

**Dataprocessing  
WoON 2009**







## Dataprocessing WoON 2009

Uitgevoerd in opdracht van  
Het Centraal Bureau voor Statistiek

Kenneth Gopal

oktober 2009 | r2009-0046KG

**ABF RESEARCH**

VERWERSDIJK 8 | 2611 NH DELFT | T [015] 27 99 300







## Inhoudsopgave

---



<b>1</b>	<b>Inleiding</b>	<b>1</b>
<b>2</b>	<b>De dataprocessing</b>	<b>2</b>
2.1	Routingcontrole	2
2.2	Typeren open vragen	3
2.3	Consistentiecontrole	4
2.4	Imputatie	4
2.5	Bepalen gewenste woonmilieus	7
2.6	Bijkomende woonlasten	8
2.7	Continuïteit gegevens Belastingdienst	9
<b>3</b>	<b>Eindproducten</b>	<b>10</b>

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).

Naast het in elkaar schuiven van deze basisonderzoeken kent het WoON een aantal belangrijke nieuwe elementen. Zo wordt er gewerkt met specifieke modules op basis van een driejaarlijkse basissteekproef van netto 41.000 respondenten. Evenals in 1998, 2002 en 2006 hebben gemeenten of regionale samenwerkingsverbanden meegedaan aan de oversampling, waarvoor nog eens 37.000 respondenten benaderd zijn. Hiermee komt het totaal aantal respondenten dat aan het WoON heeft deelgenomen op 78.000. Zij hebben allen deelgenomen aan de basismodule, module Woningmarkt.

De gegevensverzameling heeft tijdens de veldwerkfase plaatsgevonden via telefonische, face-to-face en internet enquêtes. Het veldwerk was voor het reguliere deel van de enquête in handen van het CBS. De oversampling is voor rekening genomen door TNS NIPO.

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 ontlenuen. 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 dataprocessing uitgevoerd van respectievelijk het WBO 2002 en het WoON 2006. Voor het WoON 2009 is de dataprocessing echter in handen van het CBS. ABF is voor een aantal onderdelen ingeschakeld, te weten:

- routingcontrole;
- typeren selectie van open vragen;
- consistentiecontrole;
- imputatie;
- bepalen gewenste woonmilieu;
- bijkomende woonlasten;
- continuïteit gegevens Belastingdienst.

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 2009.

## 2 De dataprocessing

---

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

### 2.1 Routingcontrole

Niet alle vragen hoeven door elke respondent te worden beantwoord. Afhankelijk van bijvoorbeeld de leeftijd, de positie in het huishouden of de verhuiscens worden daarop afgestemde vragenblokken gesteld.

Met ruim vierhonderd vragen verspreid over 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.

Bij het WBO 2002 en het WoON 2006 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 2009 is een dergelijke controle uitgevoerd.

Hier toe is eerst de vragenlijst door de drie betrokken partijen (CBS, TNS NIPO en ABF) onafhankelijk van elkaar geautomatiseerd. Vervolgens zijn de vragenlijsten met elkaar vergeleken waarna de gevonden verschillen zijn teruggekoppeld en zo nodig gecorrigeerd, opdat alle partijen dezelfde routing in de vragenlijst hanteren.

De routingcontrole onderscheidt in feite 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 nagebouwd in de syntax van de Surfox-inputfiles. De vragenlijst is ten opzichte van het WoON 2006 op veel onderdelen ongewijzigd gelaten. Bij het programmeren van de vragenlijst is daarom de geprogrammeerde vragenlijst van het WoON 2006 als uitgangspunt genomen. Voor alle vragen is goed gecontroleerd of de routing bij de betreffende vraag inderdaad ongewijzigd gebleven is.

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 bestonden niet uit daadwerkelijk uitgevoerde enquêtes, maar uit duizenden door de computer gegenereerde enquêtes. Zowel het CBS als TNS NIPO heeft een proefbestand aangeleverd van respectievelijk 100.000 en 25.000 records. Op beide bestanden heeft ABF de controle uitgevoerd waarna ABF haar bevindingen richting het CBS en TNS NIPO gecommuniceerd heeft.

### 2.1.3 Correctie en terugkoppeling

ABF heeft een eerste check gedaan op de verschillen en deze gerapporteerd. In een klein aantal gevallen heeft ABF aangegeven welke selecties zij gehanteerd heeft, zodat CBS en TNS NIPO konden controleren op welke punten hun selecties verschilden ten opzichte van ABF.

Nadat de oplossingen voor de verschillen van de routingtest gevonden waren, is op basis van de verbeterde vragenlijst door beide veldwerkpartijen een nieuw proefbestand verstuurd, waarna ABF opnieuw haar controles uitgevoerd heeft. Dit proces is net zo lang herhaald totdat er geen verschillen meer optraden. Bij geen gevonden verschillen zijn immers de geprogrammeerde vragenlijsten gelijk aan elkaar. In totaal is het controleproces drie keer herhaald voor het CBS en voor TNS NIPO één keer.

## 2.2 Typeren open vragen

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

De typeringen zijn in principe een verlengstuk van de veldwerkfase en worden vrij vroeg in het traject opgenomen, omdat zij mogelijk in het vervolg (consistentie, imputatie) een rol kunnen spelen. Het typen van de open antwoorden is eenmalig uitgevoerd, waarbij voor elke tekstregel in een apart bestand is vastgelegd wat de bijbehorende code is – zo deze te herleiden was naar een bestaande antwoordcode. Indien een open antwoord niet eenduidig getypeerd kon worden is de code 'anders' gehandhaafd.

De volgende variabelen zijn door ABF getypeerd: (1) soort hypotheek, (2) reden om niet te verhuizen en (3) gewenste wijk/buurt in huidige gemeente en (4) gewenste wijk/buurt in andere gemeente. In Tabel 2.1 is een overzicht opgenomen voor het aantal getypeerde gevallen per variabele.



Tabel 2.1 Aantal cases per open vraag met antwoord 'anders' dat getypeerd is

Vraag	Variabele	Aantal 'anders'	Aantal getypeerd
13.11	AsrtHypo	1177	469
25.2	NietVH	4105	1654
28.3	GBuurt1	222	17
28.5	GSpecBr	117	21

De overige tekstvariabelen zijn door het CBS zelf getypeerd. Op verzoek is echter voor een zestal variabelen bij wijze van achtervang een 'extra' imputatie uitgevoerd. Deze variabelen worden in eerste instantie dus wel getypeerd, maar in het geval dit niet op een eenduidige wijze mogelijk is wordt de geïmputeerde waarde overgenomen door het CBS. Een overzicht van deze variabelen is opgenomen in Tabel 2.2.

Tabel 2.2 Variabelen die geïmputeerd zijn als achtervang voor typering

Vraag	Variabele	Imputeren/Typen naar
6.6	Anders	SoortWon, SrtWoonR, AardAdr
10.2	Omschre	CodeVerh
36.6a	NSAndOp	NVOplOp, NSOplOp
36.2	VSOplOp	VOplOp
39.6a	NSAndPa	NVOplPa, NSOplPa
39.2	VSOplPa	VOplPa

## 2.3 Consistentiecontrole

Tijdens het veldwerk wordt door het CBS en TNS NIPO 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 (voorbeeld: 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, opdat de samenhang na de correctie wél consistent is.

Het CBS heeft in samenspraak met VROM deze afweging gemaakt en vervolgens de consistentiecontroles aangeleverd. Deze zijn overwegend hetzelfde gebleven ten opzichte van de controles die in 2006 zijn uitgevoerd. ABF heeft de controles verwerkt in haar eigen syntax voor Surfox. Om te voorkomen dat tijdens de imputatie nieuwe inconsistenties geïntroduceerd worden, worden de consistentiecorrecties voorafgaand aan de imputaties uitgevoerd.

## 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 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 in het verleden eveneens betrokken geweest bij de imputaties van het WBO 1998, WBO 2002 en het WoON 2006.

## 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 brengen 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.

Dé 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 zijn opgenomen in een 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

### 2.4.2.1 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. Ondanks deze routingcontrole vooraf bleek dat door onbekende oorzaken een klein aantal verschillen bestond tussen de geprogrammeerde routing en de routing die in het bestand is gevolgd.

Verder zijn er geen grote verschillen geconstateerd tussen de geprogrammeerde routing en de routing in het bestand. In een klein aantal gevallen komen TOWA's voor (Ten Onrechte Wel Antwoord). Deze zijn op NVT (Niet Van Toepassing) gezet, zodat de routing in het analysebestand precies klopt met de voorgeschreven routing (in de vragenlijst).

### 2.4.2.2 Registerdata

Om bij de imputaties gebruik te kunnen maken van zoveel mogelijk relevante extra informatie is registerdata aan het veldwerkbestand gekoppeld. Hierbij gaat het om GBA-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' 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 niet tot onlogische bijschattingen 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 in feite afgestemd op een centrale variabele (huurprijs) waarbij gebruik gemaakt wordt van RHM. Indien eenmaal een geschikte donor gevonden is, worden alle samenhangende variabelen overgenomen.

## 2.4.4 Uitvoering imputatie

### 2.4.4.1 Volgorde van bijschatting

Het 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, vorige woning, gewenste woning, inkomen ondervraagde en inkomen partner.

Elk blok kent zijn centrale variabelen (huur- en koopprijs bij de huidige woning; netto-inkomen bij het inkomensblok). Deze vragen zijn meestal 'bedragen'. Hierin zit de meeste non-respons. De non-respons neemt toe naarmate de vragenlijst vordert. Dit heeft niet alleen te maken met afgebroken interviews maar ook met de privacygevoeligheid van de vragen.

De opzet van het WoON zorgt ervoor dat sequentieel de blokken kunnen worden afgelopen. Binnen elk blok worden eerst de vragen met weinig non-respons bijgeschat. Per blok wordt toegewerkt naar de centrale variabelen.

#### **2.4.4.2 Consistentiecontroles**

Hoewel de consistentiecontroles voor de imputatie zijn uitgevoerd en hoe goed de relatietabellen ook zijn gemaakt, het kan altijd gebeuren 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 gedraaid.

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 vrijwel altijd 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, waarbij bovendien geen inconsistenties optreden, als donor gaan dienen, worden deze records buiten de donorgroepen gelaten.

#### **2.4.4.3 Gebruikte software**

De imputaties zijn uitgevoerd met het pakket Surfox dat ABF zelf ontwikkeld heeft. Ook voorgaande WBO's en het WoON 2006 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 bijschattingen, worden bewaard in een outputfile.

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. Voor de selectie van variabelen die worden geïmputeerd wordt aangesloten bij de vragenlijst van het WoON 2006. In principe worden de vragenlijstvariabelen die in het WoON 2006 werden geïmputeerd ook nu geïmputeerd. Daarom zijn ook de relatietabellen van 2006 als uitgangspunt voor 2009 gehanteerd.

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.

## **2.5 Bepalen gewenste woonmilieus**

Het gewenste woonmilieu is een belangrijke variabele voor het Ministerie van VROM. 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. Indien de buurt niet bekend is, en ook niet te typeren valt 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 (plus of men binnen of buiten de bebouwde kom wil wonen), 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.

In de vragenlijst van het WoON 2009 is een aantal nieuwe vragen opgenomen met betrekking tot de kenmerken van de gewenste woonomgeving in relatie tot voorzieningen. Deze zijn betrokken bij de imputatie van het gewenste woonmilieu.

## 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 water, gas, warmte en elektriciteit (nutsvoorzieningen).

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 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

#### 2.6.1.1 Heffingen

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

#### 2.6.1.2 WOZ-waarden

Het CBS beschikt over de WOZ-waarde van alle adressen in Nederland. Op basis van het adres van de respondenten heeft het CBS de WOZ-waarden als registerdata gekoppeld aan het veldwerkbestand.

## 2.6.2 Waterschapsheffingen

### 2.6.2.1 Heffingen

ABF 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 per gemeente de omvang van de tarieven. Deze gegevens zijn bewerkt zodat deze gebruikt konden worden in de rekenprogrammatuur.

### 2.6.2.2 Grondslagen

De watersysteemheffing ingezetenen betreft een bedrag per woonruimte (huishouden). De watersysteemheffing gebouwd is gebaseerd op de WOZ-waarde (zie ook paragraaf 2.6.1.2). 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 in eigen beheer verzameld. De verbruikgegevens met betrekking tot energie (elektriciteit, gas en warmte) heeft ABF via het CBS verkregen.

### 2.6.3.1 Tarieven

ABF 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 houdt een overzicht bij welke waterbedrijven in welke gemeente leverancier is.

### 2.6.3.2 Verbruik

De verbruikgegevens zijn door het CBS gekoppeld, mits de respondent daar toestemming voor heeft gegeven én deze volgens het WoON in een woning woont. Als die toestemming er niet is, zijn de verbruikcijfers door het CBS geïmputeerd. SenterNovem is ingeschakeld om de door het CBS geleverde energieverbruiken met gegevens over warmtelevering via stadsverwarming of blokverwarming aan te vullen en om de gegevens op een hoger kwaliteitsniveau te brengen.

## 2.7 Continuïteit gegevens Belastingdienst

In het WBO 2002 en het WoON 2006 heeft ABF de berekening van de inkomens uitgevoerd, op basis van de gekoppelde gegevens van de Belastingdienst. Voor het WoON 2009 worden de inkomens door het CBS berekend. Om er voor te zorgen dat geen trendbreuk optreedt bij de berekening van de inkomens, is ABF gevraagd om zitting te hebben in de begeleidingscommissie.

In totaal zijn er vier bijeenkomsten geweest waarin het CBS verslag deed van de werkzaamheden en informatie omtrent de berekeningen is uitgewisseld. Helaas bleek het in de laatste maanden, waarin de definitieve resultaten bekend zijn geworden en ook een vergelijking met het WoON 2006 is uitgevoerd, niet mogelijk om de begeleidingscommissie volledig bij het traject te betrekken. ABF heeft zich dan ook geen oordeel kunnen vormen over de continuïteit van de gegevens.



## 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 wordt een extra bestand opgeleverd waarin de extra aangemaakt variabelen zijn opgenomen, zoals de typering en de woonlasten.

Daarnaast wordt 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 programma's zijn gebruikt voor het uitvoeren van de verschillende onderdelen van de dataprocessing.