Uitvoering Woningmarktmodule 2021

José Gouweleeuw

projectnummer 305492
DVO
18 maart 2020

samenvatting Op 26 maart 2020 stond er een informatiebijeenkomst gepland voor potentiële deelnemers aan de oversampling van de Woningmarktmodule 2021. Vanwege het Coronavirus is deze bijeenkomst echter afgelast. In dit document wordt beschreven wat er in deze bijeenkomst zou zijn verteld over alle zaken die samenhangen met de waarneming: de waarneemstrategie, de benaderstrategie, steekproefkaders en steekproeftrekking en dataprocessing.

trefwoorden Benaderstrategie, dataprocessing, oversampling, steekproefkader, steekproeftrekking, waarneemstrategie, woningmarktmodule

1. Inleiding

Ten behoeve van de WoON oversampling wordt er al jaren een bijeenkomst georganiseerd om potentiële belangstellenden te informeren over de gang van zaken rond deze oversampling. In 2020 stond een dergelijke bijeenkomst gepland voor 26 maart, maar vanwege het Corona virus kan deze niet doorgaan. In dit document worden de zaken beschreven die bij eerdergenoemde bijeenkomst zouden zijn verteld die samenhangen met de waarneming. Het gaat hierbij om de waarneemstrategie, de benaderstrategie, de steekproeftrekking en de dataprocessing.

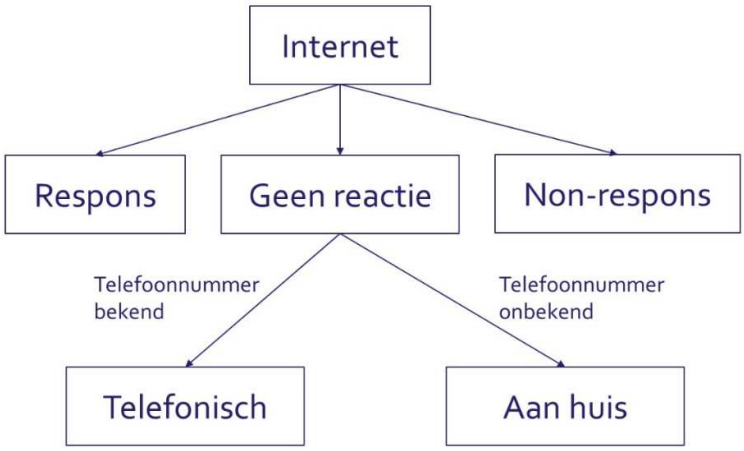
Deze notitie is als volgt opgebouwd. Paragraaf 2 gaat over de waarneem- en benaderstrategie. In paragraaf 3 wordt ingegaan op de steekproef. Hierbij is aandacht voor zowel de landelijke steekproef als de steekproefwensen van de deelnemers aan de oversampling, zowel de specifieke als de niet-specifieke wensen. De dataprocessing wordt beschreven in Paragraaf 4.

2. Waarneemstrategie en benaderstrategie

Voor de Woningmarktmodule 2021 wordt, op dezelfde manier als in de vorige editie (2018), een mixed mode strategie gebruikt. Deze is schematisch weergegeven in Figuur 2.1. Alle personen die in

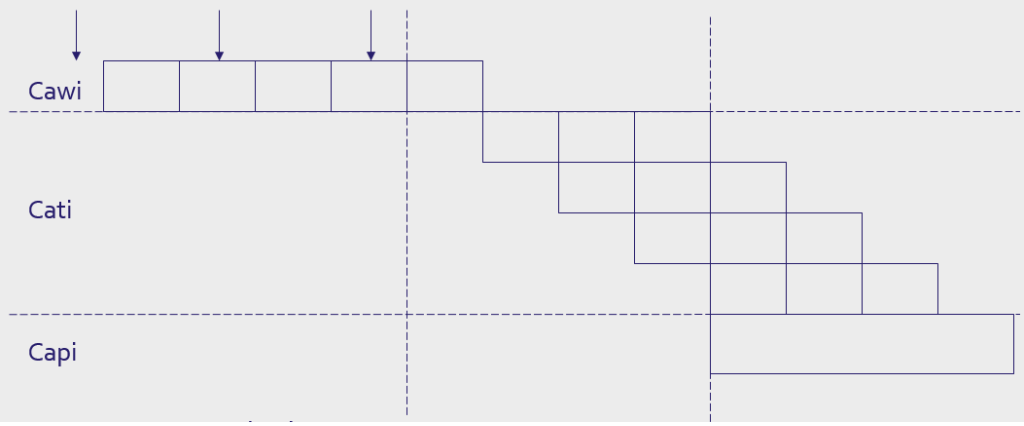
de steekproef zijn geselecteerd worden in eerste instantie uitgenodigd om de vragenlijst op internet in te vullen. Dit wordt de CAWI-waarneming genoemd. De steekproefpersonen krijgen hiervoor een brief met inloggegevens thuisgestuurd. Een deel van de aangeschreven personen zal de vragenlijst invullen, dit leidt dus tot respons. Ook zijn er personen die bij het CBS aangeven dat zij niet mee kunnen of willen doen en niet meer willen worden benaderd. Dit is non-respons. Personen die in de periode dat de vragenlijst openstaat niet reageren (dus geen vragenlijst invullen en ook niet aangeven dat zij niet mee kunnen of willen doen) gaan door naar een andere mode van waarneming. Wanneer van een steekproefpersoon een telefoonnummer beschikbaar is (van voldoende kwaliteit), dan wordt deze persoon telefonisch herbenaderd. De overige personen worden aan huis benaderd. Bij de eerste aanschrijfbrief die een steekproefpersoon ontvangt (voor de CAWI-waarneming) zit een onvoorwaardelijke incentive: een VVV-cadeaukaart van 5 euro. Dit was in 2018 ook het geval.

Figuur 2.1. Waarneemstrategie voor de Woningmarktmodule 2021.



De waarneming voor het landelijke deel van de Woningmarktmodule 2021 loopt van augustus 2020 tot en met april 2021. De totale steekproef wordt hiervoor in zeven porties verdeeld. Deze porties variëren in omvang van 8.388 personen tot 12.582 personen. De CAWI waarneming voor deze porties start vanaf augustus 2020 tot en met februari 2021. Voor één portie is in Figuur 2 schematisch aangegeven hoe de waarneming wordt uitgevoerd. Ieder blokje in deze figuur staat voor een periode van een week. Er wordt gestart met CAWI. Vlak voor de start van het veldwerk krijgen alle steekproefpersonen een brief met inloggegevens waarin wordt gevraagd om deel te nemen aan het onderzoek, na twee weken volgt er een eerste rappel, en na vier weken een tweede rappel. In totaal blijft de internet waarneming 5 weken open staan. Na afloop van de CAWI-periode wordt gekeken welke personen er niet gerespondeerd hebben en ook niet hebben aangegeven dat zij niet mee kunnen of willen doen. Wanneer van deze personen een telefoonnummer (van voldoende kwaliteit) beschikbaar is, gaan zij door naar telefonische waarneming (CATI), anders worden zij aan huis bezocht (CAPI). De personen die via CATI worden benaderd worden in 4 gelijke porties verdeeld. De eerste portie start direct na afsluiting van de CAWI-periode, de tweede een week later, de derde start twee weken na afsluiting van CAWI en de laatste portie start drie weken later. Elke portie heeft een doorlooptijd van drie weken. Een maand na het afsluiten van het CAWI veldwerk start de benadering van de personen die CAPI worden bezocht, en dit heeft een doorlooptijd van een maand. Eén en ander is schematisch weergegeven in Figuur 2.2.

Figuur 2.2. Schematische weergave van de waarneming over de tijd.



De aanschrijfbrieven voor de CAWI-waarneming worden op dinsdag verstuurd. Hiermee wordt ervoor gezorgd dat de brieven vlak voor het weekend aankomen, en dat steekproefpersonen de tijd hebben om de enquête in het weekend in te vullen. Voor de rappelbrieven geldt hetzelfde. Zoals hierboven beschreven eindigt de CAWI-waarneming na 5 weken. Vanaf dat moment is het niet meer mogelijk om de vragenlijst in te vullen via internet. Dit wordt gedaan om te voorkomen dat er dubbel wordt gerespondeerd. Het is dus niet mogelijk dat er iemand wordt gebeld of aan huis bezocht, die al heeft meegedaan. Bij de CATI waarneming wordt er per portie onderscheid gemaakt in 65plussers en 64minners. De 65plussers worden in eerste instantie overdag gebeld, de 64minners in eerste instantie 's avonds. Per persoon worden er maximaal 9 contactpogingen gedaan. Dit is altijd verspreid over 3 dagen en 3 dagdelen. Bij de CAPI waarnemingen worden er in totaal 6 contactpogingen gedaan, verspreid over de maand. Alleen wanneer er bij de 6^{e} poging een afspraak wordt gemaakt, wordt een steekproefpersoon nog een 7^{e} keer bezocht.

Wanneer de internetwaarneming wordt afgesloten, wordt er voor de nog te benaderen personen onderzocht of deze nog steeds in leven zijn. Personen waarvan in de Basisregistratie Personen (BRP) bekend is dat zij inmiddels zijn overleden worden uit de steekproef verwijderd, en dus niet gebeld of bezocht.

Tot slot moet het volgende worden opgemerkt over de landelijke Woningmarktmodule. Er ligt een wens voor het aantal respondenten voor 19 gebieden[1], en er is met behulp van responskansen uitgerekend hoe groot de CAWI-steekproef overal moet zijn om met de in deze paragraaf beschreven benaderstrategie tot de gewenste respons te komen. Om voldoende zekerheid te hebben dat de responswensen in de verschillende gebieden worden gehaald, wordt er een kleine extra steekproef uitgezet. Deze extra steekproef wordt alleen via internet waargenomen, en zorgt ervoor dat de gewenste respons per gebied met grote waarschijnlijkheid wordt gehaald.

Met deze benaderstrategie is op basis van de gerealiseerde respons uit de Woningmarktmodule in 2018 uitgerekend wat de verwachte samenstelling van de respons naar mode zal zijn in 2021. Deze is weergegeven in Tabel 2.1. De verhoudingen zijn anders dan in het WoON 2018 het geval was. Dit

[1] Groningen, Friesland, Drenthe, Overijssel minus Twente, Overijssel, Regio Arnhem/Nijmegen, Gelderland minus Regio Arnhem/Nijmegen, omgeving gemeente Utrecht, Provincie Utrecht minus omgeving Utrecht, Regio Amsterdam/Zaanstreek, Noord-Holland minus Regio Amsterdam/Zaanstreek, Regio Den Haag, Regio Rotterdam-Rijnmond, Zuid-Holland minus Den Haag en Rotterdam-Rijnmond, Zeeland, Zuidoost-Noord-Brabant, de rest van de Provincie Noord-Brabant, Limburg, Flevoland.

heeft te maken met het feit dat het steeds lastiger wordt om kwalitatief goede telefoonnummers te krijgen. Dit heeft invloed op de responskans bij CATI.

Tabel 2.1. Verwacht aandeel van de respons naar mode.

Mode	Verwacht aandeel landelijke respons
CAWI	68%
CATI	18%
CAPI	14%
Totaal	100%

3. Steekproef, landelijk en oversampling

De doelpopulatie van de Woningmarktmodule bestaat uit alle in Nederland woonachtige personen, die behoren tot particuliere huishoudens en die op 1 januari 2021 minstens 18 jaar oud zijn. Met 'in Nederland woonachtige personen' wordt op het CBS bedoeld 'alle personen die als ingezetene staan ingeschreven in de Basisregistratie Personen (BRP). Er is een lijst van al deze personen (met bijbehorende demografische kenmerken) en uit deze lijst kan een steekproef worden getrokken. In de steekpoef voor de Woningmarktmodule worden personen die partner in een huishouden zijn met een twee keer zo kleine kans getrokken ten opzichte van de overige personen. Op deze manier wordt er voor gezorgd dat alle huishoudens ongeveer dezelfde kans hebben om in de steekproef te komen. Immers, het huishouden van een gehuwd of ongehuwd stel komt in de steekproef wanneer één van de beide partners wordt geselecteerd. Het huishouden van een alleenstaande komt alleen maar in de steekproef wanneer hij/zij zelf wordt geselecteerd. Wanneer iedereen dezelfde kans zou hebben om in de steekproef te komen, zou het huishouden van de alleenstaande een twee keer zo kleine kans hebben om in de steekproef te komen. Door partners in een huishouden nu met een twee keer zo kleine kans te trekken, wordt dit als het ware rechtgetrokken.

Het CBS probeert de enquêtedruk op de bevolking zoveel mogelijk te spreiden. Wanneer mensen namelijk vaker worden gevraagd om met een onderzoek mee te doen, zullen zij op een bepaald moment steeds minder geneigd zijn om te responderen. Daar waar mogelijk probeert het CBS ervoor te zorgen dat personen binnen een bepaalde periode niet te vaak worden benaderd. Dit komt de respons over het algemeen ten goede. Er is een systeem voor steekproeftrekking opgezet, dat hiervoor kan zorgen. Een belangrijke randvoorwaarde hierbij is dat de steekproeven niet te groot mogen zijn. Daarom mag de steekproef voor de oversampling van de Woningmarktmodule niet groter zijn dan 14% van de doelpopulatie.

Voor de oversampling van de Woningmarktmodule kan een aanvrager een specifieke of en niet-specifieke aanvraag hebben. Een niet-specifieke aanvraag is eenvoudig: er wordt dan alleen aangegeven dat een bepaald aantal respondenten in een bepaald gebied (bijvoorbeeld een gemeente) gewenst is. Bij een specifieke aanvraag wordt er hierbinnen ook nog een uitsplitsing gemaakt naar verschillende onderdelen van het betreffende gebied. Denk hierbij bijvoorbeeld aan een gewenst aantal respondenten in de verschillende wijken van de gemeente. Of de gemeente wordt op basis van wijken en buurten opgedeeld in Noord, Oost, Zuid en West, en voor elk van deze deelgebieden is er een gewenst aantal respondenten. Hierbij is het wel van belang dat de indeling in deelgebieden op de één

of andere manier in het steekproefkader kan worden ingebouwd. Dit houdt in dat iedere persoon aan een uniek deelgebied moet kunnen worden toegewezen. Dit kan bijvoorbeeld op basis van wijken, maar ook op basis van postcode. Ook is het van belang dat de gewenste respons per deelgebied wel enige omvang heeft, om betrouwbare uitspraken over dit gebied te kunnen doen.

Voor de oversampling wordt per gebied waarvoor een steekproef is aangevraagd (dit kan de hele gemeente zijn, maar dus ook een deelgebied), een aselecte steekproef getrokken, waarbij partners in het huishouden dus een twee keer zo kleine kans hebben als de overige personen.

Tot slot moet er per gebied worden aangegeven hoe groot het gewenste respons-aantal in ieder gebied is. Het CBS zal dan met behulp van responskansen in die gebieden uitrekenen hoe groot de steekproef moet zijn. Hierbij wordt ervoor gezorgd dat dit gewenste aantal respondenten met grote zekerheid wordt gehaald. Heel precies wiskundig geformuleerd betekent dit dat de linkergrens van het betrouwbaarheidsinterval samenvalt met het gewenste aantal respondenten. Dit wordt op dezelfde manier gedaan als bij de landelijke steekproef. Er wordt eerst uitgerekend met behulp van responskansen in het betreffende gebied hoe groot de steekproef moet zijn om op de gewenste aantallen uit te komen. Vervolgens wordt de steekproef een stukje vergroot om met grotere zekerheid op dit aantal uit te komen. Dit extra stukje steekproef wordt alleen via internet waargenomen.

4. Dataprocessing

Wanneer de waarneming is afgelopen vindt de dataprocessing plaats. Allereerst moeten alle waarnemingen worden samengevoegd tot één bestand, met één datamodel. Dit betekent dat de waarnemingen die zijn gedaan in de verschillende modes (CAWI, CATI en CAPI) allemaal geüniformeerd moeten worden.

Een tweede stap is het toevoegen van informatie uit registers. Het gaat hierbij om dezelfde registers als in de vorige editie:

- Demografische gegevens worden toegevoegd uit de Basisregistratie personen (BRP)
- WOZ-waarde, inkomensgegevens en gegevens over huurtoeslag worden toegevoegd uit registers van de Belastingdienst
- Energieverbruik wordt toegevoegd uit registers van de netbedrijven.
- Huurpunten worden toegevoegd, hiervoor zijn diverse bronnen.
- Het bouwjaar en de oppervlakte van de woning word toegevoegd uit de Basisregistratie Adressen en Gebouwen (BAG)

Na het toevoegen van de registerinformatie moeten er ook controles en consistentie checks worden uitgevoerd. Hiervoor wordt de registerinformatie ook gebruikt. Een voorbeeld van een dergelijke check: de datum waarop een respondent aangeeft op het adres te zijn komen wonen, moet na de datum van immigratie liggen.

Verder zitten er in de vragenlijst vragen met open antwoorden, deze moeten worden getypeerd om statistiek te kunnen maken. Bij de vraag naar type woning, is er de mogelijkheid om ‘anders’ aan te vinken en dan zelf te omschrijven wat voor woning het is. Dit moet getypeerd worden. Ook worden er afleidingen en classificaties gemaakt van de variabelen in het ruwe databestand, en worden er

doelvariabelen afgeleid uit één of meerdere variabelen. Een voorbeeld hiervan is de variabele VORM-EIG. Deze combineert de vorm (een- of meergezinswoning) en het eigendom van de woning van de respondent (huur- versus koopwoning) in één variabele.

Het is mogelijk dat een respondent een bepaalde vraag heeft overgeslagen. In dat geval wordt het antwoord geïmputeerd. Deze imputatie is een zogenaamde 'random hot deck imputatie'. Dit betekent dat er wordt gekeken naar een aantal respondenten die 'lijken' op de respondent met het ontbrekende antwoord, en die de betreffende vraag wel hebben ingevuld. Uit deze respondenten wordt er eentje geloot en het antwoord van deze respondent wordt overgenomen.

De laatste stap die moet worden uitgevoerd is de weging. Hierbij krijgt iedere respondent een gewicht om op basis van de respondenten uitspraken te kunnen doen over alle inwoners van Nederland van 18 jaar of ouder in particuliere huishouden. Het gewicht houdt rekening met het steekproefontwerp en met selectiviteit in de respons. Er moet rekening worden gehouden met het steekproefontwerp omdat niet iedereen dezelfde kans heeft gehad om in de steekproef te komen. Partners hebben een twee keer zo kleine kans, en verder is de kans om in de steekproef te komen in bijvoorbeeld Flevoland groter dan de kans om in de steekproef te komen in Limburg. Hier wordt in de weging voor gecorrigeerd. Verder is er sprake van non-respons, en deze is niet aselect. Er wordt geanalyseerd welke groepen er over- en ondervertegenwoordigd zijn in de respons, en hier wordt in de weging voor gecorrigeerd.

Tot slot wordt de plausibiliteit gecontroleerd. Er wordt gekeken naar de trends, en hoe deze zich verhouden tot de vorige onderzoeken, en de uitkomsten van het onderzoek worden vergeleken met registerdata. Wanneer deze plausibiliteitscheck is afgerond, kan het bestand worden opgeleverd.