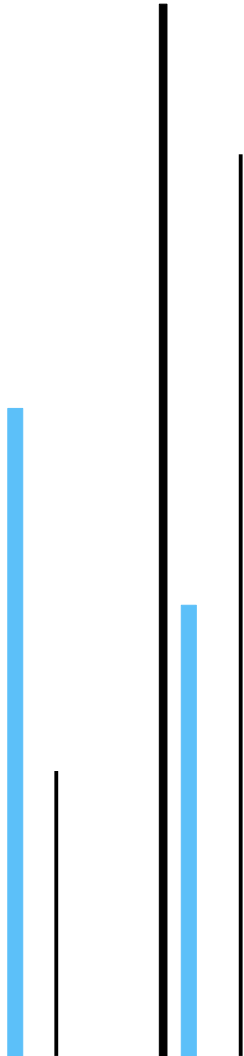


Dataprocessing WoON 2015

Onderzoeksdocumentatie



Dataprocessing WoON 2015

Onderzoeksdocumentatie

Uitgevoerd in opdracht van CBS

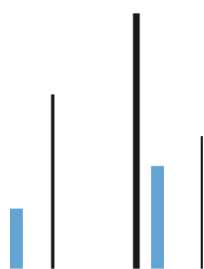
Kenneth Gopal

25 maart 2016 | r2016-0024KG | 14183-DPC

ABF Research | Verwersdijk 8 | 2611 NH | Delft | 015 - 27 99 300

Copyright ABF Research 2015

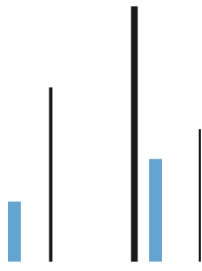
De informatie in dit rapport is met de grootste zorg samengesteld. ABF Research aanvaardt geen aansprakelijkheid voor eventuele fouten, onnauwkeurigheden of onvolledigheden. Het gebruik van (onderdelen van) dit rapport is toegestaan mits de bron duidelijk wordt vermeld.



Inhoudsopgave

1	Inleiding.....	7
2	Werkzaamheden dataprocessing.....	9
2.1	Routingcontrole	9
2.2	Typeren open vragen	10
2.3	Consistentiecontrole	11
2.4	Imputatie.....	12
2.5	Bepalen gewenste woonmilieus	16
2.6	Bijkomende woonlasten.....	16
3	Eindproducten	19
	Bijlage 1 Overzicht typering.....	21
	Bijlage 2 Consistentiecontroles.....	31
	Bijlage 3 Bijschatting Hypothekenblok.....	33
	Bijlage 4 Bijschatting huurpunten.....	37

1



Inleiding

Het Woon Onderzoek Nederland (WoON) is het grootste landelijk onderzoek in Nederland op het gebied van wonen en woonomgeving. Sinds 2006 geldt het WoON als de vervanger van het Woningbehoefte Onderzoek (WBO) en de Kwalitatieve Woning Registratie (KWR). De initiatiefnemers van dit onderzoek zijn het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK) en Centraal Bureau voor de Statistiek (CBS).

Het onderzoek is modulair van opzet. De basissteekproef telt netto 41.300 respondenten. Evenals in voorgaande edities hebben gemeenten of regionale samenwerkingsverbanden meegedaan aan de oversampling, waarvoor nog eens 21.300 respondenten benaderd zijn. Hiermee komt het totaal aantal respondenten dat aan de basismodule van het WoON 2015 heeft deelgenomen op ruim 62.600. Deze wordt ook wel de module Woningmarkt genoemd.

De gegevensverzameling heeft tijdens de veldwerkfase plaatsgevonden via telefonische, persoonlijke en internetenquêtes. Het gehele veldwerk was in handen van CBS.

Na de veldwerkfase volgt de fase van de dataprocessing. De term dataprocessing wordt gebruikt om de werkzaamheden samen te vatten die nodig zijn om aan de antwoorden die respondenten geven op vragen in de WoON-interviews plausibele statistische informatie te ontleen. Met andere woorden, er wordt in deze fase vanuit een ruw veldwerkbestand toegewerkt naar een onderzoeksbestand dat zo veel mogelijk ontdaan is van bestandsvervuiling.

De dataprocessing voor het WoON omvat de volgende onderdelen: typeren open vragen, uitvoeren consistentiecontroles, imputeren (bijschatten) item non-respons, bepalen afgeleide variabelen, draaien van inkomensprogramma, bepalen bijkomende woonlasten, wegen en ophogen, plausibiliteitcontroles en non-respons analyse.

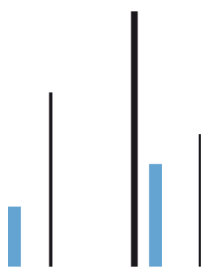
In 2002 en 2006 heeft ABF Research de volledige dataprocessing uitgevoerd van respectievelijk het WBO 2002 en het WoON 2006. Voor het WoON 2009 en 2012 lag de regie over de dataprocessing echter in handen van CBS, en is ABF Research op een aantal onderdelen ingeschakeld.

Dat is voor de nieuwste editie eveneens het geval. ABF Research heeft ten behoeve van de dataprocessing van het WoON 2015 de volgende werkzaamheden uitgevoerd:

- De controle van de routing in de vragenlijst;
- Typering van open vragen;
- De imputatie van ontbrekende waarden;
- Consistentiecontrole;
- De afleiding van het gewenste woonmilieu;
- Afleiding van de bijkomende woonlasten.

In dit rapport wordt verslag gedaan van de werkzaamheden die ABF Research uitgevoerd heeft in het kader van de dataprocessing van de module Woningmarkt van WoON 2015.

2



Werkzaamheden dataprocessing

De verschillende onderdelen van de dataprocessing van de module Woningmarkt waar ABF Research aan gewerkt heeft worden in dit hoofdstuk stapsgewijs behandeld.

2.1 Routingcontrole

Niet alle vragen hoeven door iedere respondent te worden beantwoord. Afhankelijk van bijvoorbeeld de leeftijd, de positie in het huishouden of de verhuishwens worden daarop afgestemde vragenblokken gesteld. Met ruim vierhonderd vragen verspreid over bijna veertig blokken is de routing in de vragenlijst van het WoON zeer complex. In het verleden is het verschillende malen voorgekomen dat na de afsluiting van de veldwerkfase nog fouten in de routing werden ontdekt. Bij gevolg moesten respondenten worden nagebeld of vragenblokken in het bestand extra worden geïmputeerd. Dit had direct nadelige gevolgen voor de opleveringstermijn, kwaliteit en de kosten.

Sinds het WBO 2002 is door ABF Research vóóraf een routingcontrole uitgevoerd om de kans op fouten in de routing van de vragenlijst zo klein mogelijk te houden. Ook voorafgaand aan het WoON 2015 is een dergelijke controle uitgevoerd. Hiertoe is eerst de vragenlijst door de betrokken partijen, te weten CBS en ABF Research, onafhankelijk van elkaar geautomatiseerd. Vervolgens zijn de vragenlijsten met elkaar vergeleken waarna de gevonden verschillen zijn teruggekoppeld en zo nodig gecorrigeerd, opdat beide partijen dezelfde routing in de vragenlijst hanteren.

Anders dan in voorgaande edities is ook de routing gecontroleerd voor antwoordcategorieën bij een aantal vragen. Afhankelijk van het antwoord op een voorgaande vraag, kan het voorkomen dat omwille van consistentiebehoud niet alle antwoorden van een vervolgvraag als keuze voorgelegd worden.

De routingcontrole beslaat drie stappen:

1. maken geautomatiseerde vragenlijst;
2. controle via proefbestand;
3. correctie en terugkoppeling.

2.1.1 Maken computergestuurde vragenlijsten

De routing van de totale vragenlijst is door ABF Research nagebouwd in de syntax van de Surfox¹-inputfiles. De vragenlijst is ten opzichte van het WoON 2009 op veel onderdelen ongewijzigd gebleven. Bij het programmeren van de vragenlijst is daarom de geprogrammeerde vragenlijst van het WoON 2009 als uitgangspunt genomen. Voor alle vragen is goed gecontroleerd of de routing bij de betreffende vraag inderdaad ongewijzigd gebleven is.

¹ Surfox is een statistisch software pakket dat door ABF Research zelf is ontwikkeld.

Bij de bouw van de geautomatiseerde vragenlijst is rekening gehouden met de volgende punten:

- Routing in de vragenlijst;
- Tekstvelden (deze worden overgeslagen);
- Controle op minimum- en maximumwaarden;
- De categorieën 'weet niet' en 'weigert'.

2.1.2 Controle via proefbestand

Om de vragenlijsten te kunnen testen is gewerkt met proefbestanden, ook wel dummybestanden genoemd. Deze bestaan niet uit daadwerkelijk uitgevoerde enquêtes, maar uit duizenden computer gegenereerde enquêtes. Ten behoeve van de toetsing van de computergestuurde vragenlijsten heeft ABF Research op diverse momenten een dummybestand van 50.000 records gegenereerd. Op dit bestand is vervolgens de routing volgens ABF Research toegepast, waarna dit gerouteerde bestand naar CBS is verzonden.

2.1.3 Correctie en terugkoppeling

CBS heeft haar routing vergeleken met de routing die door ABF Research is toegepast op het dummybestand. In nauw overleg hebben beide partijen de geconstateerde verschillen besproken.

Nadat de oplossingen voor de verschillen van de routingtest gevonden waren, is op basis van de verbeterde vragenlijst door ABF Research een nieuw proefbestand verstuurd, waarna CBS opnieuw controles uitgevoerd heeft. Dit proces is net zo lang herhaald totdat er geen verschillen meer optraden.

2.2 Typeren open vragen

De module Woningmarkt kent hoofdzakelijk keuzevragen en een aantal intervalvragen, bijvoorbeeld voor jaartallen en bedragen. Voor enkele vragen is een open antwoordmogelijkheid gegeven, waarbij de enquêteur (of de respondent zelf in geval van CAWI mode) het antwoord als tekst invoert. Meestal zijn deze open vragen voorafgegaan door een keuzevraag waarbij de categorie 'anders' een van de antwoordmogelijkheden was en ook gekozen is. Deze open antwoorden moeten worden gecodeerd of getypeerd naar eenduidige codes die aansluiten op de reeds bestaande antwoordcategorieën van de betreffende vraag.

Voor WoON 2015 zijn 26 variabelen getypeerd. De typering is een verlengstuk van de veldwerkfase en zijn vrij vroeg in het traject opgenomen, omdat zij mogelijk in het vervolg (routing, consistentie en imputatie) een rol kunnen spelen. Het typeren van de open antwoorden is eenmalig uitgevoerd, waarbij voor elke tekstregel in een apart bestand is vastgelegd wat de bijbehorende code is als deze te herleiden was naar een bestaande antwoordcode van de voorloopvraag. Indien een open antwoord niet eenduidig getypeerd kon worden is de code 'anders' in de voorloopvraag gehandhaafd, evenals de bijbehorende tekst van het open antwoord. In sommige gevallen verwijst de tekst (in)direct naar 'weet niet' of 'weigert'. Voor twee variabelen zijn deze antwoorden als 'missings' beschouwd en is de voorloopvraag vervolgens bijgeschat (waarover meer in paragraaf 2.4).

Tijdens het typen is per vraag eveneens gekeken of de te typen omschrijvingen aanleiding geven om een nieuwe antwoordcategorie aan de vraag toe te voegen. Dit kan het geval zijn wanneer een substantieel deel van de respondenten die een omschrijving geeft zich niet kan vinden in de bestaande set van antwoordcategorieën. Alle bevindingen zijn overlegd met CBS en BZK; zie hiervoor ook Bijlage 1.

2.2.1 Bijzondere gevallen

Voor de meeste vragen is de typering van de omschrijvingen vanzelfsprekend verlopen. In een aantal gevallen was speciale aandacht vereist:

- Voor de genoemde omschrijvingen met betrekking tot de verhuurder is nagegaan of dit kon worden herleid tot een bestaande corporatie of anders tot een van categorieën die als type verhuurder genoemd zijn. Voor de typering naar bestaande categorieën is gebruik gemaakt van een lijst met bestaande verhuurders.
- Respondenten verwarren soms de naam van het hypotheekproduct met de hypotheekvorm. Het kan dan ook geregeld voorkomen dat het product in de omschrijving genoemd wordt. Voor de meest omschreven hypotheekvormen die niet eenvoudig tot een van de bestaande antwoordcategorieën kon worden omgezet is via internet informatie verzameld om het type hypotheek te bepalen. Indien het type alsnog niet eenduidig kon worden achterhaald is de code 'anders' van toepassing gebleven.
- In het geval de typering betrekking hebben op landen, woonplaatsen of buurten zijn deze tijdens het veldwerk automatisch gecodeerd tot een bestaande wijk- en buurtcode volgens de indeling van CBS. Voor de antwoorden die niet via de codeermodule afgevangen werden heeft ABF Research bekeken of deze alsnog niet gecodeerd konden worden tot een bestaande land-, woonplaats- of buurtcode.
- De vragen met betrekking tot de tussenvorm en ontwikkelingen die de respondent verontrusten zijn nieuw en geheel open van vorm. Voor deze twee vragen heeft ABF Research eerst de belangrijkste categorieën vastgesteld en deze voorgelegd aan CBS en BZK. Vervolgens zijn de gegeven omschrijvingen hiernaar getypeerd.

2.3 Consistentiecontrole

Tijdens het veldwerk wordt door CBS al rekening gehouden met de toegestane maximum- en minimumwaarden. Door een aantal waarden van variabelen te combineren kunnen andere inconsistenties aan het licht komen. Het kan hierbij onder meer gaan om de combinaties van verschillende jaartallen (de respondent heeft zich bijvoorbeeld gevestigd op een adres vóór het bouwjaar van de woning). Indien een dergelijke inconsistentie gevonden wordt moet een zeer goede afweging worden gemaakt of de gegeven antwoorden gecorrigeerd moeten worden, zodat de samenhang na de correctie wél consistent is.

CBS heeft in samenspraak met BZK deze afweging gemaakt en vervolgens de consistentiecontroles aangeleverd. Deze zijn overwegend hetzelfde gebleven ten opzichte van de controles die in voorgaande edities zijn uitgevoerd. ABF Research heeft de controles verwerkt in haar eigen syntax voor Surfox. Om te voorkomen dat tijdens de imputatie nieuwe inconsistenties ontstaan, worden de consistentiecorrecties voorafgaand aan de imputaties uitgevoerd.

In de bijlage is een overzicht te vinden van deze consistentiecontroles. Hierin is ook een overzicht opgenomen van het aantal gevonden inconsistenties per controle. Na bewerking zijn alle inconsistenties uit het bestand verwijderd.

2.4 Imputatie

Een van de belangrijkste onderdelen in het traject om van een ruw bestand tot een geschikt onderzoeksbestand te komen is de bijschatting van ontbrekende waarden, ook wel imputatie genoemd. Doorgaans geven niet alle respondenten op alle vragen een antwoord. In sommige gevallen is dit terecht als gevolg van de routing (“Niet Van Toepassing”). Echter in veel gevallen blijkt dit niet terecht waardoor men met item non-respons te maken heeft (“Ten Onrechte Geen Antwoord”). Daarnaast worden voor de meeste vragen in het WoON de antwoordcategorieën ‘weigert’ en ‘weet niet’ ook als non-respons beschouwd.

Ontbrekende waarden zijn met behulp van imputatietechnieken bij te schatten. ABF Research is sinds het WBO 1998 altijd betrokken geweest bij de imputaties van ontbrekende antwoorden.

2.4.1 Waarom imputeren?

Bij bijna alle enquêtes heeft men te maken met non-respons. Indien een respondent weigert deel te nemen, spreekt men van unit non-respons. Via ophogingstechnieken worden weegvariabelen gemaakt of aangepast om dit soort non-respons te corrigeren. Hierbij gaat men er dan van uit dat de groep die meedoet hetzelfde gedrag vertoont als diegenen die wel hebben meegedaan en – in het geval van het WoON – dezelfde leeftijd, geslacht, geboorteland, positie in het huishouden en woongebied hebben. Indien eenmaal niet meer informatie bekend is van de respondenten is herweging de beste correctiemethode.

Anders ligt het bij respondenten die wel hebben meegedaan, maar die niet op alle vragen hebben geantwoord. Deze vorm van non-respons wordt item non-respons genoemd. Soms weet men het antwoord echt niet en in andere gevallen weigert men een antwoord te geven. Dit laatste gebeurt vaker bij variabelen waarbij bedragen worden gevraagd (huurprijs, woningwaarde, inkomen, hypotheek).

Indien een onderzoeksbestand item non-respons bevat liggen er twee grote gevaren op de loer. Het eerste gevaar is dat iedere gebruiker zijn eigen weg gaat om een oplossing te zoeken voor het verwijderen van de missende waarden. Voor de hand liggende methodes hiervoor zijn onder andere: herwegen per tabel, schatten met behulp van een regressiemodel, of alle ontbrekende waarden vervangen door het gemiddelde. Enerzijds kleven aan alle bovenstaande methoden bezwaren (tijdrovend, onderschatting van de spreiding, of verkeerde veronderstelling over het gedrag van de non-respondent), en anderzijds brengen verschillende onderzoekers zodoende ook verschillende getallen naar buiten voor dezelfde vraagstukken (variabelen).

Het tweede gevaar is dat een onderzoeker niets doet met de ontbrekende waarden en dus impliciet veronderstelt dat non-respons verdeeld is conform de respons – men moet zich echter wel realiseren dat niet bijschatten in feite ook een vorm van bijschatten is. Menig onderzoek heeft echter uitgewezen dat deze veronderstelling meestal onjuist is, omdat non-respons vaak selectief is. Deze hangt meestal samen

met bepaalde kenmerken van de respondent. Als er voor gekozen wordt om de non-respons op een of ander manier bij te schatten, dan wordt expliciet verondersteld dat de bijschatting tot een meer betrouwbare voorspelling leidt. De methode waarmee de non-respons wordt bijgeschat bepaalt dus in hoge mate de kwaliteit van het bestand in termen van geldigheid en betrouwbaarheid.

De methode om de ontbrekende waarden bij te schatten is de methode van imputatie. Little en Rubin uit de Verenigde Staten hebben vanaf 1980 baanbrekend werk gedaan op dit terrein. Er bestaan diverse imputatietechnieken waarmee praktisch aan alle bezwaren van bestaande 'oplossingen' tegemoet wordt gekomen.

Alle technieken hebben gemeen dat de 'verdeling' van de geïmputeerde waarden goed blijft. Indien een persoon een vraag niet heeft beantwoord, wordt naar andere personen gezocht die de vraag wel hebben beantwoord. De personen die zoveel mogelijk dezelfde hiermee samenhangende kenmerken hebben als de persoon met het ontbrekende antwoord, vormen de donorgroep. Uit deze donorgroep wordt willekeurig een persoon getrokken. De wel ingevulde waarde door deze donor wordt bij de ontbrekende waarde ingevuld. De kenmerken op basis waarvan de donorgroep samengesteld wordt vormen een zogenaamde relatietabel.

Imputatiemethoden hebben de eigenschap dat niet getracht wordt om de beste schatting op recordniveau te vinden, maar om de verdeling van de bij te schatten doelvariabele zo veel mogelijk te handhaven gegeven de kenmerken van de donoren. Dit heeft het grote voordeel dat de variantie van de bijgeschatte variabele aanzienlijk minder onderschat wordt in vergelijking met andere technieken.

2.4.2 Voorbereidingen imputatie

Routing vragenlijst

Een belangrijk onderdeel voordat met de imputaties begonnen kan worden is een correcte routing. Door gebruik van computergestuurde vragenlijsten is het mogelijk om een zeer complexe routingstructuur in de vragenlijst aan te leggen. Voor het imputatieproces was het noodzakelijk de routing geheel en perfect na te bouwen. Voor elke niet beantwoorde vraag moet namelijk worden nagegaan of deze een NVT (Niet Van Toepassing) is of TOGA (Ten Onrechte Geen Antwoord). Een bijschatting van een routingvraag heeft consequenties voor alle daaropvolgende vragen. Deze zijn eerst NVT, maar kunnen door de bijgeschatte routingvraag TOGA worden (vervolg non-respons). Eerder is in paragraaf 2.1 beschreven op welke wijze de routing is gecontroleerd.

Registerdata

Om bij de imputaties gebruik te kunnen maken van zoveel mogelijk relevante extra informatie heeft CBS registerdata aan het veldwerkbestand gekoppeld. Hierbij gaat het om BRP-gegevens, de WOZ-waarde, de huurtoeslag, inkomensbegrippen, energieverbruik en regionale indelingen.

2.4.3 Imputatiemethoden

De methode zoals in paragraaf 2.4.1 is uiteengezet wordt toegepast bij bijna alle imputaties. Deze methode heet Random Hotdeck Methode (RHM). In een aantal gevallen is gebruik gemaakt van simultaan

imputeren. Indien een groep variabelen ongeveer hetzelfde aantal ontbrekende waarden hebben die daarbij sterk met elkaar samenhangen, dan worden deze tegelijkertijd bijgeschat. Onderwerpen die door een gewijzigde routing van 'Niet van Toepassing' (NVT) naar 'Ten Onrechte Geen Antwoord' (TOGA) worden getransformeerd komen hiervoor meestal in aanmerking.

Een goed voorbeeld hiervan betreft de imputatie van de huur die een respondent voor de huidige woning betaalt. Voor sommige huurders zijn zaken als water, verwarming en elektriciteit inbegrepen en voor andere weer niet. Dit moet echter wel in de huurprijs zichtbaar blijven omdat een huurprijs inclusief andere zaken doorgaans hoger is dan exclusief. In het geval de huurprijs ten onrechte niet opgegeven is, blijven ook de mogelijk inbegrepen posten 'leeg'. Om tot een logische bijschatting voor de huurprijs te komen gelet op de inbegrepen posten, zal hier rekening mee gehouden moeten worden door de inbegrepen posten gelijktijdig eveneens bij te schatten.

De imputatie wordt afgestemd op een centrale variabele (huurprijs) waarbij gebruik gemaakt wordt van RHM. Indien eenmaal een geschikte donor gevonden is, worden alle samenhangende variabelen simultaan overgenomen.

2.4.4 Uitvoering imputatie

Volgorde van bijschatting

WoON is zo opgezet dat de relatief eenvoudig te beantwoorden vragen in het begin staan en de privacygevoelige vragen aan het einde van de vragenlijst: eerst de demografische gegevens over het huishouden, dan achtereenvolgens de vragen over de huidige woning, eventueel de vorige woning, eventueel de gewenste woning, en tot slot werk en opleiding van zowel de ondervraagde als van de partner.

Elk blok kent zijn centrale variabelen, zoals de huur- en koopprijs bij de huidige woning. Deze vragen hebben meestal betrekking op bedragen. Doorgaans komt non-respons bij 'bedragen-vragen' wat vaker voor. De non-respons neemt ook toe naarmate de vragenlijst vordert. Dit heeft niet alleen te maken met afgebroken interviews maar ook met de privacy-gevoeligheid van de vragen. De opzet van WoON zorgt ervoor dat sequentieel de blokken kunnen worden afgelopen.

Consistentiecontroles

Hoewel de consistentiecontroles voor de imputatie zijn uitgevoerd, kan het alsnog voorkomen dat een geïmputeerde waarde in bepaalde combinaties met andere waarden toch niet consistent is. Daarom wordt zekerheidshalve na elke imputatie van een variabele het relevante deel van het consistentieprogramma herhaald.

Ook extreme waarden kunnen bij een bepaalde omvang tot relevante beleidsinformatie leiden. Het probleem daarbij is dat de extreme waarden op landelijk niveau wel tot goede gemiddelden en verdelingen kunnen leiden, maar op regionaal niveau voor grote uitschieters kunnen zorgen. De uitkomsten van het WoON zullen doorgaans als gemiddelden (en niet als mediaan) gepresenteerd worden, waardoor de extreme waarden – vanuit regionaal oogpunt – extra moeten worden gecontroleerd. Om te voorkomen dat extreme waarden als donor gaan dienen, worden deze records buiten de donorgroepen gelaten. Dit voorkomt overmatig imputatie van extreme waarden.

Gebruikte software en output

De imputaties zijn uitgevoerd met het pakket Surfox dat ABF Research ontwikkeld heeft. Ook de voorgaande onderzoeken zijn hiermee bijgeschat. Surfox kent een interactieve versie en een batchversie Surbox. Met Surfox zijn de imputatiemodellen gemaakt, de zogenaamde relatietabellen, waaruit de verklaringsgraad is af te lezen. De resultaten, zoals het gemiddeld aantal donorrecords en een theoretisch percentage voor correcte bijstellingen, worden bewaard in een outputfile.

In totaal zijn 436 variabelen (met inbegrip hulpvariabelen) in WoON 2015 bijgeschat. Het laagste percentage correct is 56,3%. Circa 95% van de imputaties hebben geleid tot een percentage correct van minimaal 70%. Rekening houdend met het aantal records dat per variabele geïmputeerd is, komt het gewogen gemiddelde van het overall percentage correct uit op 83%.

Tabel 2.1 Aantal imputaties per klasse van percentage correct

Klasse	Aantal
minder dan 50%	0
50% - 60%	2
60% - 70%	19
70% - 80%	91
80% - 90%	190
90% - 100%	123
100%	11
Totaal	436

Het maken van de relatietabellen is handwerk en moet daarom interactief gebeuren. Na de imputatie van een variabele kan de routing ervoor zorgen dat variabelen die eerst NVT waren, worden getransformeerd naar TOGA. Ook deze wijzigingen moeten handmatig worden uitgevoerd. Nadat op deze wijze het hele bestand is bijgeschat, zijn alle handelingen achter elkaar gezet in een inputbestand. Dit inputbestand is geschikt voor de batchversie Surbox. Omdat na elk blok controles worden uitgevoerd met betrekking tot extreme waarden, inconsistenties en dergelijke is de inputfile in blokken opgeknipt. Bij geconstateerde fouten was het daarom niet nodig om het gehele bestand opnieuw te draaien. De output met daarin de relatietabellen inclusief verklaringsgraad en de imputatieresultaten beslaan ongeveer 50.000 regels.

Imputatie hypotheekblok

Ten opzichte van de voorgaande vragenlijst is het blok met betrekking tot de koopwoning op een aantal punten drastisch gewijzigd. Dit betreft voornamelijk de vragen die betrekking hebben op de hypotheek. Voor het stappenplan dat uiteindelijk gevolgd is om de vragen in het hypotheekblok wordt verwezen naar de bijlage.

Imputatie van WOZ-waarde en huurpunten

Op verzoek van CBS en BZK zijn de ontbrekende WOZ-waarde en huurpunten (uiteraard alleen voor respondenten in een huurwoning) bijgeschat. In totaal gaat het om 1499 bijstellingen van de WOZ-waarde, waarvoor uiteindelijk een gemiddeld percentage correct van 69% is behaald. Voor de bijstelling van de huurpunten is een aparte imputatieprocedure uitgevoerd. Deze is in de bijlage van dit rapport opgenomen.

2.5 Bepalen gewenste woonmilieus

Het gewenste woonmilieu is een belangrijke variabele voor BZK. Het is dan ook aan de ene kant van groot belang dat voor de berekening van deze variabele aangesloten wordt bij de werkwijze die in het verleden gehanteerd is. Aan de andere kant heeft men te maken met veranderingen in de vragenlijst om op een nog betere manier de gewenste woonmilieus af te leiden.

Het gewenste woonmilieu is deels afgeleid en deels geïmputeerd. Aan iedereen die wil of moet verhuizen is gevraagd naar welke woonplaats en naar welke wijk of buurt binnen die woonplaats men wil verhuizen. Als een gewenste wijk of buurt bekend is (en dus gecodeerd kan worden), dan wordt hier het (gewenste) woonmilieu aan gekoppeld. De gekoppelde woonmilieus zijn gebaseerd op de meeste recente typologie die beschikbaar is, te weten de typologie van 2012.

Is de buurt niet bekend en ook niet te typeren naar een bestaande wijk- of buurtcode, dan wordt het woonmilieu geïmputeerd op basis van de antwoorden op vragen als de gewenste regio, de gewenste woonplaatsgrootte, de gewenste afstand tot het centrum van de woonplaats, de gewenste bouwperiode van de buurt, het gewenste woningtype van de buurt en de gewenste menging in de buurt. Voor een betrouwbare imputatie is het daarom van groot belang dat zoveel mogelijk respondenten een gewenste buurt noemen.

2.6 Bijkomende woonlasten

De vaste woonuitgaven betreffen de uitgaven van huurders en eigenaren aan respectievelijk huur en hypotheek. Daarnaast geven bewoners geld uit voor het feitelijk gebruik van de woning. Deze uitgaven worden aangeduid als de bijkomende woonlasten. Het gaat hierbij om heffingen van gemeenten en waterschappen (heffingen van Openbaar Publiekrechtelijke Lichamen ofwel OPL) en uitgaven aan nutsvoorzieningen zoals water, gas, warmte en elektriciteit.

Het WoON bevat een grote hoeveelheid aan informatie over de woonsituatie en woonuitgaven van de respondent. Daarnaast wordt een deel van de informatie uit aanvullende bronnen gehaald. Indien een bronhouder voor een gemeente, waterschap of nutsbedrijf geen tarieven heeft kunnen aanleveren, heeft ABF Research deze gegevens in eigen beheer verzameld. De aanvullende informatie betreft:

- de gemeentelijke lasten;
- de waterschapsheffingen;
- de kosten van energie en water.

2.6.1 Gemeentelijke lasten

Heffingen

ABF Research heeft per gemeente de tarieven verzameld van de relevante gemeentelijke heffingen, te weten onroerend zaakbelasting, rioolrechten en afvalstoffenheffing. Deze gegevens zijn bewerkt zodat deze gebruikt konden worden in de rekenprogrammatuur.

2.6.2 Waterschapsheffingen

Heffingen

ABF Research heeft per waterschap de tarieven verzameld van de relevante waterschapsheffingen, te weten watersysteemheffing ingezetenen, watersysteemheffing gebouwd en zuiveringsheffing. Op basis van het werkgebied en het takenpakket van het waterschap berekent ABF Research per gemeente de omvang van de tarieven. Deze gegevens zijn bewerkt zodat deze gebruikt konden worden in de rekenprogrammatuur.

Grondslagen

De watersysteemheffing ingezetenen betreft een bedrag per woonruimte (huishouden). De watersysteemheffing gebouwd is gebaseerd op de WOZ-waarde. En de zuiveringsheffing wordt bepaald aan de hand van het aantal leden binnen het huishouden. Er geldt een tarief per vervuilingseenheid (VE). De heffing van een eenpersoonshuishouden is gebaseerd op 1 VE en die van een meerpersoonshuishouden op 3 VE.

2.6.3 Verbruik en kosten van energie en water

De tarieven en heffingen heeft ABF Research in eigen beheer verzameld. De verbruiksgegevens met betrekking tot energie (elektriciteit, gas en warmte) heeft ABF Research via CBS verkregen.

Tarieven

ABF Research heeft de tarieven van gas, warmte en elektriciteit van diverse energiebedrijven verzameld. In verband met het vrijgeven van de energiemarkt per 1 juli 2004 wordt aangaande de leveringstarieven uitgegaan van een landelijk gemiddelde. Vewin levert de tarieven van de waterbedrijven. ABF Research houdt een overzicht bij welke waterbedrijven in welke gemeente leverancier is.

Verbruik

De verbruiksgegevens zijn door CBS gekoppeld, mits de respondent daar toestemming voor heeft gegeven en de respondent volgens het WoON in een woning woont. CBS heeft verbruiksgegevens geïmputeerd als er geen verbruik beschikbaar is uit registratie of als het om een bedrijfsruimte gaat. Het gasverbruik is ook geïmputeerd als er stadsverwarming aanwezig is in de woning of als het gasverbruik nul is.

3

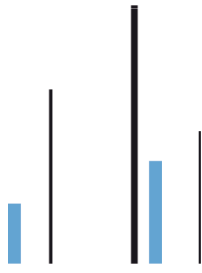


Eindproducten

Het uiteindelijke resultaat van de dataprocessing die ABF Research verricht heeft is een eindbestand dat exact hetzelfde ingericht is als het veldwerkbestand, maar dan gevuld met de bewerkte enquêtedata. Verder zijn separaat de extra aangemaakte variabelen opgeleverd, waaronder alle relevante variabelen die tijdens de berekening van de woonlasten zijn aangemaakt.

Daarnaast is ook een vlaggenbestand gemaakt. In dit bestand worden de waarden van de originele variabelen vergeleken met de waarden na de dataprocessing. Het vlaggenbestand bevat alle variabelen van de module Woningmarkt, maar elke variabele heeft een waarde die aangeeft of de waarde voor het betreffende record een ongewijzigde waarde, een consistentieaanpassing, of een bijgeschatte waarde bevat.

Tot slot is ook alle syntax beschikbaar gesteld. Dit betreft zowel de syntax voor SPSS als voor Surfox. Beide softwareprogramma's zijn gebruikt tijdens de uitvoering van de verschillende onderdelen van de dataprocessing.



Bijlage 1 Overzicht typeringen

Als onderdeel van de dataprocessing van het WoON 2015 heeft ABF Research 26 vragen met een open karakter getypeerd. In dit document wordt per getypeerde variabele een overzicht gegeven van het aantal geslaagde typeringen voor de omzetting naar reeds bestaande antwoordcategorieën. Indien nuttig is aanvullend ook aangegeven welke typeringen gemaakt hadden kunnen worden als het antwoordcategorie (thema, zoekterm, kernwoord) beschikbaar was geweest. Een aantal typeringen gaan over meerdere vragen tegelijkertijd. Hiertoe behoren de vragen/variabelen: Anders, Anders2, EigHuurC_Andrs, NSAndOP, NSAndPA, TusV_Omschr, VAnders en VEigHuurC_andrs.

ActBrn_Andrs (17.42) – totaal 6511 strings (6009 niet getypeerd)

Geslaagde typeringen

- 113 strings getypeerd als Dicho_ActBrn_1.
- 34 strings getypeerd als Dicho_ActBrn_2.
- 205 strings getypeerd als Dicho_ActBrn_3.
- 20 strings getypeerd als Dicho_ActBrn_4.
- 139 strings getypeerd als Dicho_ActBrn_5.

Nieuwe typeringen

- 108 Dicho_ActBrn_9 'Mantelzorg geven'.
- 240 Dicho_ActBrn_10 'Wegbrengen of ophalen met de auto'.
- 467 Dicho_ActBrn_11 'Vuil naar de container brengen of de container wegzetten/terugzetten'.
- 773 Dicho_ActBrn_12 'Pakketjes/post aannemen of post meenemen uit de brievenbus'.
- 338 Dicho_ActBrn_13 'Eten koken/brengen'.
- 192 Dicho_ActBrn_14 'Stoep/straat/portiek schoonmaken of sneeuwvrij maken'.
- 103 Dicho_ActBrn_15 'Helpen bij administratie'.
- 1150 Dicho_ActBrn_16 'Oppassen op de kinderen, bijles geven aan de kinderen of de kinderen uit school halen'.
- 456 Dicho_ActBrn_17 'Spullen uitlenen'.

Nieuwe variabele labels

- Dicho_ActBrn_1 veranderd in “Boodschappen doen/dragen/brengen”.
- Dicho_ActBrn_5 veranderd in “Oogje in het zeil houden, planten verzorgen, post aannemen of brievenbus legen bij afwezigheid”.

ActBrt_Andrs (17.42) – totaal 2335 strings (2053 niet getypeerd)

Geslaagde typeringen

- 26 strings getypeerd als Dicho_ActBrt_1.

83 strings getypeerd als Dicho_ActBr_t_2.

91 strings getypeerd als Dicho_ActBr_t_3.

4 strings getypeerd als Dicho_ActBr_t_4.

57 strings getypeerd als Dicho_ActBr_t_5.

49 strings getypeerd als Dicho_ActBr_t_6.

Nieuwe typeringen

575 Dicho_ActBr_t_10 'Ik ben actief in een buurtvereniging'.

93 Dicho_ActBr_t_11 'Sociale controle'. (Heeft een wat negatieve lading, wellicht neutralere term gebruiken).

Anders (7.6) – totaal 924 strings (294 niet getypeerd)

Geslaagde typeringen

126 strings getypeerd als 1. Flat, appartement, etagewoning, boven- of benedenwoning (SrtWon, 7.2).

88 strings getypeerd als 2. Rijtjeshuis, tussenwoning, hoekwoning (SrtWon, 7.2).

138 strings getypeerd als 3. Half-vrijstaande woning (SrtWon, 7.2).

6 strings getypeerd als 4. Vrijstaande woning (SrtWon, 7.2).

34 strings getypeerd als 5. Boerderij, woning met tuindersbedrijf (SrtWon, 7.2).

6 strings getypeerd als 6. Woning met aparte winkel, kantoor-, praktijk- of bedrijfsruimte (SrtWon, 7.2).

32 strings getypeerd als 8. Geen van deze (SrtWon, 7.2).

1 string getypeerd als 1. Ja (Bowigebr, 7.3).

7 strings getypeerd als 2. Nee (Bowigebr, 7.3).

47 strings getypeerd als 1. Maisonnette (SrtApp, 7.8).

23 strings getypeerd als 2. Bovenwoning (SrtApp, 7.8).

30 strings getypeerd als 3. Benedenwoning (SrtApp, 7.8).

105 strings getypeerd als 4. Appartement / etagewoning (SrtApp, 7.8).

46 strings getypeerd als 5. Flat (SrtApp, 7.8).

24 strings getypeerd als 1. Kamer(s) in een studentenhuis of bij een hospita (SrtWonAndrs, 7.4.5).

20 strings getypeerd als 3. Recreatie- of vakantiewoning (SrtWonAndrs, 7.4.5).

1 string getypeerd als 4. Woonwagen, caravan (SrtWonAndrs, 7.4.5).

21 strings getypeerd als 7. Verzorgings- of verpleeghuis (SrtWonAndrs, 7.4.5).

4 strings getypeerd als 1. (\$5/\$6/\$8: Rijtjeshuis / tussenwoning) (TypWon, 7.7).

20 strings getypeerd als 2. (\$5/\$6/\$7/\$8: Hoekwoning) (TypWon, 7.7).

81 strings getypeerd als 3. (\$4/\$5/\$7/\$8: 2 onder 1 kap woning) (TypWon, 7.7).

29 strings getypeerd als 4. (\$4/\$5/\$7/\$8: Geschakelde woning) (TypWon, 7.7).

10 strings getypeerd als 1. Een eigen opgang of ingang aan straat (SrtFlat, 7.9).

1 strings getypeerd als 2. Een voordeur in een gedeeld portiek (SrtFlat, 7.9).

1 strings getypeerd als 3. Een voordeur aan een al dan niet overdekte galerij (SrtFlat, 7.9).

Anders2 (26.2.3) – totaal 341 strings (161 niet getypeerd)

Geslaagde typeringen

29 strings getypeerd als 1. Flat, appartement, etagewoning, boven- of benedenwoning (GSrtWon, 26.1).

138 strings getypeerd als 2. Rijtjeshuis, tussenwoning, hoekwoning, half-vrijstaande woning, vrijstaande woning (GSrtWon, 26.1).

9 strings getypeerd als 3. Boerderij, woning met tuindersbedrijf (GSrtWon, 26.1).

1 strings getypeerd als 4. Woning met aparte winkel, kantoor-, praktijk- of bedrijfsruimte (GSrtWon, 26.1).

1 strings getypeerd als 7. Woning die voorzieningen deelt, zoals was- en drooggelegenheid of hobbyruimte met andere woningen uit het project (centraal wonen) (GSrtWon, 26.1).

1 strings getypeerd als 8. Geen van deze (GSrtWon, 26.1).

1 strings getypeerd als 3. Recreatie- of vakantiewoning (GSrtWonAndrs, 26.2.3).

4 strings getypeerd als 1. Rijtjeshuis / tussenwoning (GWonTyp, 26.11).

7 strings getypeerd als 3. 2 onder 1 kap woning (GWonTyp, 26.11).

4 strings getypeerd als 4. Geschakelde woning (GWonTyp, 26.11).

50 strings getypeerd als 5. Vrijstaande woning (GWonTyp, 26.11).

40 strings getypeerd als 6. Geen van deze (GWonTyp, 26.11).

15 strings getypeerd als 4. Andere woonvorm (GSrtBejW1, 5.3.1).

9 strings getypeerd als 1. Een ouderenwoning, bejaardenwoning, seniorenwoning (GSrtBejW2, 5.3.2).

6 strings getypeerd als 2. Een aanleunwoning (GSrtBejW2, 5.3.2).

ASrtHypVorm (14.7.1) – totaal 270 strings (170 niet getypeerd)

Geslaagde typeringen

41 strings getypeerd als Dicho_HypVorm_1

35 strings getypeerd als Dicho_HypVorm_2

16 strings getypeerd als Dicho_HypVorm_3

13 strings getypeerd als Dicho_HypVorm_4

2 strings getypeerd als Dicho_HypVorm_5

1 strings getypeerd als Dicho_HypVorm_6

8 strings getypeerd als Dicho_HypVorm_7

4 strings getypeerd als Dicho_HypVorm_8

ASrtHypVorm2 (14.7.2) – totaal 116 strings (98 niet getypeerd)

Geslaagde typeringen

11 strings getypeerd als Dicho_HypVorm_13

3 strings getypeerd als Dicho_HypVorm_14

1 string getypeerd als Dicho_HypVorm_15

1 string getypeerd als Dicho_HypVorm_16

0 strings getypeerd als Dicho_HypVorm_17

0 strings getypeerd als Dicho_HypVorm_18

0 strings getypeerd als Dicho_HypVorm_19

2 strings getypeerd als Dicho_HypVorm_20

ASrtHypVorm3 (14.7.3) – totaal 28 strings (28 niet getypeerd)

-

EigHuurC_Andrs (3.3.1) – totaal 343 strings (128 niet getypeerd)

Geslaagde typeringen

10 strings getypeerd als 1. Ja (EigHuurA, 3.1).

13 strings getypeerd als 2. Nee (EigHuurA, 3.1).

13 strings getypeerd als 1. Ja (EigHuurB, 3.2).

0 strings getypeerd als 2. Nee (EigHuurB, 3.2).

3 strings getypeerd als 1. Corporatie, woningbouwvereniging of gemeentelijk woonbedrijf (WieVerh, 10.1).

1 strings getypeerd als 2. Rijk, de provincie of het waterschap (WieVerh, 10.1).

2 strings getypeerd als 3. Pensioenfonds, verzekeringsmaatschappij, belegger of een makelaar (WieVerh, 10.1).

98 strings getypeerd als 1. Woon gratis (EigHuurC, 3.3).

2 strings getypeerd als 2. Woon antikraak (EigHuurC, 3.3).

27 strings getypeerd als 1. Woon bij familie in huis (Gratis, 3.3.2).

2 strings getypeerd als 2. Woon bij een ander huishouden in huis (Gratis, 3.3.2).

24 strings getypeerd als 3. Woon in een woning van familie, vrienden of kennissen (Gratis, 3.3.2).

15 strings getypeerd als 4. Woon in een woning van de werkgever (Gratis, 3.3.2).

0 strings getypeerd als 5. Woon in een leegstaande woning, kantoor of bedrijfspand (Gratis, 3.3.2).

29 strings getypeerd als 6. Andere reden (Gratis, 3.3.2).

92 strings getypeerd als antwoordoptie 1 [Gratis] in EigHuurC (3.3) en als antwoordoptie 1 [InFamH] in Gratis (3.3.2).

31 getypeerd als antwoordoptie 1 [Gratis] in EigHuurC (3.3) en als verschillende antwoordopties in Gratis (3.3.2).

50 getypeerd als antwoordoptie 1 [Gratis] in EigHuurC (3.3) en als antwoordoptie 6 [Andrs] in Gratis (3.3.2).

NSAndOP (33.4) – totaal 1352 strings (968 niet getypeerd)

Geslaagde typeringen

1 string getypeerd als 2. Vmbo, lwoo, vso (NivActOP).

2 strings getypeerd als 3. Havo (NivActOP).

7 strings getypeerd als 5. Mbo (pdb, mba) (NivActOP).

9 strings getypeerd als 6. Hbo (associate degree) (NivActOP).

6 strings getypeerd als 7. Universitaire opleiding (NivActOP).

1 string getypeerd als 2. Theoretische of gemengde leerweg (NA_Vmbo_OP).

5 strings getypeerd als Nederlandse taalvaardigheid / inburgeringscursus (NSOpIOP).

354 strings getypeerd als Vakgerichte of bedrijfsgerichte opleiding of cursus (NSOpIOP).

Nieuwe typeringen

260 strings kunnen getypeerd worden als “Een (buitenlandse) taalcursus”. “Een (buitenlandse) taalcursus” is toegevoegd aan NSOpIOP als categorie 4.

NSAndPA (35.4) – totaal 433 strings (328 niet getypeerd)

Geslaagde typeringen

1 string getypeerd als 2. Vmbo, lwo, vso (NivActPA).

3 strings getypeerd als 5. Mbo (pdb, mba) (NivActPA).

3 strings getypeerd als 6. Hbo (associate degree) (NivActPA).

4 strings getypeerd als 7. Universitaire opleiding (NivActPA).

1 string getypeerd als 2. Theoretische of gemengde leerweg (NA_Vmbo_PA).

1 string getypeerd als 1. Nederlandse taalvaardigheid / inburgeringscursus (NSOpIPA).

93 strings getypeerd als 2. Vakgerichte of bedrijfsgerichte opleiding of cursus (NSOpIPA).

Nieuwe typeringen

102 strings kunnen getypeerd worden als “Een (buitenlandse) taalcursus”. “Een (buitenlandse) taalcursus” is toegevoegd aan NSOpIPA als categorie 4.

NtParhw_Andrs (20.24) – totaal 213 strings (180 niet getypeerd)

Geslaagde typeringen

11 strings getypeerd als Dicho_NtParhw _1.

5 strings getypeerd als Dicho_NtParhw _2.

10 strings getypeerd als Dicho_NtParhw _3.

7 strings getypeerd als Dicho_NtParhw _4.

Nieuwe typeringen

67 Dicho_NtParhw_11 'Reden hing niet samen met particuliere verhuurder'.

NtSochw_Andrs (20.29) – totaal 105 strings (89 niet getypeerd)

Geslaagde typeringen

5 strings getypeerd als Dicho_NtSochw _1.

1 strings getypeerd als Dicho_NtSochw _2.

1 strings getypeerd als Dicho_NtSochw _3.

1 strings getypeerd als Dicho_NtSochw _4.

1 strings getypeerd als Dicho_NtSochw _5.

2 strings getypeerd als Dicho_NtSochw _6.

5 strings getypeerd als Dicho_NtSochw _7.

Nieuwe typeringen

49 Dicho_NtSochw_11 'Reden hing niet samen met sociale verhuurder'.

OmgvVH_Andrs (20.18.1) – totaal 109 strings (87 niet getypeerd)

Geslaagde typeringen

6 strings getypeerd als Dicho_OmgvVH_1.

1 string getypeerd als Dicho_OmgvVH_2.

4 strings getypeerd als Dicho_OmgvVH_5.

10 strings getypeerd als Dicho_OmgvVH_6.

1 strings getypeerd als Dicho_OmgvVH_7.

2 strings getypeerd als Dicho_OmgvVH_8.

Omschre (10.3) – totaal 1200 strings (789 niet getypeerd)

Geslaagde typeringen

20 strings getypeerd als '2. Rijk, de provincie of het waterschap' (WieVerh, 10.1).

377 strings getypeerd als '3. Pensioenfonds, verzekeringsmaatschappij, belegger of een makelaar' (WieVerh, 10.1).

14 strings getypeerd als '6. (\$1: Zorginstelling)' (WieVerh, 10.1).

OudwVH_Andrs (20.19.1) – totaal 62 strings (51 niet getypeerd)

Geslaagde typeringen

1 string getypeerd als Dicho_OudwVH_1.

3 strings getypeerd als Dicho_OudwVH_5.

4 strings getypeerd als Dicho_OudwVH_6.

2 strings getypeerd als Dicho_OudwVH_8.

1 string getypeerd als Dicho_OudwVH_9.

RedNietVH2_Andrs (25.3) – totaal 2139 strings (1826 niet getypeerd)

Geslaagde typeringen

95 strings getypeerd als Dicho_RedNietVH1_1.

332 strings getypeerd als Dicho_RedNietVH1_2.

26 strings getypeerd als Dicho_RedNietVH1_3.

7 strings getypeerd als Dicho_RedNietVH1_4.

45 strings getypeerd als Dicho_RedNietVH1_6.

1 strings getypeerd als Dicho_RedNietVH2_1.

71 strings getypeerd als Dicho_RedNietVH2_2.

1 string getypeerd als Dicho_RedNietVH2_3.

12 strings getypeerd als Dicho_RedNietVH2_4.

12 strings getypeerd als Dicho_RedNietVH2_5.

Nieuwe typeringen

836 Dicho_RedNietVH2_9 'Familie, vrienden en/of kennissen wonen dichtbij'.

123 Dicho_RedNietVH2_10 'Er zijn veel voorzieningen in de nabijheid'.

TOMgvVH_Andrs (25.22.1) – totaal 744 strings (633 niet getypeerd)

Geslaagde typeringen

5 strings kunnen getypeerd worden als Dicho_TOMgvVH_1.

3 strings kunnen getypeerd worden als Dicho_TOMgvVH_2.

15 strings kunnen getypeerd worden als Dicho_TOMgvVH_3.

12 strings kunnen getypeerd worden als Dicho_TOMgvVH_4.

19 strings kunnen getypeerd worden als Dicho_TOMgvVH_5.

55 strings kunnen getypeerd worden als Dicho_TOMgvVH_6.

8 strings kunnen getypeerd worden als Dicho_TOMgvVH_7.

18 strings kunnen getypeerd worden als Dicho_TOMgvVH_8.

3 strings kunnen getypeerd worden als Dicho_TOMgvVH_9.

Nieuwe typeringen

72 Dicho_TOMgvVH_13 'Er is te weinig groen'.

52 Dicho_TOMgvVH_14 'Er zijn te weinig voorzieningen in de buurt'.

21 Dicho_TOMgvVH_15 'Het is te rustig'.

124 Dicho_TOMgvVH_16 'Het is te druk'.

72 Dicho_TOMgvVH_17 'Het is ongezellig / de sfeer is slecht'.

TRedenVH_Andrs (25.18) – totaal 4363 strings (3278 niet getypeerd)

Geslaagde typeringen

43 strings getypeerd als Dicho_TRedenVH_1.

7 strings getypeerd als Dicho_TRedenVH_2.

75 strings getypeerd als Dicho_TRedenVH_3.

180 strings getypeerd als Dicho_TRedenVH_4. Strings alleen toegevoegd als respondent expliciet aangeeft om vanuit financiële overwegingen van een huurwoning naar een koopwoning te gaan. TFinanVH (25.20.1) krijgt antwoordcategorie 2 [Andrs] in dit geval. Als respondent aangeeft 'gewoon' een woning te willen kopen, dan: TRedenVH (25.17) krijgt antwoordcategorie 5 [Woning] en TWoningVH (25.21) krijgt antwoordcategorie 4 [Koop].

219 strings getypeerd als Dicho_TRedenVH_5.

564 strings getypeerd als Dicho_TRedenVH_6.

9 strings getypeerd als Dicho_TRedenVH_7.

Typeringen naar Dicho_TRedenVH_3, Dicho_TRedenVH_4, Dicho_TRedenVH_5 en Dicho_TRedenVH_6 zijn 'vooruit getypeerd' naar, respectievelijk, TWerkVH, TFinanVH, TWoningVH en TOMgvVH.

De vier onderstaande mogelijke typeringen zijn getypeerd naar TRedenVH (25.17) antwoordcategorie 5 [Woning] en de betreffende categorie in TWoningVH (25.21):

457 strings kunnen getypeerd worden als "Ik wil een kleinere woning".

67 strings kunnen getypeerd worden als “Ik wil een kleinere/geen tuin”.

688 strings kunnen getypeerd worden als “Ik wil een grotere woning”.

278 strings kunnen getypeerd worden als “Ik wil een (grotere) tuin”.

347 strings getypeerd als antwoordcategorie 1 [Gezo] in TRedenVH (25.17).

314 respondenten geven “Huwelijk/samenwonen, scheiding of zelfstandig wonen op” als reden. Volgens de routing zouden deze antwoorden niet gegeven kunnen worden. Deze antwoorden zijn terug gecodeerd naar de betreffende categorie in TPersoms (25.16).

TusV_Omschr (3.5) – totaal 536 strings (524 niet getypeerd)

Geslaagde typering

2 strings getypeerd als 2. Nee (EigHuurA, 3.1).

2 strings getypeerd als 2. Nee (EigHuurB, 3.2).

10 strings getypeerd als 2. Nee (TussenV, 3.4).

VAnders (7.6) – totaal 256 strings (157 niet getypeerd)

Geslaagde typering

28 strings getypeerd als 1. Flat, appartement, etagewoning, boven- of benedenwoning (VSrtWon, 21.4).

10 strings getypeerd als 2. Rijtjeshuis, tussenwoning, hoekwoning (VSrtWon, 21.4).

7 strings getypeerd als 3. Half-vrijstaande woning (VSrtWon, 21.4).

1 strings getypeerd als 4. Vrijstaande woning (VSrtWon, 21.4).

1 strings getypeerd als 5. Boerderij, woning met tuindersbedrijf (VSrtWon, 21.4).

1 strings getypeerd als 6. Woning met aparte winkel, kantoor-, praktijk- of bedrijfsruimte (VSrtWon, 21.4).

16 strings getypeerd als 8. Geen van deze (VSrtWon, 21.4).

4 strings getypeerd als 1. Maisonnette (VSrtApp, 21.3.2).

6 strings getypeerd als 2. Bovenwoning (VSrtApp, 21.3.2).

4 strings getypeerd als 3. Benedenwoning (VSrtApp, 21.3.2).

39 strings getypeerd als 4. Appartement / etagewoning (VSrtApp, 21.3.2).

2 strings getypeerd als 5. Flat (VSrtApp, 21.3.2).

14 strings getypeerd als 1. Kamer(s) in een studentenhuis of bij een hospita (VSrtWonAndrs, 21.4).

2 strings getypeerd als 3. Recreatie- of vakantiewoning (VSrtWonAndrs, 21.4).

5 strings getypeerd als 7. Verzorgings- of verpleeghuis (VSrtWonAndrs, 21.4).

1 strings getypeerd als 1. (\$5/\$6/\$8: Rijtjeshuis / tussenwoning) (VTypWon, 21.3.3).

3 strings getypeerd als 3. (\$4/\$5/\$7/\$8: 2 onder 1 kap woning) (VTypWon, 21.3.3).

5 strings getypeerd als 4. (\$4/\$5/\$7/\$8: Geschakelde woning) (VTypWon, 21.3.3).

2 strings getypeerd als 1. Een eigen opgang of ingang aan straat (VSrtFlat, 21.7).

2 strings getypeerd als 2. Nee (VLift, 21.8).

VEigHuurC_andrs (3.3.1) – totaal 227 strings (206 niet getypeerd)

Geslaagde typering

9 strings getypeerd als 1. Ja (VEigHuurA, 21.1).

8 strings getypeerd als 2. Nee (VEigHuurA, 21.1).

8 strings getypeerd als 1. Ja (VEigHuurB, 21.2.1).

0 strings getypeerd als 2. Nee (VEigHuurB, 21.2.1).

32 strings getypeerd als 1. Woonde gratis (VEigHuurC, 3.3).

1 strings getypeerd als 2. Woonde antikraak (VEigHuurC, 3.3).

166 getypeerd als [Gratis] in VEigHuurC (21.2.2).

VSOplOP (33.7) – totaal 1676 strings (1090 niet getypeerd)

Geslaagde typering

4 strings getypeerd als 2. Lagere school (NivBehOP).

91 strings getypeerd als 3. Lbo of vso (NivBehOP).

8 strings getypeerd als 5. Mavo (NivBehOP).

13 strings getypeerd als 6. Havo (NivBehOP).

28 strings getypeerd als 7. Vwo, gymnasium of atheneum (NivBehOP).

349 strings getypeerd als 8. Mbo (NivBehOP).

67 strings getypeerd als 9. Hbo (NivBehOP).

26 strings getypeerd als 10. Universitair (NivBehOP).

VSOplPA (35.7) – totaal 650 strings (401 niet getypeerd)

Geslaagde typering

3 strings getypeerd als 2. Lagere school (NivBehPA).

50 strings getypeerd als 3. Lbo of vso (NivBehPA).

8 strings getypeerd als 5. Mavo (NivBehPA).

4 strings getypeerd als 6. Havo (NivBehPA).

5 strings getypeerd als 7. Vwo, gymnasium of atheneum (NivBehPA).

140 strings getypeerd als 8. Mbo (NivBehPA).

31 strings getypeerd als 9. Hbo (NivBehPA).

8 strings getypeerd als 10. Universitair (NivBehPA).

WZ_BrtMaats (32.10.1) – totaal 16856 (n.v.t.).

Geslaagde typering

788 strings kunnen getypeerd worden als Dicho_WaarZorgen_1.

325 strings kunnen getypeerd worden als Dicho_WaarZorgen_2.

59 strings kunnen getypeerd worden als Dicho_WaarZorgen_3.

2190 strings kunnen getypeerd worden als Dicho_WaarZorgen_4.

Nieuwe typeringen

2745 Dicho_WaarZorgen_11 'Economische/financiële ontwikkelingen'.

2447 Dicho_WaarZorgen_12 'Ontwikkelingen in de gezondheidszorg'.

3455 Dicho_WaarZorgen_13 'Verharding/verruwing van de maatschappij'.

3374 Dicho_WaarZorgen_14 'Externe dreiging'.

ZkParhw_Andrs (20.22) – totaal 99 strings (61 niet getypeerd)

Geslaagde typeringen

1 strings getypeerd als Dicho_ZkParhw _1.

11 strings getypeerd als Dicho_ZkParhw _2.

20 strings getypeerd als Dicho_ZkParhw _3.

7 strings getypeerd als Dicho_ZkParkw _4.

ZkSochw_Andrs (20.27) – totaal 192 strings (175 niet getypeerd)

Geslaagde typeringen

4 strings getypeerd als Dicho_ZkSochw_1.

7 strings getypeerd als Dicho_ZkSochw_2.

1 strings getypeerd als Dicho_ZkSochw_3.

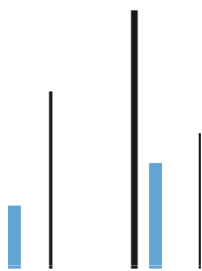
8 strings getypeerd als Dicho_ZkSochw_4.

80 strings getypeerd als antwoordcategorie 3 [InschTijd] in NtSochw (20.28).

39 strings getypeerd als antwoordcategorie 1 [NietAanm] in NtSochw (20.28).

Nieuwe typeringen

15 Dicho_ZkSochw_8 'Reden hing niet samen met sociale verhuurder'.

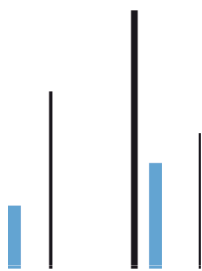


Bijlage 2 Consistentiecontroles

1. Het jaar dat de respondent op het huidige adres is komen wonen ligt niet voor het geboortjaar van de respondent.
2. Het verschil tussen het jaar van aankoop van de huidige woning en het jaar van komen wonen op dat adres mag niet meer dan vijf jaar zijn.
3. Het jaar dat de koopwoning is gekocht ligt niet voor het geboortjaar van de respondent.
4. Indien de woonkamer op de zevende verdieping of hoger ligt, is er altijd een lift aanwezig.
5. Indien de woonkamer op de zevende verdieping of hoger kan er geen sprake zijn van een eigen ingang aan de straat.
6. Het jaar dat de respondent op het vorige adres is gaan wonen ligt niet na het jaar dat de respondent op het huidige adres is komen wonen, en na het geboortjaar van de respondent.
7. Het bouwjaar van de vorige woning ligt niet voor het jaar dat de respondent op het vorige adres is gaan wonen.
8. Het bedrag aan maandelijkse hypotheekrente is niet meer dan 10% van de totale hypotheeksom.
9. De maandelijkse rentelasten mogen niet meer zijn dan totale lasten inclusief aflossing.
10. De koopprijs van een voormalige huurwoning mag niet meer dan 100 maal de (voormalige) betaalde jaarhuur is en meer dan 500.000 euro bedragen.
11. Het aantal verdiepingen met een rookmelder is niet groter dan het aantal verdiepingen van de woonruimte.
12. Het aantal verdiepingen van de woonruimte is niet groter is dan het aantal verdiepingen van het gebouw waar het deel van uitmaakt.
13. Bij bewoning van een eengezinsvorm kan er geen sprake zijn van beneden- of burens.
14. De verwachting dat men in het buitenland gaat wonen moet stroken met de toekomstige woonplaats.
15. Voor een drietal 'blokragen' (de eigen woning, de buurt en de burens) is een controle uitgevoerd op antwoordpatronen. Alle subvragen binnen een blok mogen niet hetzelfde antwoord hebben.

Tabel B2.1 Aantal gevonden inconsistenties per controle (voor bewerking)

Controle	Omschrijving	Aantal
1	Jaar op huidig adres komen wonen (2)	0
2	Jaar koopwoning gekocht (1)	274
3	Jaar koopwoning gekocht (2)	5
4	Lift bij hooggelegen woonkamer	0
5	Eigen ingang aan straat bij hooggelegen woonkamer	1
6	Jaar op vorig adres komen wonen	0
7	Bouwjaar vorige woning	0
8	Maandelijks bedrag hypotheekrente	11
9	Betaalde rente p/m vs betaalde hypotheeklasten p/m	0
10	Aankoopprijs voormalige huurwoning	2
11	Aantal verdiepingen met rookmelder	825
12	Aantal verdiepingen woning vs woongebouw	0
13	controle of eengezinswoning voorkomt in combinatie met benwon	275
14	Verhuizing buitenland	117
15	Controle op antwoordpatronen	164



Bijlage 3 Bijschatting Hypothekenblok

Voordat de syntax wordt gedraaid, zijn de guldenbedragen al naar euro's geconverteerd. Dus zelfs al staat de munteenheid variabele op 2 (gulden), dan zal toch het bijbehorende bedrag al in euro's zijn uitgedrukt. Alle munteenheid variabelen die op 2 gulden) staan, worden derhalve hard op 1 (euro) gezet.

Stap 0

Het probleem met Blok 14 zit hem in de consistentie tussen variabelen binnen één loop (Hypo[i].Jaar t/m Hypo[i].LpThyp) en in de consistentie tussen de loops en de totaalbedragen (TotBetHyp en TotBetRente). Met de rest van Blok 14 is weinig mis. Daarom veranderen we ook niets aan de aanpak van deze overige variabelen. Eerst worden daarom voor alle records de ontbrekende waarden op deze overige variabelen bijgeschat. Het gaat om:

- AankPrs t/m NHG (14.01 t/m 14.10) en
- Schenk t/m Verkwaar (14.27 t/m 14.03)

Dit zijn dus de variabelen die buiten de loop vallen en bovendien geen directe relatie hebben met de som van afzonderlijke Hypotheeklasten over alle loops. Deze variabelen worden, waar nodig, op dezelfde wijze met dezelfde relatietabellen bijgeschat als voorheen.

Stap 1

Dan wordt voor alle records waarvoor Blok 14 niet gevraagd is en voor records waarbij sprake is van 0 hypotheek blok 14 afgerond. Dat gebeurt eenvoudigweg door de variabelen Hypo[i].Jaar t/m BnTotBetRente (14.19 t/m 14.26) allemaal op NVT te zetten.

Stap 2

Het komt regelmatig voor dat een respondent heeft aangegeven meerdere hypotheek te hebben en vervolgens bij het invullen van de tweede loop geen geldige waarde voor hypotheeksom opgeeft. De hele tweede loop blijft dan leeg in het veldwerkbestand, terwijl deze volgens de routing wel gevuld zou moeten zijn: variabele Hypotheek heeft immers waarde 2. In plaats van de volledige tweede loop bij te schatten is er van uit gegaan dat de respondent per abuis waarde 2 heeft ingevuld en dat al zijn hypotheekgegevens al in de eerste loops vermeld zijn.

De variabele Hypotheek wordt voor deze records dus van waarde 2 teruggezet naar waarde 1. De rekenregels die hierbij gebruikt worden zijn:

1. Als sprake is van twee of meer hypotheek en bij de tweede hypotheek is de hypotheeksom niet ingevuld of heeft een extreme waarde (kleiner dan 1.000 of groter dan 2.000.000), dan terugzetten naar 1 hypotheek. (komt 1612 keer voor)
2. Als de som van hypotheeksommen (indien meerdere opgegeven) groter is dan $1.7 * \text{aankoopprijs}$ (indien bekend), dan terugzetten naar 1 hypotheek. (komt 173 keer voor)

3. Als de som van eerste twee hypotheeksommen (indien uit hetzelfde jaar) groter is dan 1.2 * aankoopprijs (indien bekend), dan terugzetten naar 1 hypotheek. (komt 243 keer voor)

Stap 3: Afhandelen records met 1 hypotheek

Vervolgens worden alle records met 1 hypotheek (Hypotheek=1) afgehandeld. Voor deze records worden de volgende stappen ondernomen:

1. Eerst alle variabelen in de 2e en 3e loop op NVT gezet.
2. Vervolgens wordt variabele Hypo[1].Jaar (14.19) bijgeschat.
3. Vervolgens wordt de hypotheeksom Hypo[1].Rente (14.12) bijgeschat voor records waarvoor deze wel opgegeven is, maar door extreme waarde (kleiner dan 1.000 of groter dan 2.000.000) toch op TOGA is gezet.
4. Vervolgens wordt bepaald voor welke records TOGA's resterend op Hypo[1].Rente. Deze personen hebben dus geen waarde opgegeven voor Hypo[1].Rente, wat betekent dat ook de rest van de eerste loop volledig ontbreekt en in zijn geheel moet worden bijgeschat.
5. Vervolgens wordt de loop volledig gemaakt voor records waarvoor Hypo[1].Rente inmiddels bekend is (ofwel plausibele waarde opgegeven ofwel extreme waarde bijgeschat in stap 3.3). Daartoe worden dus voor deze records, waar nodig, de variabelen Hypo[1].HypVorm t/m Hypo[1].LptHyp bijgeschat.
6. In stap 3.7 worden straks de resterende TOGA's op Hypo[1].Rente bijgeschat. Eerst wordt een aantal potentiële donoren voor deze bijbeschrijving uitgesloten op basis van de volgende rekenregels:
 - a. Regel 1: Volledig aflossingsvrije hypotheek en openstaande schuld (registratie) wijkt te veel af (<70% of >110%) van hypotheeksom.
 - b. Regel 2: Volledig aflossingsvrije hypotheek en jaarlijkse betaalde rente (registratie) wijkt teveel af (<90% of >110%) van opgegeven bruto maandelijkse rentelast.
 - c. Regel 3: Volledig aflossingsvrije hypotheek en opgegeven bruto maandelijkse rentelast komt niet overeen (<70% of >110%) met schatting op basis van hypotheeksom en jaarlijks rentepercentage.
 - d. Regel 4: Volledig aflossingsvrije hypotheek en opgegeven netto maandelijkse rentelast is hoger dan schatting van bruto maandelijkse rentelast op basis van hypotheeksom en jaarlijks rentepercentage.
 - e. Regel 5: Lineaire hypotheek en bruto totale maandlasten wijken teveel af (<80% of >120%) van schatting van bruto totale maandlasten. Schatting van bruto totale maandlasten wordt gemaakt door bruto rentelasten en bruto aflossingslasten apart te schatten. De schatting van bruto rentelast wordt gedaan m.b.v. rentepercentage en resterende openstaande bedrag (=hypotheeksom-afgelost tot nu toe). De (grote) schatting van bruto aflossingslast wordt gedaan m.b.v. de hypotheeksom en de totale looptijd in maanden (=sysjr+rusterende looptijd-jaar van afsluiten).
 - f. Regel 6: Lineaire hypotheek en opgegeven netto totale maandlasten zijn hoger dan schatting van bruto totale maandlasten (zie regel 3.6e)
 - g. Regel 7: Er wordt afgelost (o.b.v. hypotheekvorm), maar de openstaande schuld (registratie) wijkt sterk af (<70% of >130%) van schatting openstaande schuld. De schatting van openstaande schuld wordt gemaakt m.b.v. hypotheeksom en tot nu toe afgelost bedrag.

Door toepassing van deze rekenregels blijven van de 25.218 potentiële donoren (records met 1 hypotheek waarvoor Hypo[1].Rente bekend is er 11.979 over.

7. Vervolgens worden met behulp van de geschikt bevonden donoren (dus donoren waarvoor inmiddels de gehele eerste loop gevuld is en zoveel mogelijk intern consistent is) de variabelen Hypo[1].Rente (en in samenhang daarmee simultaan ook de variabelen Hypo[1].RenteMnt t/m Hypo[1].LpTHyp) bijgeschat.
8. Ten slotte worden voor de records met 1 hypotheek de variabelen TotBetHyp t/m BNTotBetRente bepaald. Deze staan volgens de routing bij 1 hypotheek simpelweg altijd op NVT.

Stap 4: Afhandelen records met meerdere hypotheeken

Wat nu nog resteert, zijn de records met meerdere hypotheeken waarvoor – indien nodig – de variabelen binnen de twee of drie hypotheekloops en de variabelen TotBetHyp t/m BNTotBetRente moeten worden bijgeschat. Deze records worden op de volgende manier afgehandeld:

1. Allereerst wordt een aantal records met 3 hypotheeken teruggezet naar 2 hypotheeken. Dit wordt gedaan als in de derde loop geen waarde (of een extreme waarde <1.000 / $>2.000.000$) op de hypotheeksom Hypo[3].Rente is opgegeven.
2. Vervolgens worden voor records met meerdere hypotheeken eerst de totale maandlasten en totale maandelijkse rentelasten (TotBetHyp t/m BNTotBetRente) bijgeschat. Dat gebeurt (in hoofdlijnen) volgens de volgende stappen:
 - a. Eerst worden TOGA's op TotBetRente zo goed mogelijk aangevuld met behulp van de geregistreerde jaarrente en/of met behulp van de opgegeven maandelijkse rentelasten (Hypo[i].BetRente) in de verschillende loops. Daarbij wordt zoveel mogelijk rekening gehouden met eventuele opgegeven maandelijkse totale lasten (Hypo[i].TotBetHyp) in de verschillende loops. Zo mag TotBetRente bijvoorbeeld nooit hoger worden dan de opgegeven som van Hypo[i].TotBetHyp. Verder mag TotBetRente bijvoorbeeld niet kleiner zijn dan 1 promille van de totale hypotheeksom/12.
 - b. Vervolgens wordt voor alle records met meerdere hypotheeken de verhouding tussen rentelasten en totale maandlasten TotBetRente/TotBetHyp bijgeschat (hulpvariabele VH2523). Donoren voor deze bijschatting zijn de records waarvoor zowel TotBetHyp als TotBetRente al gevuld is en de verhouding tussen deze twee variabelen plausibel (tussen 0,1 en 1) is.
 - c. De hulpvariabele VH2523 wordt vervolgens gebruikt om voor de records waarvoor TotBetHyp wel bekend is, maar TotBetrente op TOGA staat de waarde van TotBetRente te bepalen.
 - d. Daarna wordt TotBetRente bijgeschat voor de records waarvoor TotBetHyp nog niet bekend is.
 - e. Ten slotte worden de TOGA's op TotBetHyp aangevuld met behulp van TotBetRente en VH2523.

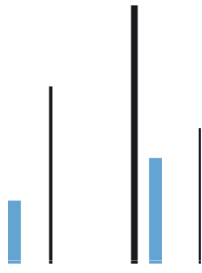
Voor alle records met meerdere hypotheeken zijn op dit moment dus bekend:

- De hypotheeksom per hypotheek (Hypo[i].Rente).
- De maandelijkse rentelasten over alle hypotheeken (TotBetRente)
- De maandelijkse totale lasten over alle hypotheeken (TotBetHyp)

3. Voor records met meerdere hypotheke(n) worden vervolgens in loop 1 de resterende TOGA's op Hypo[1].Jaar t/m Hypo[1].LpTHyp bijgeschat (exclusief Hypo[1].Rente die al bekend is en Hypo[1].BetHyp en Hypo[1].BetRente die later in 4.6 en 4.7 bepaald worden). Van Hypo[1].BetHyp en Hypo[1].BetRente worden in dit stadium alleen NVT's en TOGA's aangemaakt. De TOGA's op deze variabelen worden in stap 4.6 en 4.7 weggewerkt m.b.v. TotBetHyp en TotBetRente.
4. Voor records met meerdere hypotheke(n) worden vervolgens in loop 2 de resterende TOGA's op Hypo[2].Jaar t/m Hypo[2].LpTHyp bijgeschat (exclusief Hypo[2].Rente die al bekend is en Hypo[2].BetHyp en Hypo[2].BetRente die later in 4.6 en 4.7 bepaald worden). Van Hypo[2].BetHyp en Hypo[2].BetRente worden in dit stadium alleen NVT's en TOGA's aangemaakt. De TOGA's op deze variabelen worden in stap 4.6 en 4.7 weggewerkt m.b.v. TotBetHyp en TotBetRente.
5. Voor records met drie hypotheke(n) worden vervolgens in loop 3 de resterende TOGA's op Hypo[3].Jaar t/m Hypo[3].LpTHyp bijgeschat (exclusief Hypo[3].Rente die al bekend is en Hypo[3].BetHyp en Hypo[3].BetRente die later in 4.6 en 4.7 bepaald worden). Van Hypo[3].BetHyp en Hypo[3].BetRente worden in dit stadium alleen NVT's en TOGA's aangemaakt. De TOGA's op deze variabelen worden in stap 4.6 en 4.7 weggewerkt m.b.v. TotBetHyp en TotBetRente. Op dit moment zijn de enige variabelen waar nog TOGA's op voorkomen dus de variabelen Hypo[i].BetHyp en Hypo[i].BetRente voor de records met meerdere hypotheke(n).
6. Vervolgens worden in de 2 (of 3) loops de maandelijkse totale lasten Hypo[i].BetHyp bepaald met behulp van TotBetHyp. Daarbij wordt ervoor gezorgd dat de som van Hypo[i].BetHyp over alle loops precies optelt tot de in stap 4.2 vastgelegde waarde voor TotBetHyp. In een aantal gevallen moet daarvoor een bestaande waarde op Hypo[i].BetHyp worden overschreven. Dat gebeurt bijvoorbeeld als sprake is van 2 loops, waarbij in loop 1 een waarde op Hypo[1].BetHyp is opgegeven die onder 4.2 plausibel is geacht als zijnde de totale lasten over alle hypotheke(n) (TotBetHyp) en in loop 2 geen waarde voor Hypo[2].BetHyp is opgegeven. Het totaal TotBetHyp wordt dan over Hypo[1].BetHyp en Hypo[2].BetHyp verdeeld naar rato van de hypotheeksommen (Hypo[1].Rente en Hypo[2].Rente). Dit komt 924 keer voor.
7. Ten slotte worden in de 2 (of 3) loops de maandelijkse rentelasten Hypo[i].BetRente bepaald met behulp van de eerder vastgelegde TotBetRente. Daarbij wordt ervoor gezorgd dat de som van Hypo[i].BetRente over alle loops precies optelt tot de in stap 4.2 vastgelegde waarde voor TotBetRente. Evenals onder 4.6 zal ook hier in een aantal gevallen een bestaande waarde op Hypo[i].BetRente moeten worden overschreven. Dit 1582 komt keer voor.

Stap 5: NVT routing draaien

Ten slotte wordt op het gehele bestand nog een keer de NVT routing losgelaten. Daarmee wordt er onder meer voor gezorgd dat de variabelen TotBetHyp en TotBetRente weer op NVT komen te staan. De onderdelen (Hypo[i].BetHyp en Hypo[i].BetRente bevatten immers geen DK/RF meer).



Bijlage 4 Bijschatting huurpunten

In het WoON 2015 is het aantal huurpunten evenals in de voorgaande edities een van de vragen die aan de huurders gesteld wordt. Deze bron van informatie wordt over het algemeen niet als betrouwbaar verondersteld, omdat huurders niet altijd even goed op de hoogte zijn van het aantal punten voor de eigen woning. Daarom is aan het WoON registerinformatie gekoppeld, die afkomstig is van NCCW en de verhuurders zelf. Daarbij speelt echter het probleem dat niet voor alle cases de registerdata bekend is. Daarom is besloten om de ontbrekende data bij te schatten. In deze notitie wordt de procedure voor de bijschatting van het aantal punten voor huurwoningen beschreven.

Methode van bijschatten

Evenals in vorige edities van het WoON wordt het ontbrekende puntental in het WoON2015 bijgeschat door gebruik te maken van **regressie analyse**. Gegeven de personen waarvoor de huurpunten bekend zijn wordt bij deze techniek eerst door middel van lineaire regressie op basis van een aantal verklarende variabelen een formule opgesteld die het aantal huurpunten van deze personen zo goed mogelijk verklaart. Er moet dus worden gezocht naar variabelen in WoON2015 die sterk gecorreleerd zijn met het aantal huurpunten.

Een voorwaarde van het gebruik van deze methode is wel dat de set van gebruikte verklarende variabelen voor alle personen waarvoor het aantal huurpunten bijgeschat dient te worden, bekend dienen te zijn. Hier staat tegenover dat vaak niet veel verklarende variabelen nodig zijn om tot een zeer redelijke mate van verklaarde variantie te komen. Een ander voordeel is dat deze methode toestaat dat puntentallen geschat worden die niet in de dataset voorkomen. De spreiding van huurpunten blijft op deze manier beter gehandhaafd. Een nadeel van regressie is dat zo nu en dan extreme waarden geschat zullen worden. Als uit de regressievergelijking bijvoorbeeld blijkt dat de kale huur sterk positief gecorreleerd is met het aantal huurpunten, dan zal een ontvanger die een kale huur van € 3.000 heeft opgegeven op basis van de regressievergelijking een onrealistisch hoog aantal punten bijgeschat krijgen. Dit kan echter gemakkelijk worden voorkomen door het geschatte aantal punten op een bepaalde grens te maximeren.

Om te voorkomen dat extreme waarden toegekend worden nemen we alleen geschatte waarden over die zich tussen 40 en 300 punten bevinden. Voor ontvangers waarvoor het geschatte aantal punten buiten deze grenzen valt passen we de zogenaamde Predictive Mean Matching (**PMM**)-methode toe. Daarbij wordt voor een ontvanger het donorrecord bepaald waarvan het geschatte aantal huurpunten (op basis van de regressievergelijking) zo dicht mogelijk bij het geschatte aantal huurpunten van de ontvanger ligt. Vervolgens wordt dan het werkelijke puntental van deze donor gekopieerd naar de ontvanger.

Vooraf

Het WoON 2015 telt in totaal 62.668 respondenten waaronder 21.203 huurders (HUKO=2). Van deze huurders zijn er 1.586 die een huisvestings situatie anders dan “woning” hebben (HVS ongelijk aan 1). Deze groep van 1.586 zal worden uitgesloten van de analyse en er zullen voor deze groep ook geen huurpunten worden bijgeschat. Er blijft dan een groep van 19.617 personen over waarvoor de huurpunten ofwel al bekend zijn (zelf opgegeven of gekoppeld) of zullen worden bijgeschat.

13.574 ontvangers

Van de 19.617 personen hebben er 1.437 zelf een waarde opgegeven voor het aantal huurpunten (variabele AANTPUNT). Daarvan bevinden zich er echter 492 onder de 40 en 15 boven de 300 punten. Deze waarden worden als niet realistisch beoordeeld en daarom als ongeldig beschouwd. Voor 5.483 van de 19.617 huurders kon geldige registerdata gekoppeld worden (variabele HUURPUNT). In dit onderzoek wordt het gekoppelde aantal punten als meer betrouwbaar verondersteld dan het door de respondent zelf opgegeven aantal punten. Voor 6.043 van de 19.617 personen is (via registerdata of zelf opgegeven data) uiteindelijk een aantal huurpunten bekend dat zich tussen 40 en 300 bevindt. Voor de overige 13.574 personen zullen de huurpunten met behulp van lineaire regressie en eventueel de PMM-methode worden bijgeschat.

5.584 bruikbare waarnemingen

Zoals hierboven beschreven zijn vóór de bijschatting voor 6.043 personen (HUKO=2 & HVS=1) het aantal huurpunten bekend. Deze groep is dan ook in principe kandidaat voor de input van de regressieanalyse. We zullen echter een deel van deze groep niet gebruiken voor de analyse. Dit heeft te maken met het optreden van onrealistische combinaties van de kale huur en het aantal huurpunten, zoals hieronder wordt toegelicht.

Op basis van het aantal huurpunten kan voor de 6.043 huurders een maximale huur bepaald worden. Dit is gebeurd op basis van de huurprijstabel voor zelfstandige woningen per 1 juli 2014.

Voor de regressieanalyse zullen alleen personen worden gebruikt waarvoor de verhouding tussen de opgegeven kale huur (KHUURII) en de hierboven bepaalde maximale huur redelijk is. We hanteren daarvoor de volgende criteria:

- Personen in het gereguleerde segment (<142 huurpunten) waarvoor de kale huur groter is dan de maximale huur op basis van het aantal huurpunten, worden niet gebruikt in de regressieanalyse.
- Als de maximale huur meer dan 2,5 maal zo hoog is als de opgegeven kale huur, wordt de persoon niet gebruikt in de regressieanalyse.

Met deze criteria vallen 459 van de 6.043 personen af, waarmee er 5.584 overblijven als input voor de regressieanalyse. Let wel, voor de 459 personen die zijn afgevallen is het aantal huurpunten wel bekend en zal dus niet worden bijgeschat.

Regressie analyse

In deze paragraaf zullen de relevante aspecten van de regressieanalyse in het kort worden beschreven.

Onafhankelijke variabelen

De afwegingen bij de keuze van de te gebruiken onafhankelijke variabelen in de regressieanalyse zijn gemaakt op basis van beschikbaarheid van de variabelen, de mogelijkheden tot het eventueel bijschatten van verklaarders en zeker ook op intuïtieve juistheid van het verband met de huurpunten. Na enig experimenteren is gekozen voor de volgende verklaarders voor het aantal huurpunten:

- Een constante wordt opgenomen in de vergelijking;
- De WOZ-waarde (variabele ABWOZ);
- De woonoppervlakte (variabele BAGOPP);
- Het bouwjaar van de woning (variabele BAGBJR);
- De kale huurprijs (variabele KHUURII);

COROP-gebieden

Aangezien de relatie tussen de verklarende variabelen en het aantal huurpunten naar verwachting regionaal flink zal differentiëren, worden per afzonderlijk COROP-gebied de regressieparameters bepaald op basis van de analysegroep in het betreffende gebied. Om een minimale vulling van 15 woningen per analysegebied te realiseren, moesten een aantal COROP-gebieden worden samengevoegd. De 10 COROP-gebieden die schaarstegebieden zijn, zijn daarbij altijd gescheiden gehouden van niet-schaarstegebieden. Bovendien zijn alleen COROP-gebieden samengevoegd die geografisch in elkaars nabijheid liggen. De volgende COROP-gebieden zijn samengevoegd:

- Oost-Groningen(1) en Delfzijl en omgeving(2)
- Noord-Friesland(4) en Zuidoost-Friesland(6)
- Noord-Drenthe(7) en Zuidoost-Drenthe(8)
- Zuidwest-Drenthe(9), Noord-Overijssel(10) en Zuidwest-Overijssel(11)
- Zeeuws-Vlaanderen(31) en Overig Zeeland(32)

Na een eerste controle op de resultaten is tenslotte nog besloten om de stad Den Haag (onderdeel van COROP-gebied Agglomeratie 's-Gravenhage(26)) als apart gebied op te nemen in de analyse. Het aantal afzonderlijke gebieden waarvoor een regressieanalyse is uitgevoerd komt daarmee uit op 35.

Resultaat regressie

Per analysegebied zijn het aantal woningen dat is gebruikt voor de regressie en de resulterende R-kwadraat opgenomen in Tabel B4.2. Over alle 35 modellen wordt gemiddeld 71,1% van de variantie in het aantal huurpunten verklaard door de WOZ-waarde, het woonoppervlakte, het bouwjaar van de woning en de kale huurprijs.

Tabel B4.2 Aantal woningen en R-kwadraat per analysegebied

analysegebied	aantal woningen	R-kwadraat
Oost-Groningen en Delfzijl en omgeving	17	0,760
Overig Groningen	112	0,716
Noord-Friesland en Zuidoost-Friesland	33	0,804
Zuidwest-Friesland	17	0,627
Noord-Drenthe en Zuidoost-Drenthe	30	0,728
Zuidwest-Drenthe, Noord-Overijssel en Zuidwest-Overijssel	75	0,777
Twente	15	0,734
Veluwe	120	0,641
Achterhoek	17	0,760
Arnhem/Nijmegen	291	0,743
Zuidwest-Gelderland	21	0,817
Utrecht	366	0,717
Kop van Noord-Holland	74	0,743
Alkmaar en omgeving	22	0,799
IJmond	14	0,894
Agglomeratie Haarlem	17	0,472
Zaanstreek	32	0,831
Groot-Amsterdam	887	0,717
Het Gooi en Vechtstreek	42	0,698
Agglomeratie Leiden en Bollenstreek	97	0,738
Agglomeratie 's-Gravenhage (minus Den Haag)	757	0,669
Delft en Westland	430	0,744
Oost-Zuid-Holland	234	0,808
Groot-Rijnmond	454	0,760
Zuidoost-Zuid-Holland	64	0,702
Zeeuws-Vlaanderen en Overig Zeeland	132	0,705
West-Noord-Brabant	284	0,636
Midden-Noord-Brabant	32	0,435
Noordoost-Noord-Brabant	17	0,745
Zuidoost-Noord-Brabant	48	0,494
Noord-Limburg	47	0,599
Midden-Limburg	54	0,686
Zuid-Limburg	131	0,667
Flevoland	104	0,673
Den Haag	497	0,853

Voor elk van de 13.574 ontvangers is vervolgens het aantal huurpunten geschat met behulp van de regressievergelijking van het analysegebied van de ontvanger. Als deze waarde binnen het interval [40,300] valt, wordt de waarde rechtstreeks toegekend aan de ontvanger. Als het puntental op basis van de regressie buiten het interval [40,300] valt, dan krijgt de betreffende ontvanger het werkelijk aantal huurpunten toegewezen van een persoon uit de analysegroep van hetzelfde gebied waarvoor het geschatte puntental het dichtst bij het geschatte puntental van de ontvanger ligt (**PMM**).

Evaluatie resultaten

Na imputatie hebben alle 19.617 personen (waarvoor HUKO=2 en HVS=1) een puntental. Het betreft dus:

- 5.584 personen met punten vooraf die gebruikt zijn voor de regressieanalyse;
- 459 personen met punten vooraf die (vanwege abnormale verhouding tussen kale huur en maximale huur) niet gebruikt zijn in de regressieanalyse;
- 13.574 personen zonder punten vooraf waarvoor het puntental met behulp lineaire regressie en eventueel de PMM-methode is bijgeschat.

Punten voor en na bijschatting

Voor toetsing van de uitkomsten worden de huurpunten vóór bijschatting vergeleken met de punten na bijschatting. Uit Tabel B4.3 valt op te maken dat door de bijschatting het gemiddeld aantal huurpunten toeneemt. In WoON2015 kwam het landelijk gewogen gemiddelde aantal huurpunten vóór de bijschatting (over de 5.584 personen die gebruikt zijn in de regressie-analyse) uit op 144 en na bijschatting (over alle 19.617 huurwoningen) op 149. Groningen en Friesland zijn de enige twee provincies waarvoor het aantal huurpunten door imputatie afneemt. Ook voor de stad Rotterdam gaat het gewogen aantal huurpunten door imputatie flink omlaag.

Tabel B4.3 Aantal huurpunten voor en na bijschatting, gewogen naar huishoudens

gemiddeld aantal huurpunten vóór bijschatting				gemiddeld aantal huurpunten ná bijschatting			
Groningen	139	Amsterdam	137	Groningen	137	Amsterdam	140
Friesland	149	s-Gravenhage	124	Friesland	146	s-Gravenhage	130
Drenthe	148	Rotterdam	140	Drenthe	154	Rotterdam	130
Overijssel	147	Utrecht	153	Overijssel	149	Utrecht	156
Flevoland	155	Overige gemeenten	148	Flevoland	163	Overige gemeenten	152
Gelderland	152			Gelderland	157		
Utrecht	161			Utrecht	164		
Noord-Holland	142			Noord-Holland	145		
Zuid-Holland	136			Zuid-Holland	136		
Zeeland	145			Zeeland	147		
Noord-Brabant	159			Noord-Brabant	164		
Limburg	146			Limburg	148		
totaal	144		144	totaal	149		149

Vergelijking met CBS-huurenquête

Voor toetsing van de bijschatting worden de huurpunten na bijschatting in het WoON2015 vervolgens vergeleken met de huurpunten uit de CBS-huurenquête van 2015. Het landelijk gewogen gemiddelde voor het aantal huurpunten in de huurenquête (alleen records met PCBS>0) bedraagt 149 punten, zie Tabel 4. In het WoON2015 komt het gewogen gemiddelde puntental na bijschatting (over de 19.617 records) ook uit op 149.

Voor de provincie Groningen is het puntengemiddelde in het bijgeschatte WoON2015 8 punten lager dan in de CBS huurenquête en voor Friesland 5 punten lager. Voor de provincie Utrecht komt het puntengemiddelde in het WoON2015 juist 5 punten hoger uit dan in de CBS huurenquête. Voor de overige 9 provincies is het verschil maximaal 2 punten. Ook op het niveau van de grote steden komen het bijgeschatte WoON2015 en de CBS huurenquête goed met elkaar overeen. Voor geen van de vier grote steden is de afwijking groter dan 3 punten. Duidelijk is dat het WoON2015 ná de bijschatting beter aansluit op de CBS huurenquête dan vóór de bijschatting.

Tabel 4: vergelijking bijgeschatte WoON2015 met CBS huurenquête 2015

gemiddeld aantal huurpunten ná bijschatting				gemiddeld aantal huurpunten CBS huurenquête 2015			
Groningen	137	Amsterdam	140	Groningen	145	Amsterdam	138
Friesland	146	s-Gravenhage	130	Friesland	151	s-Gravenhage	128
Drenthe	154	Rotterdam	130	Drenthe	155	Rotterdam	128
Overijssel	149	Utrecht	156	Overijssel	150	Utrecht	153
Flevoland	163	Overige gemeenten	152	Flevoland	164	Overige gemeenten	153
Gelderland	157			Gelderland	159		
Utrecht	164			Utrecht	159		
Noord-Holland	145			Noord-Holland	147		
Zuid-Holland	136			Zuid-Holland	137		
Zeeland	147			Zeeland	148		
Noord-Brabant	164			Noord-Brabant	162		
Limburg	148			Limburg	149		
totaal	149		149	totaal	149		149