



Weegrapport

Energiemodule ENO2018

Arthur Kierkels, Ralph Meijers

maart 2019

samenvatting De respons op de Energiemodule, bestaande uit een bewonersenquête en een woningopname, is gewogen zodat op basis hiervan uitspraken kunnen worden gedaan over de populatie. De weging corrigeert voor selectiviteit in respons, en voor over- en ondervertegenwoordiging in de uitgezette steekproef. Eerder is voor de bewonersenquête apart een weegmodel opgesteld, waarbij rekening is gehouden met een verwachte uitval van maximaal een derde van enquêterespons bij technische inspectie van de woning. De uitval is echter groter en selectiever gebleken dan vooraf aangenomen, waarom het weegmodel aangepast is moeten worden. In navolging van de weging van WoON2018 betreft het een weging op persoonsniveau waarbij totalen van enkele aan de woning gerelateerde kenmerken op huishoudensniveau passend worden gemaakt.

Inleiding

In deze notitie wordt de weging voor de Energiemodule (ENO) 2018 beschreven. Dit onderzoek is een vervolg op de module woningmarkt van het Woon Onderzoek Nederland (WoON) 2018. De opdrachtgever van beide onderzoeken is het ministerie van Binnenlandse Zaken en Koninkrijksrelaties (BZK). De ENO2018 bestaat uit twee delen: een bewonersenquête waarin gegevens worden verzameld over gedrag van respondenten met effect op energieverbruik, en een daaropvolgende technische inspectie van de woning. Omdat de woningopnames meer tijd kostten dan gepland, is de respons op alleen de enquête op verzoek van BZK al eerder apart gewogen (Kierkels en Meijers, Weegrapport Bewonersenquête ENO2018, januari 2019).

Eén streven bij het ontwikkelen van de weging van de bewonersenquête, was dat het weegmodel ook op de respons van de gehele Energiemodule toegepast kon worden. De verwachting was dat bij minimaal 5.000 van de 7.426 responsen op de bewonersenquête er ook een technische inspectie van de woning zou volgen, i.e. een uitval van maximaal een derde van de beschikbare respons. Het opgestelde model kende een minimaal aantal van 17 responsen per weegcel, en bij aselechte 'nonrespons' op de woningopnames werden geen grote problemen verwacht. Helaas bleek de uitval groter en minder aselekt dan gehoopt (zie **Tabel 1** voor responspercentages per stratum); het uiteindelijke bestand bevatte 4.506 responsen. Om die reden is nu voor een minder uitgebreid weegmodel gekozen. Tevens is het niet langer verantwoord om stratumtotalen passend te maken op persoons- én huishoudensniveau, waarom ook de wijze van bepalen van startgewichten is aangepast.

Stratificatie

Bij de Energiemodule 2018 worden 30 strata onderkend, gedefinieerd middels de volledige uitkruising van vijf bouwjaarklassen (1930 of eerder, 1931-1959, 1960-1980, 1981-1995, en na 1995), drie eigendomtypes (koop, en sociale of particuliere huur), en twee woningvormen (ééngezins of meergezins).

Tabel 1: Responspercentages ENO2018 naar stratum; bewonersenquête geperceerd op uitzet, woningopname geperceerd op enquêterespons, en (onder de streep) woningopname geperceerd op uitzet

eigendom	vorm	bouwjaar				
		voor 1931	1931-'59	1960-'80	1981-'95	na 1995
koopwoning	eengezins	36,1%	38,4%	39,4%	39,1%	35,7%
		64,8%	62,9%	63,8%	67,4%	61,5%
	meergezins	23,4%	24,1%	25,1%	26,4%	22,0%
		36,1%	33,4%	32,1%	36,4%	40,3%
		70,2%	60,6%	61,3%	68,9%	67,3%
		25,3%	20,3%	19,7%	25,1%	27,2%
sociale huurwoning	eengezins	34,3%	36,1%	37,7%	39,0%	28,6%
		55,4%	59,9%	56,9%	56,1%	55,6%
	meergezins	19,0%	21,6%	21,4%	21,9%	15,9%
		36,8%	36,8%	39,7%	38,0%	38,3%
		52,8%	46,6%	57,5%	56,2%	62,8%
		19,4%	17,1%	22,9%	21,3%	24,0%
particuliere huurwoning	eengezins	27,3%	25,3%	32,8%	32,6%	34,0%
		39,4%	27,5%	47,1%	59,7%	49,1%
	meergezins	10,8%	7,0%	15,5%	19,5%	16,7%
		27,9%	28,7%	32,8%	35,6%	37,8%
		50,4%	47,3%	59,5%	57,4%	52,9%
		14,1%	13,6%	19,5%	20,4%	20,0%

Doelpopulatie en steekproefkader

De doelpopulatie bestaat uit de Nederlandse woningvoorraad, voor zover het zelfstandige woningen betreft, en de bewoners daarvan. Steekproefkader was een selectie van respondenten op WoON2018 die hebben aangegeven voor vergelijkbaar onderzoek opnieuw benaderd te mogen worden; geselecteerd hieruit zijn die respondenten die zelfstandige woningen bewoonden en die behoorden tot de huishoudkern, *volgens de op het moment van maken van die selectie beschikbare gegevens*. Niet tot het kader behoorden respondenten van wie huisvestingssituatie, plaats in het huishouden, of stratum van de woning niet bekend was, alsook respondenten die sinds deelname aan WoON verhuisd zijn; in dat laatste geval zouden gegevens verkregen uit ENO betrekking hebben op een andere woning dan waarover in WoON is geënquêteerd.

Randtotalen voor weging

De randtotalen voor ENO kunnen niet direct uit registers bepaald worden, en zijn bepaald door optelling van eindgewichten uit de weging van WoON; daarbij zijn alleen eindgewichten meegenomen van die WoON-respondenten die *volgens de op het moment van weging van WoON beschikbare gegevens* een zelfstandige woning bewoonden, die behoorden tot de kern van hun huishouden, en voor wie het stratum van de woning kon worden vastgesteld. Op deze manier zijn de randtotalen feitelijk puntschattingen op basis van de gewogen WoON-respons, welke niet per se gehele getallen zijn. Als randtotalen op huishoudensniveau dienen vergelijkbaar de puntschattingen op huishoudensniveau, welke tot stand komen door optelling van WoON-eindgewichten vermenigvuldigd met een kwantitatieve huishoudfactor; deze factor is 1 is voor elementen in een éénpersoons huishoudkern, en ½ voor elementen in een meerpersoons huishoudkern. Na weging kunnen op basis van ENO uitspraken worden gedaan over een doelpopulatie van afgerond 7.471.018 huishoudens in zelfstandige woningen en 11.651.122 huishoudkernleden daarin.

Verrijking van steekproef- en weegkader

Zowel bij de steekproeftrekking voor, alsook bij de weging van de respons op ENO is gebruikgemaakt van achtergrondkenmerken. Peildatum is voor alle variabelen dezelfde geweest als bij WoON; i.e. in de meeste gevallen 1 januari 2018, en anders het meest recente beschikbare moment vóór die datum

(e.g. voor huishoudinkomen). Het steekproefkader bestond uit die elementen in het weegkader die bij WoON-respons hebben aangegeven voor vergelijkbaar onderzoek opnieuw benaderd te mogen worden, en de ENO-respons is dus een deelverzameling van WoON-respondenten. Met het oog op vergelijkbaarheid met WoON zijn ENO-respons en weegkader voorafgaand aan de weging niet verrijkt met extra variabelen, en is uitsluitend gebruikgemaakt van kenmerken die bij de weging van WoON beschikbaar waren.

De steekproeftrekking voor ENO heeft eerder plaatsgevonden dan de weging van WoON, waardoor er bij de weging van WoON meer actuele registers beschikbaar waren dan op het moment van afbakening van het steekproefkader. Het kan daarom voorkomen dat variabelen voor enkele elementen in de steekproef een andere waarde hebben dan in het weegkader. Voor vrijwel alle weegvariabelen zijn in zowel het weegkader alsook in de ENO-respons die waardes gehanteerd die bij de weging van WoON zijn gebruikt. Uitzondering hierop zijn waardes van *stratum*, evenals die van de stratumdefiniërende variabelen *bouwjaar*, *eigendom* en *vorm*: in het weegkader wordt het stratum afgeleid uit de definiërende variabelen, en voor de respons worden juist die variabelen uit het op het moment van steekproeftrekking vastgestelde stratum afgeleid.

Nonrespons

Eén doel van weging is het corrigeren voor selectiviteit door nonrespons. Hiertoe wordt gezocht naar personeigen achtergrondkenmerken die het responsgedrag zouden kunnen verklaren. Doorgaans is herkomst sterk verklarend, waarbij personen met een niet-Westerse migratieachtergrond minder vaak responderen, en is het zo dat vrouwen iets beter responderen dan mannen. Bij de bewonersenquête was een afwijkend patroon zichtbaar: mannen respondeerden significant vaker dan vrouwen, en niet-Westerse personen juist niet significant minder vaak. Als respons op de gehele ENO2018 tellen echter alleen die responsen die zijn gevolgd door een technische inspectie van de woning. Gepercentreerd op de enquêterespons was er wel een significant lagere deelname aan die inspecties onder niet-Westersen, waardoor de totale respons op ENO2018 toch niet alleen onder mannen significant hoger was dan onder vrouwen, maar ook onder niet-Westersen significant lager dan onder Nederlanders en andere Westersen (zie **Tabel 2**). Ook leeftijd is vaak een goede indicatie voor responsbereidheid. Bij zowel de enquête, alsook bij de woningopnames werd vaker deelgenomen door ouderen. Ditzelfde is bijgevolg dan ook zichtbaar in de totale respons (zie **Tabel 3**); het verschil tussen mannen en vrouwen is ontstaan in de eerste fase (enquête). NB: in beide **Tabellen 2 en 3** betreft het ongewogen cijfers, en is tussen haakjes steeds het aantal aangeschrevene doelpersonen per cel gegeven.

Tabel 2: Afgeronde woningopnames naar herkomst en geslacht, gepercentreerd op uitzet ENO2018

	Nederlanders		Westersen		niet-Westersen		totaal	
mannen	25,9%	(8.130)	23,3%	(858)	17,4%	(536)	25,2%	(9.524)
vrouwen	19,9%	(8.989)	20,0%	(1.036)	18,0%	(646)	19,8%	(10.671)
totaal	22,7%	(17.119)	21,5%	(1.894)	17,7%	(1.182)	22,3%	(20.195)

Tabel 3: Afgeronde woningopnames naar leeftijd en geslacht, gepercentreerd op uitzet ENO2018

leeftijd	18-24	25-34	35-44	45-54	55-64	65-74	75+	totaal
mannen	11,0%	16,6%	18,7%	18,4%	27,4%	38,3%	35,8%	25,2%
	(263)	(1.364)	(1.414)	(1.807)	(2.041)	(1.777)	(858)	(9.524)
vrouwen	7,5%	12,7%	14,6%	15,4%	24,1%	29,6%	25,7%	19,8%
	(346)	(1.668)	(1.606)	(2.052)	(2.066)	(1.762)	(1.171)	(10.671)
totaal	9,0%	14,4%	16,5%	16,8%	25,7%	34,0%	30,0%	22,3%
	(609)	(3.032)	(3.020)	(3.859)	(4.107)	(3.539)	(2.029)	(20.195)

Weegmethode

Alle wegen t.b.v. ENO2018 zijn uitgevoerd met het R-package *weeg*, een uitbreiding op *survey*; hierbij is de *weeg*-implementatie van de lineaire regressieschatter gebruikt.

Startgewichten

Van ieder element in de ENO-respons is een eindgewicht uit de weging van WoON bekend. Uitgaand hiervan zijn de ENO-startgewichten in twee stappen bepaald. Allereerst zijn de WoON-eindgewichten gedeeld door empirisch vastgestelde correctiefactoren per stratum, zijnde de genormaliseerde quotiënten van de aantal elementen per stratum in de ENO-uitzet en in het *weegkader* (zie **Tabel 4**). Daarna is een weging uitgevoerd met het model *geslacht*(2) x *leeftijd*(14) + *herkomst*(3) + *provincie*(16) om tot de startgewichten te komen. Hiermee is daarin al gecorrigeerd voor de grootste vertekening door nonrespons. Zie **Tabel 5** voor de statistische gegevens van gewichten op persoons- en huishoudensniveau in deze fasen van het weegproces, en de bovenste panelen in **Figuur 1** voor de verdelingen.

Tabel 4: Eerste correctiefactoren (per stratum) bij de bepaling van startgewichten ENO2018.

Empirisch vastgestelde steekproef fracties, zodanig genormaliseerd dat de waarde bij oververtegenwoordiging groter is dan één, en bij ondervertegenwoordiging kleiner dan één.

eigendom	vorm	bouwjaar				
		voor 1931	1931-'59	1960-'80	1981-'95	na 1995
koopwoning	eengezins	0,910	0,807	0,817	0,832	0,908
	meergezins	1,574	1,547	1,045	1,547	0,870
sociale huurwoning	eengezins	1,801	1,133	0,952	1,009	1,266
	meergezins	1,962	1,470	1,100	1,175	1,247
particuliere huurwoning	eengezins	1,469	1,507	1,292	1,279	1,310
	meergezins	1,427	1,385	1,266	1,241	1,082

Weegmodel

Het voor de Energiemodule ENO2018 gehanteerde weegmodel is het volgende:

$$\text{Stratum}(30) + \text{bouwjaar}(5) \times \text{eigendom}(3) + \text{eigendom}(3) \times \text{vorm}(2) + \text{vorm}(2) \times \text{bouwjaar}(5) + \text{geslacht}(2) \times \text{leeftijd}(14) + \text{herkomst}(3) + \text{inkomen}(5) + \text{WOZ}(8) + \text{WOZ}(8) + \text{provincie}^+(16) + \text{provincie}^+(16) + \text{stedelijkheid}(5) + \text{huishoudtype}(5) + \text{huishoudomvang}(5)$$

Het getal tussen haakjes geeft steeds het aantal categorieën per variabele. Onderstreepte termen zijn passend gemaakt op huishoudensniveau; de overige op persoonsniveau. De variabelen zijn beschreven in de volgende paragraaf.

Nota Bene. Doordat de kwantitatieve huishoudfactor precies twee waarden aanneemt, gedraagt deze zich als een categoriale wanneer een (uitkruising van) categoriale variabele(n) op persoons- én huishoudensniveau is opgenomen, i.e. de som *variabele*(n) + *variabele*(n) bewerkstelligt hetzelfde effect als de uitkruising *huishoudfactor*(2) x *variabele*(n). Bovenstaand model heeft een minimale celvulling van 11. Passend maken van *Stratum* op beide niveaus zou de kleinste cel (*Stratum* 11) effectief splitsen in één van omvang 6 en één van omvang 5.

Weegvariabelen

Hier volgt een opsomming van de in het weegmodel gebruikte variabelen:

- ____ de kwantitatieve huishoudfactor uit WoON2018:
 - o 1 wanneer respondent tot een éénpersoons huishoudkern behoort; of
 - o ½ wanneer respondent tot een meerpersoons huishoudkern behoort (partner).

- *Stratum*, de uitkruising tot 30 strata van volgende variabelen:
 - o *bouwjaar* van de woning, ingedikt in vijven als:
 - 1930 of eerder, 1931-1959, 1960-1980, 1981-1995, of na 1995;
 - o *eigendom*: koopwoning, sociale huurwoning, of particuliere huurwoning;
 - o *vorm*: ééngezinswoning of meergezinswoning.
- *geslacht* van respondent: man of vrouw;
- *herkomst* van respondent: Nederlander, of Westerse of niet-Westerse migratieachtergrond;
- *leeftijd* van respondent, ingedikt in veertien als: 18- tot en met 24-jarigen, 25- tot en met 84-jarigen in vijfjaarsklassen, en 85-plussers. Voor zowel weegkader als voor ENO-respons is de leeftijd op moment van WoON-enquêteering gebruikt. Bij de WoON-weging is de respons opgehoogd naar de leeftijd op het peilmoment;
- *inkomen*: het besteedbaar huishoudinkomen volgens het voorlopige inkomensbestand 2017, ingedeeld in kwintielen. Onbekende inkomens zijn in het weegproces van WoON2018 geïmputeerd;
- *WOZ-waarde* van de woning, ingedikt in achten (7 + 1) als: 1-150.000€, 150.001-200.000€, 200.001-250.000€, 250.001-300.000€, 300.001-400.000€, 400.001-500.000€, ≥500.001€, of onbekend of "0€", als in voorgaande edities van de energiemodule;
- *provincie**: regionale indeling in zestien, te weten naar provincie, of naar gemeente indien één der vier grootste gemeenten (Amsterdam, Rotterdam, Den Haag, Utrecht);
- *stedelijkheid* van de gemeente: zeer sterk, sterk, matig, weinig, of niet stedelijk;
- *huishoudtype*: éénpersoonshuishouden, paar zonder kinderen, paar met kinderen, één-oudergezin, of ander type huishouden;
- *huishoudomvang*: aantal personen in huishouden tot vijf of meer.

Resultaat

Weging van de respons op ENO2018 leidt tot eindgewichten tussen 11,3 en 20.661,0 met een gemiddelde van 2.585,7; zie **Tabel 5** voor statistische eigenschappen hiervan, en het laatste paneel in **Figuur 1** voor een histogram.

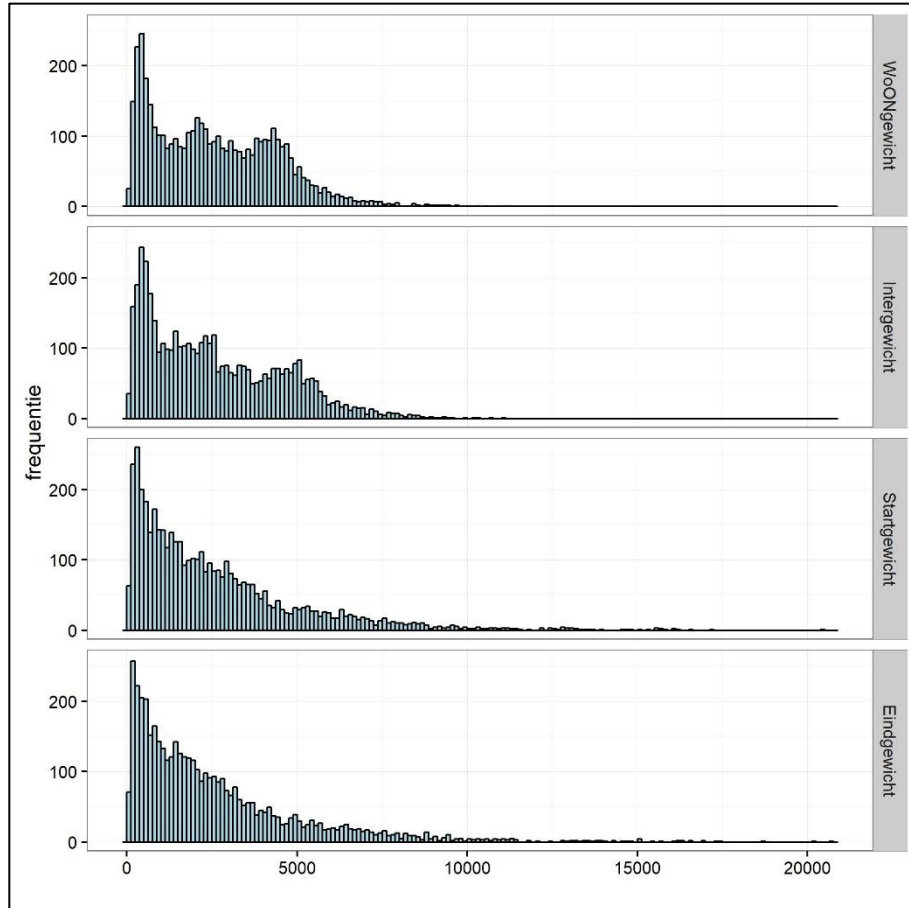
Tabel 5: Statistische gegevens van gewichten in alle fases van het weegproces voor ENO2018

niveau		μ	σ	min.	kwart. 1	mediaan	kwart. 3	max.
WoON-gewichten	pers.	2.585,7	1.830,2	7,3	920,7	2.339,6	3.977,6	11.205,5
	(geschaald) huish.	1.685,0	1.153,9	7,3	627,9	1.759,7	2.374,2	10.568,4
'tussengewichten'	pers.	2.585,7	1.948,1	8,7	834,6	2.206,9	4.057,1	11.075,4
	(correctie f) huish.	1.638,6	1.114,7	8,7	587,1	1.647,9	2.443,4	9.164,6
startgewichten	pers.	2.585,7	2.465,9	5,9	779,4	1.893,1	3.543,8	20.427,0
	ENO2018 huish.	1.655,4	1.518,4	5,9	525,4	1.298,9	2.269,4	15.696,9
eindgewichten	pers.	2.585,7	2.556,7	11,3	761,3	1.839,7	3.470,3	20.661,0
	ENO2018 huish.	1.658,0	1.569,5	11,3	523,7	1.238,2	2.280,9	18.651,5

Effect op marges

Door te wegen wordt in het algemeen de vertekening van schatters verkleind. Ook de variantie van schatters kan worden beïnvloed door gebruik te maken van ongelijke gewichten. De variantie van de regressieschatter voor doelvariabelen verandert ongeveer met een factor $(1 + CV_w^2) (1 - R^2)$, met R de correlatiecoëfficiënt tussen de betreffende doelvariabele en de hulpvariabelen van het weegmodel, en CV_w de variatiecoëfficiënt van de eindgewichten (i.e. het quotiënt van standaardafwijking en gemiddelde van de gewichten). Voor de Energiemodule ENO2018 geldt $CV_w = 0,989$ zodat in het geval dat doel- en hulpvariabelen niet correleren ($R = 0$) de marges op de regressieschatter toenemen met

een factor 1,406 (= wortel($1 + CV_w^2$)). Wanneer de absolute waarde van de correlatiecoëfficiënt R groter dan 0,703 is, dan volgt $(1 + CV_w^2) (1 - R^2) \leq 1$ en daarmee een verkleining van de marge op de schatting van de betreffende doelvariabele.



Figuur 1: Verdelingen van gewichten in verschillende fases van het weegproces. Achtereenvolgens de geschaalde WoON-gewichten, gewichten na correctie voor verschillende insluitkansen ('Intergewicht'), de startgewichten na 'voorweging', en de uiteindelijke eindgewichten.