signaal in de periode voor de eerste HPE-breuk op het niveau van de GREG-schatting voor de eerste peiling liggen, wat verandert in 2010 door de breuk. De signaalschattingen op het niveau van het nieuwe design liggen op het niveau van de GREG-schattingen voor de eerste peiling; ook dit is vergelijkbaar met het resultaat voor de trend.

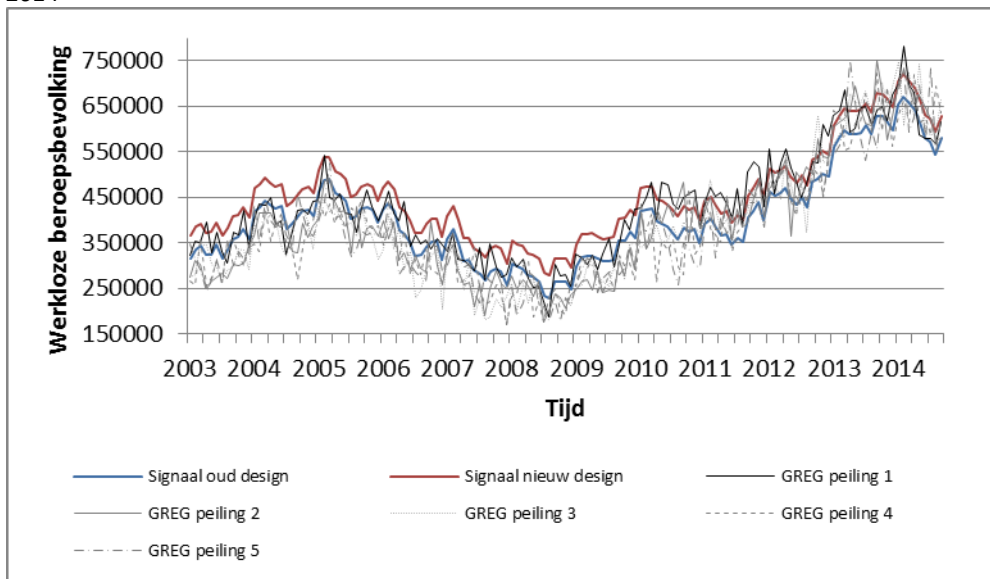
Figuur 5.3 laat de standaardfouten van de gefilterde trend van de werkloze beroepsbevolking zien. Bij de berekening van de standaardfout wordt geen rekening gehouden met de standaardfout van de directe schatting voor de breuken in de eerste en tweede peiling uit de paralleltrajecten. Hierdoor wordt de standaardfout vanaf 2010 enigszins onderschat. De standaardfouten voor de trend op het niveau van het oude en het nieuwe design zijn gelijk, omdat het verschil tussen beide trends juist deze breukschattingen zijn, waarvan de standaardfout genegeerd wordt.

De standaardfout van de gefilterde trend voor tijdstip t hangt af van de steekproeffouten van de inputreeksen (de GREG-schattingen), de lengte van de inputreeks tot tijdstip t en de complexiteit van het model. De daling van de standaardfout in de periode tot 2009 wordt veroorzaakt doordat de inputreeksen steeds langer worden en het model in deze periode niet verandert. De lichte stijging in 2009 hangt samen met de stijging van de steekproeffouten in deze periode. In 2010 is een sterke stijging van de standaardfout te zien. Dit komt doordat in deze periode de interventievariabelen voor het modeleren van de breuken in de vervolpeilingen ingeschakeld worden. Iedere keer als een vervolpeiling overgaat van het oude design naar het tussendesign, wordt een nieuwe regressiecoëfficiënt aan het model toegevoegd. Deze coëfficiënt wordt geschat op basis van de waarnemingen in de daarop volgende perioden. In het begin is de onzekerheid rondom deze schatting hoog, waardoor de standaardfout van de gefilterde trend en het signaal ook toeneemt. Nadat de eerste HPE-breuk voor alle peilingen geïntroduceerd is, daalt de standaardfout weer omdat steeds meer informatie beschikbaar komt om de nieuwe modelparameters beter te schatten. Vanaf het laatste kwartaal van 2012 is weer een stijging van de standaardfout te zien, veroorzaakt door de tweede HPE-breuk in de derde peiling. Deze toename houdt aan totdat de vijfde peiling is overgegaan op het nieuwe design.

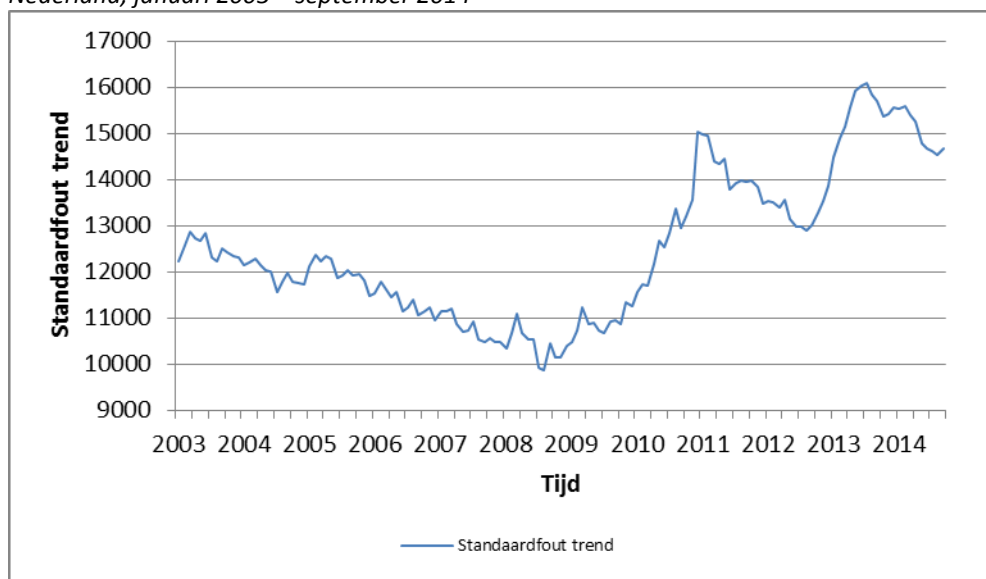
Figuur 5.1: Gefilterde trend op het niveau van het oude design en het nieuwe design en GREG-schattingen, werkloze beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



Figuur 5.2: Gefilterde signaal op het niveau van het oude design en het nieuwe design en GREG-schattingen, werkloze beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



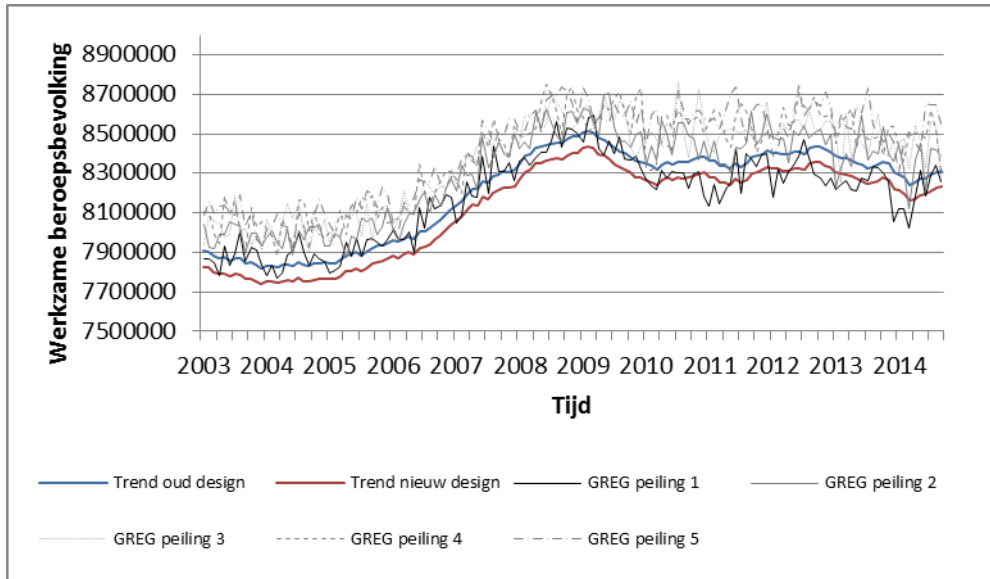
Figuur 5.3: Standaardfouten voor de gefilterde trend, werkloze beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



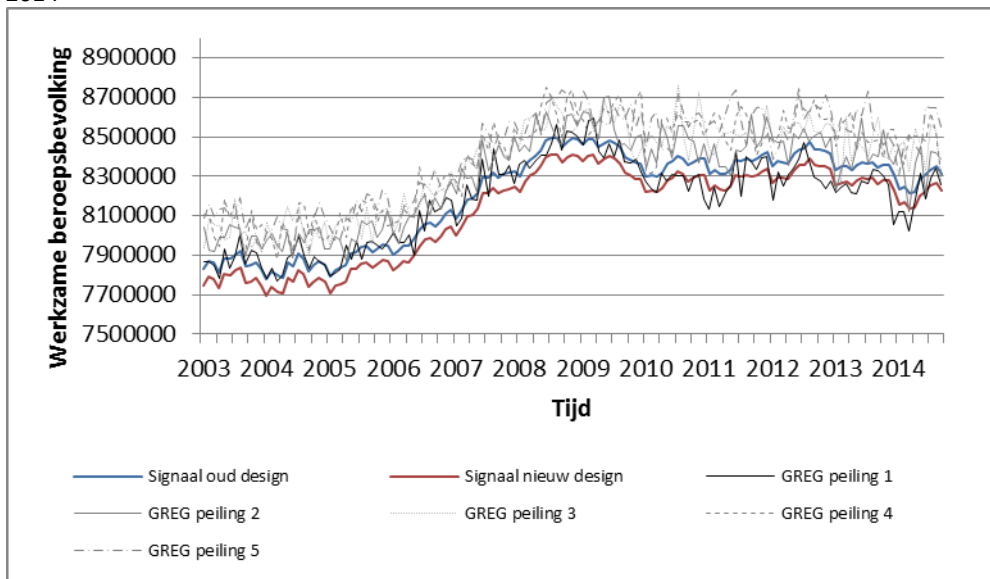
Figuren 5.4 en 5.5 laten de gefilterde trend en het gefilterde signaal voor de werkzame beroepsbevolking voor Nederland zien, op het niveau van het oude design (blauwe lijn) en op het niveau van het nieuwe design (rode lijn). Ook aan deze figuren zijn de reeksen van de GREG-schattingen toegevoegd (zwarte lijnen). Ook voor de werkzame beroepsbevolking geldt dat trend en signaal in de periode voor de eerste HPE-breuk op het niveau van de GREG-schatting voor de eerste peiling liggen. Dat verandert weer in 2010 doordat door de breuken de GREG-schattingen voor de eerste peiling lager komen te liggen ten opzichte van het oude design. De gefilterde trend en het gefilterde signaal op het niveau van het nieuwe design liggen vanaf 2012 op het niveau van de GREG-schattingen voor de eerste peiling.

Figuur 5.6 laat de standaardfout van de trendschattingen voor de werkzame beroepsbevolking zien. Het patroon is vergelijkbaar met dat van de standaardfout voor de trend van de werkloze beroepsbevolking, met een daling tot 2009, een lichte stijging in 2009, een sterke stijging in 2010, daarna weer een daling tot aan het begin van het tweede herontwerp in 2012. Gedurende de periode van het tweede herontwerp stijgt de standaardfout wederom. Vanaf de tweede helft van 2013 daalt de standaardfout weer.

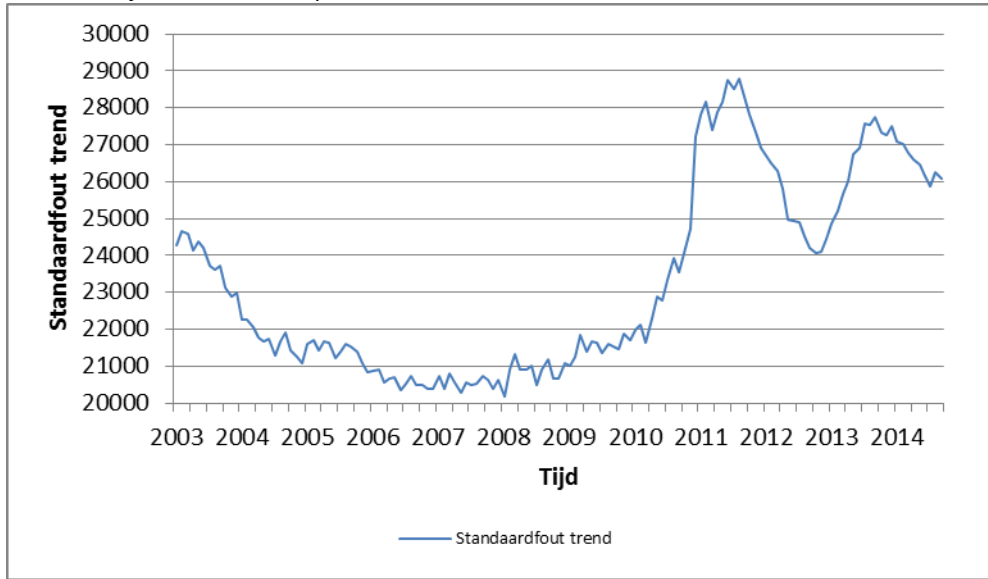
Figuur 5.4: Gefilterde trend op het niveau van het oude design en het nieuwe design en GREG-schattingen, werkzame beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



Figuur 5.5: Gefilterde signaal op het niveau van het oude design en het nieuwe design en GREG-schattingen, werkzame beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



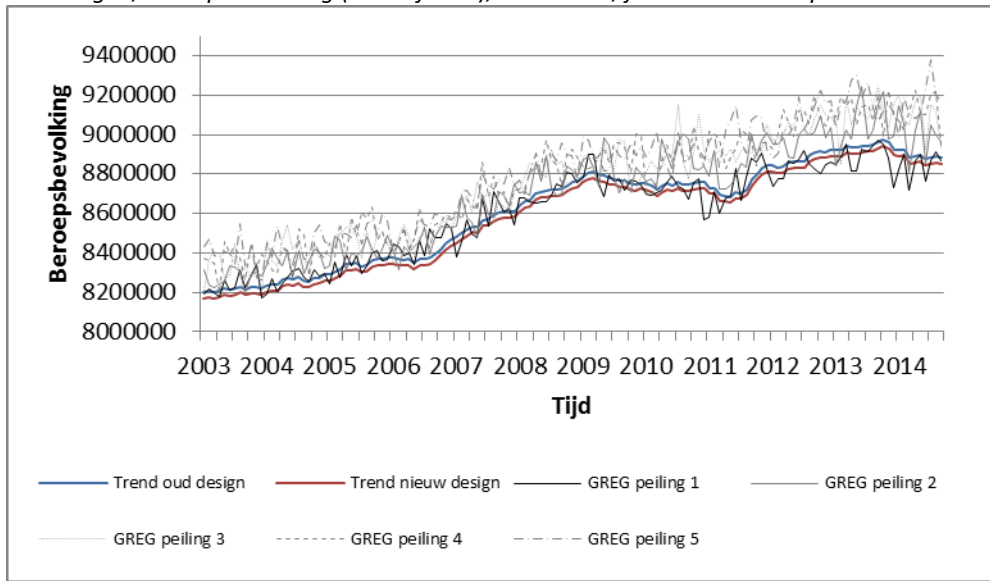
Figuur 5.6: Standaardfout voor de gefilterde trend, werkzame beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



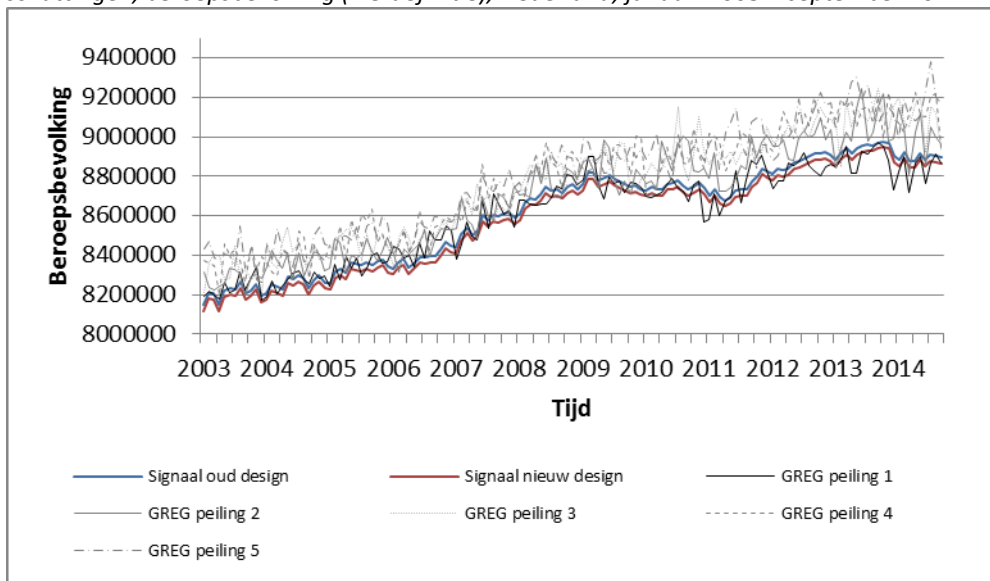
Figuren 5.7 en 5.8 laten de gefilterde trend en het gefilterde signaal voor de beroepsbevolking van Nederland zien op het niveau van het oude design (rode lijn) en het niveau van het nieuwe design (rode lijn). Het verschil tussen het oude en het nieuwe design is voor deze variabele relatief klein, wat samenhangt met de kleine breuk die hier gevonden is (tabel 4.10). Ook aan deze figuren zijn de reeksen van de GREG-schattingen toegevoegd (zwarte lijnen). Ook voor de beroepsbevolking geldt dat trend en signaal in de periode voor de eerste HPE-breuk op het niveau van de GREG-schatting voor de eerste peiling liggen. Dat verandert weer in 2010 door de breuk. Ook hier liggen de modelschattingen voor trend en signaal in de periode 2011/2012 hoger dan de GREG-schattingen voor de eerste peiling. De trendschatting en signaalschatting op het niveau van het nieuwe design liggen op het niveau van de GREG-schattingen voor de eerste peiling. Omdat de breuken voor deze variabele relatief klein zijn, is dit in de figuren nauwelijks zichtbaar.

Figuur 5.9 laat de standaardfout van de modelschatting voor de trend van de beroepsbevolking zien. Het patroon is vergelijkbaar met dat van de standaardfout voor de trend van de werkloze en werkzame beroepsbevolking.

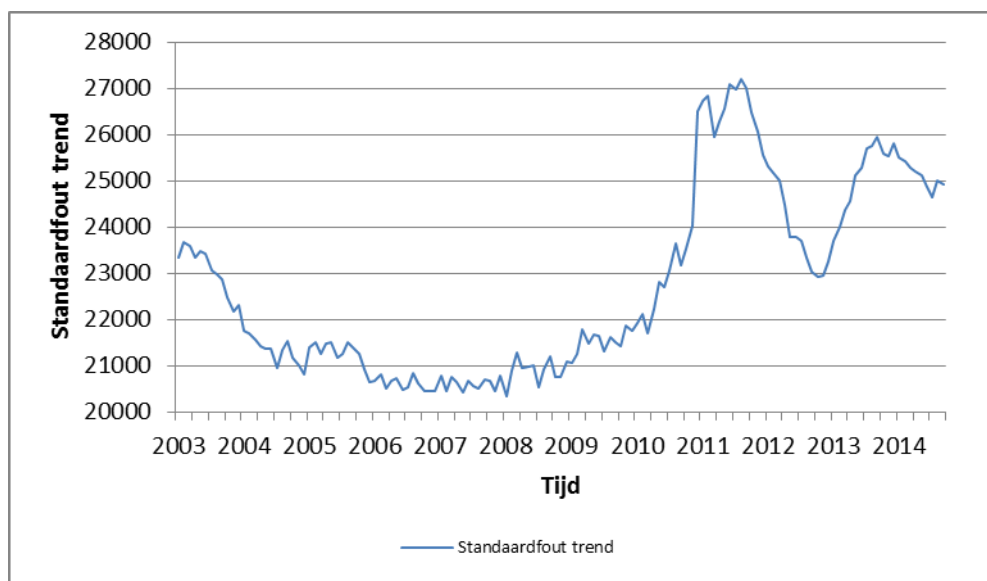
Figuur 5.7: Gefilterde trend op het niveau van het oude design en het nieuwe design en GREG-schattingen, beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



Figuur 5.8: Gefilterde signaal op het niveau van het oude design en het nieuwe design en GREG-schattingen, beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



Figuur 5.9: Standaardfout voor de gefilterde trend, beroepsbevolking (ILO definitie), Nederland, januari 2003 – september 2014



6. Conclusies

Voor het maken van officiële statistieken gebruiken statistische bureaus zoals het CBS vaak schattingsmethodieken afkomstig uit de traditionele steekproeftheorie zoals de gegeneraliseerde regressieschatter. Deze methodiek kent een aantal belangrijke beperkingen. Ten eerste resulteert de gegeneraliseerde regressieschatter bij een kleine steekproefomvang in grote varianties. Bij de EBB doet deze situatie zich onder andere voor bij het maken van maandcijfers over de beroepsbevolking. Ten tweede kan de gegeneraliseerde regressieschatter slechts in beperkte mate omgaan met de vertekening in de uitkomsten van de verschillende herbenaderingen van een roterend panelontwerp en systematische verschillen in de uitkomsten die worden veroorzaakt door veranderingen in de waarneemstrategie.

Voor de EBB is een schattingsmethodiek ontwikkeld waarbij gebruik wordt gemaakt van een multivariaat structureel tijdreeksmodel. Aan de hand van deze methodiek kunnen voldoende nauwkeurige schattingen over de werkloze, werkzame en totale beroepsbevolking op maandbasis worden gemaakt. Dit komt omdat via het tijdreeksmodel op efficiënte wijze gebruik wordt gemaakt van steekproefinformatie uit voorgaande perioden. Daarnaast houdt het model rekening met de vertekening tussen de opeenvolgende peilingen en de autocorrelatie ten gevolge van het panelontwerp. Om bovengenoemde redenen is deze methodiek zeer geschikt om officiële cijfers over de beroepsbevolking op maandbasis te publiceren. Deze methodiek is in 2010 geïntroduceerd.

Ten gevolge van veranderingen in de waarneemstrategie in 2010 en in 2012 ontstaan methodebreuken in de inputreeksen voor het tijdreeksmodel. Om modelmisspecificatie te voorkomen, moet het tijdreeksmodel worden aangepast om rekening te houden met deze methodebreuken. De omvang van deze methodebreuken wordt zo nauwkeurig mogelijk geschat door voor de eerste peiling een zo groot mogelijk paralleltraject uit te voeren bij de overgang in 2010 en in 2012. In het tijdreeksmodel worden de uitkomsten gebenchmarkt op de eerste peiling, omdat de eerste peiling als de meest betrouwbare wordt beschouwd. Om die reden is ervoor gekozen om het beschikbare budget voor het paralleltraject te gebruiken om de breuk in de eerste peiling zo nauwkeurig mogelijk waar te nemen. Bij de overgang in 2012 was het paralleltraject voor de tweede peiling ook van voldoende omvang om een directe schatting voor de methodebreuk te geven. Voor de overige peilingen wordt via interventievariabelen rekening gehouden met methodebreuken. Het tijdreeksmodel geeft modelschattingen voor de omvang van deze breuken. Deze zijn kort na de overgang nog instabiel. Deze instabiliteit werkt niet door in de puntschattingen omdat het tijdreeksmodel uitkomsten benchmarkt naar de eerste peiling. De uitbreidingen van het tijdreeksmodel en de instabiele schattingen voor de breuken direct na de overgang in de vervolgpeilingen zorgen wel voor een toename van de standaardfout van de gefilterde trend en het gefilterde signaal.

Via het structurele tijdreeksmodel kunnen schattingen berekend worden op het niveau van de waarneemstrategie van voor 2010. Hiermee zijn de maandcijfers over de beroepsbevolking door de tijd heen vergelijkbaar. Het is ook mogelijk om schattingen te berekenen die op het niveau van de nieuwe waarneemstrategie zijn. Vanaf januari 2015 worden cijfers over de beroepsbevolking gepubliceerd op het niveau van het nieuwe design.

7. Referenties

- Bailar, B.A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, 70, pp. 23-30.
- Binder, D.A., en J.P. Dick (1990). A method for the analysis of seasonal ARIMA models. *Survey Methodology*, 16, pp. 239-253.
- Brakel, J.A. van den, en K. van Berkel (2012). Transitietraject Herontwerp Persoonsenquêtes bij de EBB. BPA-nr. PPM-2012-02-24-JBRL-CBRL.
- Brakel, J.A. van den, en S. Krieg (2009). Estimation of the Monthly Unemployment Rate through Structural Time Series Modelling in Rotating Panel Design. *Survey Methodology*, 16, 2, 177-190.
- Brakel, J.A. van den, en S. Krieg (2010). Schatten van maandcijfers over de beroepsbevolking. CBS publicatie, 6 mei 2010, Heerlen. <http://www.cbs.nl/NR/rdonlyres/008529E3-597E-497C-BAAC-27574253D803/0/ebbmaandcijfersschatten.pdf>
- Brakel, J.A. van den, en S. Krieg (2012). Dealing with small sample sizes, rotation group bias and discontinuities in a rotating panel design. BPA nr. PPM-2012-11-13-JBRL. Centraal Bureau voor de Statistiek, Heerlen.
- Brakel, J.A. van den en J. Roels (2010). Intervention analysis with state-space models to estimate discontinuities due to a survey redesign. *Annals of Applied Statistics*, 4, 1105-1138.
- Dataverzameling – monitoren en analyse (2010). Veldwerkverslag PDV December 2010. Interne CBS nota, Centraal Bureau voor de Statistiek, Heerlen.
- Dataverzameling – monitoren en analyse (2011). Veldwerkverslag DVU Juli/Augustus 2011. Interne CBS nota, Centraal Bureau voor de Statistiek, Heerlen.
- Dataverzameling – monitoren en analyse (2013). Veldwerkverslag DVU December 2012. Interne CBS nota, Centraal Bureau voor de Statistiek, Heerlen.
- Doornik, J.A. (2007). *An Object-oriented Matrix Programming Language Ox 5*. London: Timberlake Consultants Press.
- Durbin, J., en Koopman, S.J. (2001). *Time series analysis by state space methods*. Oxford: Oxford University Press.
- Harvey, A.C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge: Cambridge University Press.
- Koopman, S.J., N. Shephard en J.A. Doornik, (1999). Statistical Algorithms for Models in State Space using SsfPack 2.2. *Econometrics Journal*, 2, 113-166.
- Koopman, S.J., N. Shephard en J. A. Doornik (2008). *Statistical Algorithms for Models in State Space using SsfPack 3.0*. London: Timberlake Consultants Press.
- Leden Taskforce EBB (2010). Rapportage 2 Responsverbetering EBB. Interne nota CBS, BPA nr. PDV-2010-N-H3, Centraal Bureau voor de Statistiek, Heerlen.

- Pfeffermann, D. (1991). Estimation and seasonal adjustment of population means using data from repeated surveys. *Journal of Business & Economic Statistics*, 9, pp. 163-175.
- Pfeffermann, D., M. Feder, en D. Signorelli (1998). Estimation of autocorrelations of survey errors with application to trend estimation in small areas. *Journal of Business & Economic Statistics*, 16, pp. 339-348.
- Särndal, C-E., B. Swensson en J. Wretman (1992). *Model Assisted Survey Sampling*. New York: Springer Verlag.