



Centraal Bureau voor de Statistiek

MSP/TNM
MNR/MOO/TMO

ENQUÊTERESULTATEN WERKLOZE BEROEPSBEVOLKING: UITBIJTERDETECTIE

R. Bikker, R. Hijman, P. Kee en G. van Leeuwen

Samenvatting: Met het oog op mogelijke uitbijters in de enquêteresultaten is het gewenst de werkloosheidscijfers volgens de Enquête beroepsbevolking vóór publicatie te confronteren met gerelateerde werkloosheidsindicatoren. Uitbijters in de uitkomsten van de enquête kunnen dan tijdig worden geïdentificeerd.

Trefwoorden: arbeidsmarkt, tijdreeksanalyse.

1. Inleiding

Maandelijks geeft het CBS via StatLine en een persbericht informatie over de gemiddelde omvang van de werkloze beroepsbevolking (WLB) in de afgelopen drie maanden. In de berichtgeving wordt het geaggregeerde cijfer uitgesplitst naar geslacht en drie leeftijdsgroepen. De WLB volgens de definitie van het CBS telt alle personen zonder werk (of met werk van minder dan twaalf uur per week), die minstens twaalf uur per week willen werken, daarvoor direct beschikbaar zijn en bovendien actief op zoek zijn naar werk. Het gaat hierbij om personen in de leeftijd vanaf 15 tot en met 64 jaar. De omvang van de WLB kan alleen worden bepaald aan de hand van een persoonsenquête. Het CBS gebruikt hiervoor de Enquête beroepsbevolking (EBB).

De 95%-betrouwbaarheidsmarges van de steekproefuitkomsten van de EBB kunnen vrij goed worden bepaald. Dit geldt ook voor de marges van mutaties. Steekproefuitschieters zijn echter bij toeval mogelijk. Zij leiden tot onjuiste

Projectnummer: 103439-Euro-02
BPA-nummer: 2005-01-MOO
Datum: 8 februari 2005

conclusies over het verloop van de werkloosheid. Het is dus van belang deze uitschieters te signaleren en te corrigeren. Hiervoor vergelijken we vóór publicatie de verandering in de werkloosheidscijfers uit de EBB met de mutaties in verwante werkloosheidsindicatoren zoals het aantal bij het CWI ingeschreven niet-werkende werkzoekenden (NWW) en het aantal WW-uitkeringen.

Het CBS hanteert vanaf augustus 2004 een tijdreeksmodel om de plausibiliteit van de te publiceren gegevens vooraf te beoordelen. Met het model, dat een empirische relatie legt tussen de EBB- en CWI-cijfers, worden de meest recente cijfers voor de WLB voorspeld. Door confrontatie van voorspelling en realisatie worden uitbijters in de EBB-uitkomsten geïdentificeerd. Een uitbijter wordt gedefinieerd als een realisatie die buiten het 95%-betrouwbaarheidsinterval van de voorspelling ligt. Dit artikel beschrijft de methodologie achter deze uitbijterdetectie.

2. De niet-werkende werkzoekenden

Het CBS ontvangt maandelijks van het CWI een bestand met daarin alle personen die de afgelopen drie maanden als werkzoekende stonden ingeschreven bij het CWI. Hieruit kan het gemiddelde NWW-cijfer over alle dagen van de afgelopen maand worden bepaald, wat een maandcijfer oplevert. Dit gemiddelde verschilt uiteraard van de maandelijks door het CWI in de Nieuwsflits Arbeidsmarkt gerapporteerde stand ultimo de maand. Personen die deel uitmaken van de NWW staan ingeschreven bij het CWI én werken volgens deze registratie niet of minder dan twaalf uur per week. Feitelijk is het echter mogelijk dat zij dit wel doen, bijvoorbeeld als de inschrijving wordt gehandhaafd omdat slechts sprake is van een tijdelijke baan. Of men beschikbaar is voor een baan van minimaal twaalf uur per week speelt geen rol in de definitie van de NWW. Het is aannemelijk dat de cijfers over de NWW achterlopen op de actuele situatie. Met name werkaanvaarding wordt niet altijd (tijdig) doorgegeven aan het CWI. Een verschil in gemiddelde vertraging tussen in- en uitschrijvingen zou betekenen dat het NWW-cijfer bij dalende werkloosheid (meer uit- dan inschrijvingen) gemiddeld meer achterloopt dan bij stijgende werkloosheid (meer in- dan uitschrijvingen).

Gegevens over de WLB en de NWW hebben slechts gedeeltelijk betrekking op dezelfde personen (zie de Vries et al., 2003). In 2003 behoorde gemiddeld bijna 50% van de WLB tot de NWW. Deze overlap bedroeg slechts gemiddeld 32% van de NWW. Personen die wel tot de WLB maar veelal niet tot de NWW behoren, betreffen onder meer degene die uitkijken naar tijdelijk werk, zoals jongeren die een vakantiebaantje zoeken. Verder vallen hieronder vaak ook de personen die geen recht hebben op een uitkering, zoals jonge schoolverlaters en herintredende vrouwen. Jongeren betreden de arbeidsmarkt veelal in de zomerperiode, waardoor vooral in dit tijdvak een duidelijk verschil optreedt in het patroon van WLB en NWW. Personen die behoren tot de NWW maar niet tot de WLB, betreffen onder meer degene die niet beschikbaar zijn voor werk van twaalf uur of meer per week. Verder vallen hieronder ook de personen die niet actief zoeken, bijvoorbeeld omdat

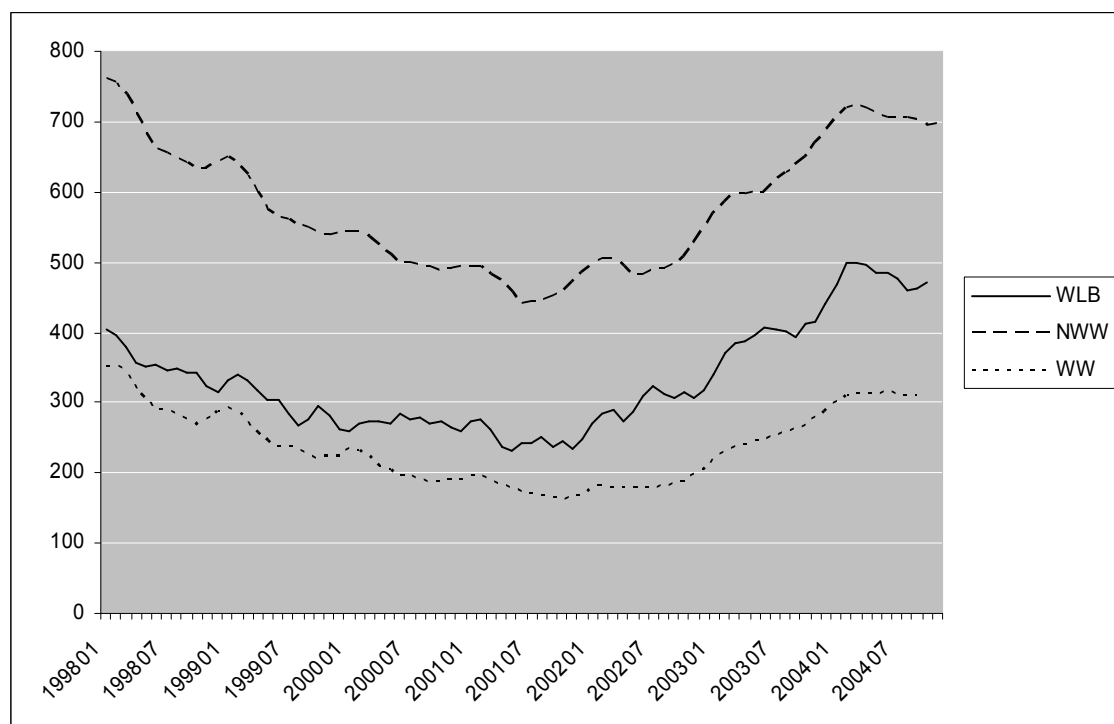
de uitkeringsregeling voor hen een ontheffing kent van de sollicitatieplicht of omdat zij zijn ontmoedigd.

3. De relatie tussen de reeksen

Het CBS meldde op 16 december dat de WLB gemiddeld 470200 personen telde in de periode september–november 2004. Het door het CBS berekende NWW-cijfer voor november 2004 bedraagt 696300. Figuur 1 laat zien, dat het cijfer voor de NWW zich op een structureel hoger niveau bevindt dan dat voor de WLB. Dezelfde figuur duidt op een vrijwel constant niveauverschil in de tijd en op een redelijk stabiele relatie tussen het verloop van deze twee cijferreeksen. Verder zien we dat de stijging van de WLB vanaf het voorjaar van 2001 en de daling medio 2004 vrijwel tegelijkertijd tot uiting komen in de cijfers uit het NWW-bestand.

In figuur 1 is ook de reeks van het aantal WW-uitkeringen weergegeven. Deze reeks beweegt zich op een structureel lager niveau (310400 ultimo september 2004) dan de NWW- en WLB-cijfers. De WW-reeks voegt vrijwel geen informatie toe aan de CWI-cijfers (de correlatiecoëfficiënt is 0,98): ontvangers van een WW-uitkering staan verplicht ingeschreven bij het CWI. Een ander nadeel is dat de gegevens over het aantal WW-uitkeringen met een vertraging van enkele maanden beschikbaar komen. Voor ons doel van confrontatie met de WLB-cijfers zou de uitkeringsinformatie dan moeten worden geëxtrapoleerd.

Figuur 2 geeft de relatie tussen de niveaus van de WLB en NWW in een spreidings-

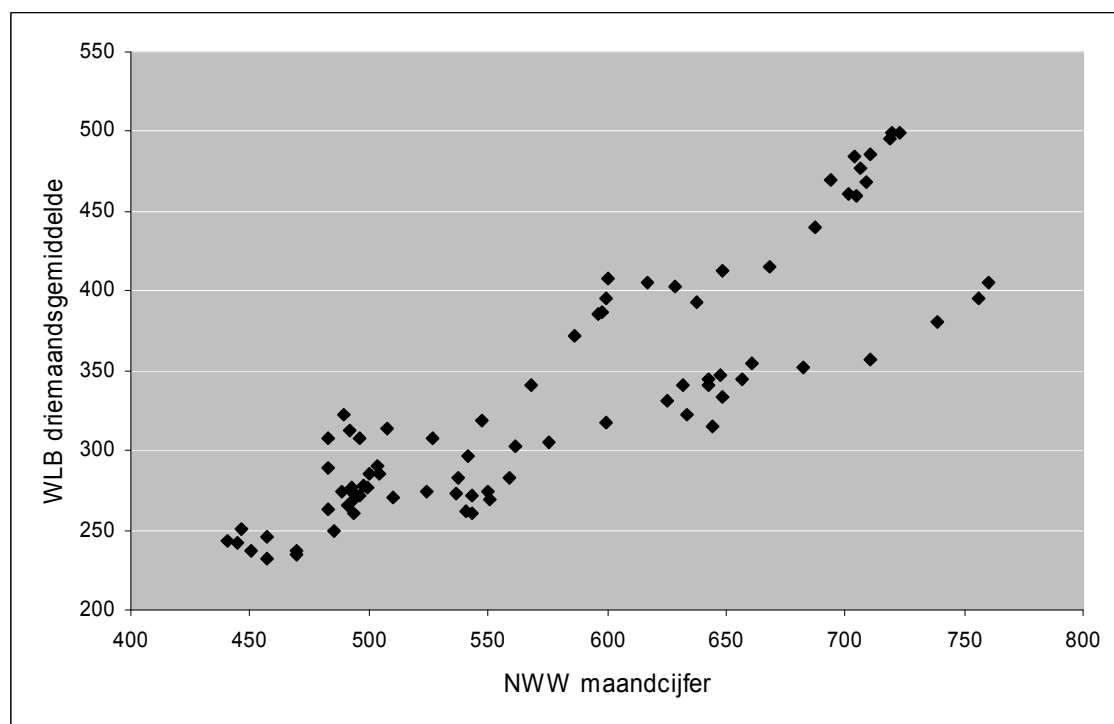


Figuur 1. Tijdreeksen van de WLB, NWW en WW.

diagram weer. De figuur toont in feite twee lineaire relaties: één tak voor de periode januari 1998–december 1999 en één tak met een steilere helling voor de periode januari 2000–november 2004. De precieze oorzaak van dit beeld is niet duidelijk. Externe factoren, zoals de oprichting van het CWI of de omslag van de conjunctuur, kunnen verantwoordelijk zijn. Ook zou het te maken kunnen hebben met een andere waarnemingsmethodiek bij de EBB vanaf 2000. Nader onderzoek is nodig om de oorzaak te achterhalen.

4. Voorspelling

Voor het voorspellen van de te publiceren EBB-uitkomsten hanteren we een tijdreeksmodel dat de empirische relatie tussen de WLB-tijdreeks en de tijdreeks van de NWW beschrijft. Hierbij maken we gebruik van data vanaf 1998. We voorspellen de WLB één maand vooruit. Voor de voorspelling voor maand t gebruiken we de realisaties van de WLB-reeks tot en met $t-1$ en van de NWW-reeks tot en met t . De voorspelling voor het macrocijfer bepalen we door aggregatie van de voorspellingen voor de zes publicatiegroepen: mannen en vrouwen in de leeftijdscategorieën 15–24, 25–44 en 45–64. De onderling consistente ‘out-of-sample’ voorspellingen voor het geaggregeerde cijfer en de gedesaggregeerde cijfers vergelijken we vervolgens met de gerealiseerde waarden.



Figuur 2. Spreidingsdiagram van de WLB- en NWW-niveaus.

5. Error-correction model

We hebben in eerste instantie gebruik gemaakt van een ‘Error-correction’ model (ECM; zie bijvoorbeeld Greene, 2000, blz. 733-735). Dit model is gebaseerd op twee opeenvolgende en onafhankelijke OLS-regressies: een lange termijn evenwichtsrelatie in niveaus tussen WLB en NWW en een korte termijn relatie in eerste verschillen met fluctuaties rondom het evenwicht. De regressieresultaten voor de evenwichtsrelatie bevestigen een significant steilere helling vanaf 2000 (zie figuur 2). Vanaf het omslagpunt van een dalende naar een stijgende werkloosheid (april 2001) is het intercept van deze relatie bovendien significant hoger. Dit is een gevolg van de eerder genoemde asymmetrische vertragingen in de in- en uitschrijvingen bij het CWI.

De vergelijking van ECM-voorspelling en realisatie voor de periode januari 2002–november 2004 resulteert in een groot aantal uitbijters bij de onderscheiden subpopulaties. De typering als uitbijter van bijna al deze realisaties is echter op grond van het historische verloop van de deelreeksen niet plausibel. Nadere analyse leert dat de voorspellingen met het ECM zeer gevoelig zijn voor de lengte van de schattingsperiode. Met de deelreeksen vanaf 1998 blijken de voorspellingen tot op heden instabiel. We concluderen dat het ECM (vooralsnog) ongeschikt is voor ons doel van uitbijterdetectie.

6. State-space model

De resultaten van de simulaties met het ECM zijn aanleiding om een alternatieve aanpak voor het voorspellen van de EBB-cijfers te onderzoeken. Deze aanpak maakt gebruik van een ‘State-space’ model (SSM; zie bijvoorbeeld Durbin en Koopman, 2001). De SSM-aanpak houdt in tegenstelling tot het parametrische ECM rekening met het stochastische karakter van de verschillen tussen de WLB en de NWW. Hierdoor zijn veranderingen in de eigenschappen van de reeksen beter te volgen. De meest recente informatie wordt in een recursief proces opgebouwd door de schatting van het simultane SSM met de methode van Maximale Aannemelijkheid.

Uitgangspunt bij de opzet van het model is het idee dat de werkloosheid is opgebouwd uit een aantal niet direct waarneembare componenten: een trend, een seizoenfactor en een onregelmatig deel (statistische ruis). We extraheren deze componenten uit de CWI-cijfers, omdat maandcijfers hiervoor geschikter zijn dan de gepubliceerde 3-maands gemiddelden volgens de EBB-reeks. Het niveau van de trend wordt gelijk gesteld aan het niveau van de voorafgaande maand vermeerderd met een stochastische helling (de verandering van de trend). Ook het seizoenpatroon veronderstellen we stochastisch. Vervolgens breiden we het model uit met een vergelijking die de EBB-cijfers koppelt aan de componenten van de CWI-reeks. Hierbij houden we rekening met stochastische niveauverschillen tussen de CWI- en EBB-reeksen van zowel seizoenafhankelijke als niet-seizoengebonden aard.

Bij aanwending van het SSM wordt een gering aantal uitbijters aangewezen voor de subgroepen: over de periode januari 2002–november 2004 slechts vijf. De overschrijding van de betrouwbaarheidsmarge van de voorspelling is bij deze realisaties gering en geloofwaardig. Voor 2004 vinden we twee uitbijters, die zich beide voordoen bij het gemiddelde over de eerste drie maanden. Tabel 1 laat zien dat voor werkloze vrouwen in de leeftijdscategorieën 15–24 en 45–64 de realisatie net buiten het 95%-betrouwbaarheidsinterval van de voorspelling ligt. Bij de gebruikelijke revisie over het afgelopen jaar zal worden bezien of herziening van deze cijfers in de rede ligt.

Voor het macrocijfer resulteert het beeld zoals weergegeven in figuur 3. De gestippelde lijnen in de figuur representeren het 95%-betrouwbaarheidsinterval van de voorspellingen. Voor de maanden januari 2003–november 2004 wordt één uitbijter gedetecteerd (december 2003).

7. Uitbijters

Een geconstateerde uitbijter kan te wijten zijn aan fouten bij de verwerking van de data. Onze methode van uitbijterdetectie vergemakkelijkt het gericht zoeken naar fouten binnen de diverse subpopulaties.

Een uitbijter kan ook ontstaan doordat recente ontwikkelingen die de omvang van de WLB beïnvloeden (nog) niet tot uiting komen in de CWI-cijfers, of vice versa. Indien geen fouten worden gevonden, moet nadere analyse dan ook uitwijzen of er plausibele verklaringen zijn voor sterk van de modelschattingen afwijkende EBB-uitkomsten. Om geforceerde verklaringen achteraf te vermijden, moet vóóraf een inschatting worden gemaakt van de te verwachten ontwikkeling in de WLB-reeksen.

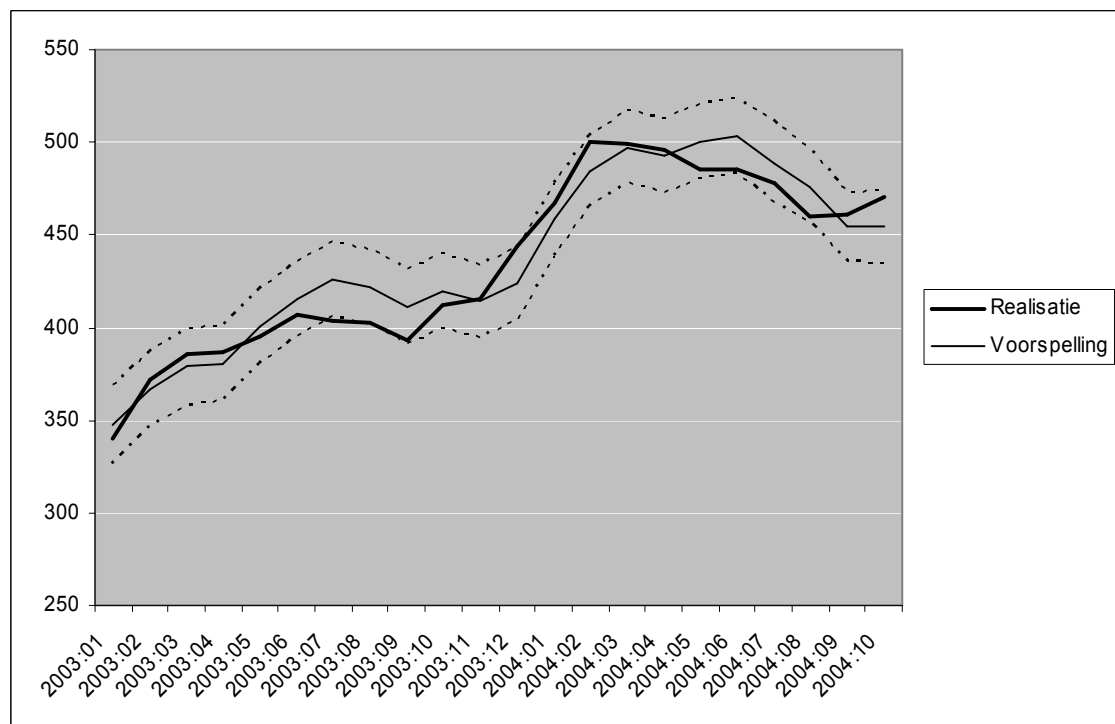
Tabel 1. Toepassing van het SSM op de deelreeksen van de WLB.

Geslacht	Leeftijd	WLB	Voor- spelling	Onder- grens	Boven- grens	Gestand. afw. ^a
Januari – maart 2004						
M+V	15–64	500,2	484,4 ^b	465,5 ^c	503,4 ^c	1,7
M	15–24	64,4	60,6	54,4	66,9	1,2
M	25–44	127,3	125,3	114,0	136,5	0,4
M	45–64	70,1	69,5	64,1	74,9	0,2
V	15–24	57,9	48,4	40,0	56,9	2,2
V	25–44	123,3	128,8	120,5	137,0	-1,3
V	45–64	57,1	51,8	46,8	56,8	2,1

^a) Gestandaardiseerde afwijking: het verschil tussen voorspelling en realisatie gedeeld door de standaardfout van de voorspelling.

^b) Som van voorspellingen voor strata WLB.

^c) Marge berekend als de wortel uit de som van de gekwadrateerde marges van voorspellingen voor strata WLB.



Figuur 3. Vergelijking SSM-voorspelling van de WLB met realisatie.

Als er geen sprake is van fouten en als ook een plausibele verklaring ontbreekt, wordt de betreffende uitbijter als steekproefuitschieter gezien en moet aanpassing plaatsvinden. Deze aanpassing dient plaats te vinden tot op de rand van het 95%-betrouwbaarheidsinterval van de voorspelde waarde, waarbij als randvoorwaarde geldt dat de gecorrigeerde waarde binnen de 95%-betrouwbaarheidsmarge van de EBB-uitkomst blijft.

8. Conclusie

Door mogelijke uitbijters in de steekproefuitkomsten bestaat er behoefte aan een objectieve en transparante methode om door het CBS gepubliceerde werkloosheidscijfers vooraf te beoordelen. Wij geven een toetsing van deze cijfers door ze te confronteren met voorspelde waarden die worden afgeleid uit de empirische relatie tussen de werkloosheidsindicatoren van het CBS en het CWI. Door hierbij gebruik te maken van de ‘State-space’ methodologie kunnen uitbijters in de werkloosheidsuitkomsten van het CBS op een gedegen wijze worden geïdentificeerd. Bovendien kan zo een eventuele aanpassing van het enquêteresultaat kwantitatief worden onderbouwd.

Literatuur

Durbin, J. en S.J. Koopman (2001) Time Series Analysis by State Space Methods. Oxford: Oxford University Press.

Greene, W. (2000) Econometric Analysis, 4th ed. New York: Prentice Hall International.

Vries, S. de, S. Lucassen, J. van der Valk, en A. Bouman (2003) Staan werklozen ingeschreven als niet-werkende werkzoekenden? Sociaal-economische maandstatistiek, 20(10), 18-22.