

DE INVLOED VAN NON-RESPONS BIJ EEN BUDGETONDERZOEK OP EEN CONSUMENTEN
PRIJSINDEXCIJFER: EEN GEVOELIGHEIDSANALYSE

H.M.P. Kersten^{*)}

Samenvatting

In dit artikel wordt ingegaan op de mogelijke invloed die non-respons bij een budgetonderzoek zou kunnen hebben op schattingen van budgetaandelen. Deze budgetaandelen zijn belangrijke bouwstenen bij de constructie van consumenten prijsindices. Nagegaan is hoe gevoelig een prijsindex zou kunnen zijn voor de doorwerking van de non-respons bij een budgetonderzoek. Deze gevoeligheidsanalyse is ook uitgevoerd onder een veronderstelling over een mogelijke ondervertegenwoordiging van hogere inkomens in de respons van het budgetonderzoek. De gevoeligheid van de door ons beschouwde prijsindex blijkt gering te zijn.

*) Hoofdafdeling Statistische Methoden,
Centraal Bureau voor de Statistiek,
Postbus 959, 2270 AZ VOORBURG.
Tel. 070-694341

De in dit artikel weergegeven opvattingen zijn die van de auteur en komen niet noodzakelijk overeen met het beleid van het Centraal Bureau voor de Statistiek. De auteur dankt de heer Kortbeek voor zijn hulp bij de berekeningen en de heer Balk voor zijn opmerkingen op een eerdere versie.

1. Inleiding

Non-respons is een verschijnsel waarvan bekend is dat het invloed kan hebben op de uitkomsten van steekproefonderzoek. Indien het optreden van non-respons samenhangt met variabelen die van belang zijn voor het onderzoek leidt non-respons tot verkeerde uitkomsten. Correctieprocedures voor non-respons hebben als voornaamste doelstelling de vertekening ten gevolge van non-respons terug te dringen.

In dit artikel wordt ingegaan op de invloed van de non-respons bij een budgetonderzoek op de geschatte budgetaandelen. De vertekende werking van deze non-respons blijft niet beperkt tot de budgetaandelen. De geschatte budgetaandelen worden in het algemeen gebruikt voor de berekening van een consumenten prijsindexcijfer en een vertekening in een geschat budgetaandeel wordt meegenomen in het indexcijfer. In dit artikel wordt vooral ingegaan op de vraag hoe groot onder bepaalde veronderstellingen, de vertekening ten gevolge van non-respons in een prijsindexcijfer in het meest ongunstige geval, zou kunnen zijn. We beperken ons tot volledige non-respons d.w.z. participatie aan het hele onderzoek blijft achterwege. Partiele non-respons, d.w.z. van personen wordt medewerking aan slechts een deel van het onderzoek verkregen, wordt buiten beschouwing gelaten. Hoewel in dat geval een deel van de gewenste informatie beschikbaar is en daardoor beter gecorrigeerd kan worden voor de nog ontbrekende informatie, zijn we in dit artikel vooral geïnteresseerd in de doorwerking van de totale non-respons in de uitkomsten en houden we dit verschijnsel buiten de gevoeligheidsanalyse.

In paragraaf 2 wordt het verschijnsel non-respons besproken en wordt ingegaan op de vertekende werking van non-respons. Hierbij komt de non-respons van een budgetonderzoek aan de orde. In paragraaf 3 wordt nagegaan hoe de vertekening in een geschat budgetaandeel kan doorwerken in een prijsindexcijfer. In paragraaf 4 worden gegevens over verbruikselasticiteiten gebruikt om de doorwerking van een mogelijke ondervertegenwoordiging door non-respons van de hogere inkomens in een budgetonderzoek te schatten. In paragraaf 5 tenslotte wordt ingegaan op de mogelijke doorwerking van non-respons op een verandering in een prijsindexcijfer.

2. De vertekening ten gevolge van non-respons

Uitgangspunt bij de berekeningen is het zogenaamde Vaste Responsmodel. Dit model veronderstelt dat de populatie in twee strata kan worden verdeeld: een stratum van respondenten en een stratum van non-respondenten. Personen die tot het responsstratum behoren zullen volgens het model altijd deelnemen aan het steekproefonderzoek wanneer ze in de steekproef zijn getrokken. Personen die tot het non-responsstratum behoren zullen dat nooit doen.

De formule voor de vertekening in de schatter voor het gemiddelde in een populatie op grond van de respons, is volgens het Vaste Responsmodel eenvoudig. Voor het bepalen van de vertekening ten gevolge van non-respons in andere schatters, zoals bijvoorbeeld schatters voor quotiënten, zijn de formules ingewikkelder. Een budgetaandeel wordt geschat met behulp van een zogenaamde gecombineerde quotiëntenschatter en de formule voor de vertekening is daardoor complex. Vóórdat ingegaan wordt op de mogelijke vertekening in de schatter voor een budgetaandeel, wordt de vertekening besproken in de schatter voor het gemiddelde om te laten zien dat de vertekening ten gevolge van non-respons twee componenten heeft: het aandeel van de non-respondenten in de populatie en het verschil dat tussen respondenten en non-respondenten bestaat ten aanzien van het te onderzoeken kenmerk. Deze twee componenten zullen we weer tegenkomen wanneer de vertekening in de schatter voor een budgetaandeel wordt besproken. Voor het schatten van een gemiddelde $\bar{Y}$ in de populatie geldt dat de vertekening in de schatter $\bar{y}_r$ voor $\bar{Y}$ op grond van de respons wordt gegeven door

$$B(\bar{y}_r) = E(\bar{y}_r) - \bar{Y} = Q(\bar{y}_r - \bar{y}_{nr}), \quad (1)$$

waarbij $B(\bar{y}_r)$ de vertekening van de schatter $\bar{y}_r$,

$\bar{y}_r$ de schatter voor $\bar{Y}$ op grond van de respons,

$E(\bar{y}_r)$ de verwachting van $\bar{y}_r$,

$\bar{y}_r$ het gemiddelde in de populatie van respondenten (het responsstratum),

$\bar{y}_{nr}$ het gemiddelde in de populatie van non-respondenten (het non-responsstratum),

Q de fractie non-respons.

Het verschil tussen $\bar{Y}_r$ en $\bar{Y}_{nr}$ wordt het kontrast K genoemd, dus $K = \bar{Y}_r - \bar{Y}_{nr}$. Wanneer de fractie non-respons groot is en het contrast ook, dan is de vertekening groot.

Het contrast kan positief, nul of negatief zijn. Indien het contrast gelijk aan nul is, is er geen sprake van een vertekening. Is het contrast positief dan is de vertekening positief, hetgeen betekent dat de schatting voor het gemiddelde gebaseerd op de respons hoger is dan het werkelijke gemiddelde. Bij een negatief contrast is de vertekening negatief, de schatting op grond van de respons is te laag. Onder het absolute contrast verstaan we de absolute waarde van het contrast.

De waarde van het contrast is doorgaans onbekend. Het betreft populatieparameters en hoewel het mogelijk is zuivere schatters voor de populatie van respondenten te berekenen, is dit niet altijd mogelijk voor de populatie van non-respondenten.

De hier besproken theorie voor de vertekening in een gemiddelde kan worden uitgewerkt voor de vertekening in de schatter voor een budgetaandeel. Het budgetaandeel Y_i voor artikelengroep i is

$$Y_i = \frac{\sum_{h=1}^N X_{ih}}{\sum_{i=1}^L \sum_{h=1}^N X_{ih}} \quad (2)$$

waarbij X_{ih} de uitgave van huishouden h aan artikelengroep i ,
 N het totale aantal huishoudens in de populatie,
 L het totale aantal onderscheiden artikelengroepen.

Het budgetaandeel van artikelengroep i in de populatie is vanzelfsprekend onbekend en een budgetonderzoek heeft tot doel de budgetaandelen voor de verschillende artikelengroepen te schatten. Zij y_{ri} de schatter voor Y_i op grond van de respons, dan geldt dat de vertekening ten gevolge van non-respons (afgezien van de onzuiverheid van de quotiëntschatter voor het responsstratum) gegeven wordt door

$$B(Y_{r_i}) = \frac{\sum_{i=1}^L \sum_{h \in NR} X_{ih}}{\sum_{i=1}^L \sum_{h=1}^N X_{ih}} (Y_{r_i} - Y_{nr_i}) = \frac{Z_{nr}}{Z} (Y_{r_i} - Y_{nr_i}) . \quad (3)$$

Hierbij geeft $Z_{nr} = \sum_{i=1}^L \sum_{h \in NR} X_{ih}$ de som van alle uitgaven in het non-respons stratum, of te wel de totale bestedingen in het non-respons stratum,

$Z = \sum_{i=1}^L \sum_{h=1}^N X_{ih}$ de som van alle uitgaven in de populatie, of te wel de totale bestedingen in de populatie,

Y_{r_i} = het budgetaandeel van artikelengroep i in het responsstratum,

Y_{nr_i} = het budgetaandeel van artikelengroep i in het non-responsstratum,

Indien het budgetaandeel van artikelengroep i in het responsstratum en het non-responsstratum gelijk is dan is de vertekening nul. Een schatting voor Y_{r_i} kan worden verkregen op grond van de respondenten. De grootte van de vertekening wordt verder bepaald door het aandeel van alle uitgaven in het non-responsstratum (Z_{nr}) in de totale uitgaven in de populatie (Z). Dit aandeel is vanzelfsprekend onbekend. Als benadering hiervoor kan men nemen N_{nr}/N , het aandeel (n.b. het aandeel huishoudens) van het non-responsstratum in de gehele populatie. Dit aandeel kan worden geschat door het aandeel van de non-respons in de steekproef, Q . Dit veronderstelt dat de gemiddelde totale uitgaven van de huishoudens in het non-responsstratum gelijk zijn aan die in het responsstratum. Is dit niet het geval, dat wil zeggen zijn de totale uitgaven in het responsstratum bijvoorbeeld gemiddeld hoger, dan geeft de benadering een te rooskleurig beeld van de vertekening. Het omgekeerde doet zich voor indien het gemiddelde in het responsstratum lager is. De vertekening in de schatter voor budgetaandeel i op grond van de respons wordt (als $Z_{nr}/Z \approx Q$).

$$B(Y_{r_i}) \approx Q(Y_{r_i} - Y_{nr_i}) = OK_i . \quad (4)$$

In tabel 1 wordt voor verschillende percentages non-respons en verschillende waarden van het absolute contrast $|K|$, de absolute waarde van de vertekening in de schatting van een budgetaandeel gegeven. Voor een budgetonderzoek is het welhaast onmogelijk om een hoger percentage respons dan circa 20 procent te behalen door de grote non-respons bij de werving en door de uitval die later optreedt. Bij de interpretatie van tabel 1 moet men bedenken dat het gemiddeld budgetaandeel, wanneer we 48 artikelengroepen onderscheiden, circa 0,021 is. De vertekening in de schatting van een budgetaandeel is bij een absoluut contrast van 0,010 bij een budgetonderzoek met een non-respons van 70% dus 0,007.

Tabel 1. De vertekening in absolute waarde voor verschillende percentages non-respons en absolute contrastwaarden bij het schatten van een budgetaandeel

percentage non-respons	absolute contrastwaarde $ K $				
	0,005	0,010	0,015	0,020	0,025
30	0,0015	0,0030	0,0045	0,0060	0,0075
40	0,0020	0,0040	0,0060	0,0080	0,0100
50	0,0025	0,0050	0,0075	0,0100	0,0125
60	0,0030	0,0060	0,0090	0,0120	0,0150
70	0,0035	0,0070	0,0105	0,0140	0,0175
80	0,0040	0,0080	0,0120	0,0160	0,0200
90	0,0045	0,0090	0,0135	0,0180	0,0225

Met behulp van tabel 1 is het mogelijk om de invloed van de non-respons op de geschatte budgetaandelen te bepalen. Het vaststellen van deze invloed kan een hulpmiddel zijn bij het interpreteren van de uitkomsten van de budgetonderzoeken en bij het bepalen van het aggregatieniveau waarop gepubliceerd wordt.

3. De mogelijke vertekening in een prijsindexcijfer

Aangezien de budgetaandelen die geschat worden op grond van een budgetonderzoek gebruikt worden voor de constructie van een consumenten prijsindexcijfer is het mogelijk dat de vertekening ten gevolge van non-respons doorwerkt in dit prijsindexcijfer. Een prijsindexcijfer is een doorgaans gewo-

gen rekenkundig gemiddelde van enkelvoudige prijsindexcijfers waarbij gewogen wordt met de geschatte budgetaandelen in een basisperiode. In formulevorm is het prijsindexcijfer voor periode t

$$P_t = \sum_{i=1}^L Y_i P_{it} . \quad (5)$$

Hierbij is P_t het samengestelde prijsindexcijfer voor periode t ,
 L het aantal onderscheiden artikelengroepen,
 Y_i het populatie-budgetaandeel van artikelengroep i ,
 P_{it} het enkelvoudige prijsindexcijfer voor periode t van artikelengroep i .

P_t is onbekend en kan worden geschat door Y_i te schatten door y_r . De waarde van P_{it} wordt geschat op grond van een ander steekproefonderzoek. We houden geen rekening met de variantie en een eventuele vertekening in de schatting voor P_{it} . De schatter voor het prijsindexcijfer op grond van de respons van het budgetonderzoek wordt nu:

$$P_{t_r} = \sum_{i=1}^L y_{r_i} P_{it} . \quad (6)$$

De vertekening in het prijsindexcijfer ten gevolge van non-respons is

$$B(P_{t_r}) = \sum_{i=1}^L B(y_{r_i}) P_{it} = \sum_{i=1}^L Q(Y_{r_i} - Y_{nr_i}) P_{it} = \sum_{i=1}^L Q K_i P_{it} . \quad (7)$$

Merk op dat wanneer de enkelvoudige prijsindexcijfers P_{it} aan elkaar gelijk zijn voor periode t , de non-respons in het budgetonderzoek geen invloed heeft op het prijsindexcijfer.

De grootheden Q (de fractie non-respons), Y_{r_i} (het budgetaandeel in het responsstratum van artikelengroep i) en P_{it} (de enkelvoudige prijsindexcijfers) zijn bekend of kunnen eenvoudig worden geschat. Ten aanzien van de vertekening $B(y_{r_i}) = Q(Y_{r_i} - Y_{nr_i})$ in de schatting voor budgetaandeel Y_i geldt dat een aantal randvoorwaarden aanwezig zijn. In de eerste plaats geldt dat

$$\sum_{i=1}^L Y_{r_i} = 1$$

en

$$\sum_{i=1}^L Y_{nr_i} = 1. \quad (8)$$

Daarnaast geldt dat $0 \leq Y_{r_i}$ voor alle i en evenzo $0 \leq Y_{nr_i}$. Verder geldt dat $Y_{nr_i} = Y_{r_i} - K_i \cdot Y_{r_i}$ kan op grond van de respons worden geschat. Bij een zekere waarde van Y_{nr_i} , ligt K_i vast en kan $B(p_r)$ worden berekend. De vraag welk contrast K_i aanwezig is laat zich moeilijk of niet beantwoorden.

We kunnen een indicatie voor de mogelijke doorwerking van de non-respons in het prijsindexcijfer verkrijgen door een ondergrens en bovengrens van (7) te berekenen. Dit is mogelijk door te stellen dat het absolute contrast van artikelengroep i ten hoogste de waarde M_i aanneemt. Dit betekent dat verondersteld wordt dat het budgetaandeel van artikelengroep i voor het non-responsstratum zich in het interval

$$(Y_{r_i} - M_i, Y_{r_i} + M_i) \quad (9)$$

bevindt. Dit leidt tot een lineair programmeringsprobleem dat er als volgt uitziet

$$\min_{Y_{nr_i}} \left[B(p_r) \right] \begin{cases} 0 \leq Y_{nr_i} & \text{voor } i=1,2,\dots,L \\ Y_{r_i} - M_i \leq Y_{nr_i} \leq Y_{r_i} + M_i & \text{voor } i=1,2,\dots,L \\ \sum_{i=1}^L Y_{nr_i} = 1 \end{cases} \quad (10)$$

Oplossen van dit probleem met behulp van lineaire programmering waarbij we Y_{r_i} schatten met y_{r_i} , geeft een onder- en bovengrens van de vertekening in het prijsindexcijfer ten gevolge van non-respons.

De grootte van het interval waarbinnen we Y_{nr_i} laten variëren moet worden vastgelegd. Het is mogelijk om voor elke artikelengroep i dezelfde breedte M_i te veronderstellen (met andere woorden $M_i=M$ voor alle i), het is ook mogelijk om voor elke artikelengroep i een zelfde relatieve breedte te veronderstellen (met andere woorden $M_i/Y_{r_i}=c$). Door gebruik te maken van de standaardafwijking van de budgetaandelen van de respondenten kan de te hanteren breedte van het interval waarbinnen Y_{nr_i} ligt ook worden vastgelegd. Zij S_i de standaardafwijking behorende bij het budgetaandeel Y_{r_i} dan veronderstellen we dat de breedte M_i evenredig is met S_i , dus $M_i=\alpha S_i$. Het interval waarbinnen we Y_{nr_i} laten variëren wordt nu $(Y_{r_i}-\alpha S_i, Y_{r_i}+\alpha S_i)$, waarbij $\underline{s}_i$ de schatter op grond van de respons voor S_i is.

De hierboven geformuleerde theorie kan worden gebruikt om de gevoeligheid van een prijsindexcijfer voor de invloed van non-respons bij een budgetonderzoek te bepalen. We gaan bij wijze van voorbeeld na wat de bovengrens en wat de ondergrens is van de vertekening in een prijsindexcijfer dat gebaseerd is op 48 artikelengroepen. Als basisjaar nemen we 1975 en we nemen de budgetaandelen zoals die berekend zijn voor werknemersgezinnen (CBS, 1979) op het aggregatieniveau van 48 artikelengroepen. De enkelvoudige prijsindices zijn de prijsindices voor deze 48 groepen zoals die berekend kunnen worden op grond van enkelvoudige prijsindices en geschatte budgetaandelen van een lager aggregatieniveau. Voor de berekening van de gevoeligheid gaan we uit van het percentage non-respons bij werknemersgezinnen in het Budgetonderzoek 1978: 74,4% (zie CBS, 1982). In tabel 2 worden voor 48 artikelengroepen de geschatte budgetaandelen van de respondenten (y_{r_i}) gegeven alsmede de enkelvoudige prijsindexcijfers in 1981 van de 48 artikelengroepen die we onderscheiden.

In tabel 3 worden de berekende budgetaandelen voor de non-respondenten gegeven die leiden tot een bovengrens van de vertekening in het prijsindexcijfer voor 1981. Ter vergelijking zijn ook de budgetaandelen van de respondenten opgenomen. In tabel 4 worden resultaten van de berekeningen die leiden tot een ondergrens gepresenteerd. De uitkomsten laten zien dat wanneer de budgetaandelen van de non-respondenten ten hoogste 100×10^{-5} van die van de respondenten verschillen ($M_i=100 \times 10^{-5}$) de bovengrens van de mogelijke vertekening 0,53 en de ondergrens -0,50 bedraagt. Dit wil zeggen

Tabel 2. De 48 onderscheiden artikelengroepen, het budgetaandeel $\times 100000$ voor de respondenten en de enkelvoudige prijsindices voor 1981 (1975=100)

artikelengroep	wegingsfactor	enkelvoudige prijsindexcijfer
1 brood, gebak en grutterwaren	3880	150,1
2 aardappelen en groente	2385	134,8
3 fruit	1900	135,3
4 suikerhoudende artikelen	2200	141,3
5 dranken (niet alcoholisch)	1170	104,0
6 dranken (zwak alcoholisch)	1227	117,1
7 dranken (sterk alcoholisch)	913	128,9
8 oliën en vetten	960	109,0
9 vlees, vleeswaren en vis	6900	130,4
10 zuivelproducten	4200	130,1
11 verteringen buitenshuis	1220	149,8
12 overige voedingsmiddelen	1045	118,7
13 huur of huurwaarde	10700	157,7
14 bijkomende kosten n.e.g.	400	143,6
15 onderhoud woning	1394	125,5
16 nagelvaste installaties	307	175,5
17 tuin en bloemen	815	127,6
18 meubelen, stoffering en linnengoed	4952	139,2
19 huishoudelijke apparaten en gereedschappen	2901	125,4
20 verwarming en verlichting	3966	209,6
21 herenkleding	1800	130,3
22 jongenskleding	800	137,4
23 dameskleding	2500	136,3
24 meisjeskleding	900	131,8
25 overige kleding	700	133,2
26 stoffen en maakloon	800	124,7
27 schoeisel	1600	167,3
28 opschik	630	167,3
29 huishoudelijke dienstverlening en reiniging	400	146,2
30 reinigingsartikelen	1500	127,0
31 lichamelijke verzorging	1500	132,8
32 geneeskundige verzorging en ziektekostenverzekering	11100	154,0
33 opleiding	1467	152,3
34 schrijfbehoeften en lectuur	1230	148,1
35 sport, spel en kampeeruitrusting	811	139,8
36 kamperen en weekendbesteding	700	137,9
37 vakanties	1156	172,6
38 muziek, zang en toneel	200	122,3
39 radio, t.v. en grammofoon	2970	99,9
40 overige ontspanning	1500	131,3
41 speelgoed	689	148,0
42 roken	1570	155,9
43 openbaar vervoer	580	160,2
44 rijwielen, bromfietsen en motoren	700	142,5
45 auto's	4470	135,7
46 overige kosten eigen vervoer	2581	169,2
47 overige verkeers- en vervoerskosten	800	102,9
48 overige bestedingen	911	146,7
Totaal	100000	142,8

Tabel 3. Budgetaandelen $\times 100000$ van de respons en de berekende budgetaandelen $\times 100000$ van de non-respons die leiden tot een bovengrens in de vertekening in het onderzochte prijsindexcijfer voor 1981 voor zes alternatieven

artikelengroep	budgetaandeel respons Y_i	berekende budgetaandelen non-respons					
		$M_1=100$ $\times 10^{-5}$	$M_1=500$ $\times 10^{-5}$	$M_1=0,1Y_i$ r_i	$M_1=0,2Y_i$ r_i	$M_1=0,1g_i$ g_i	$M_1=0,2g_i$ g_i
1 brood, gebak en grutterwaren	3880	3780	3380	3492	3104	3744	3608
2 aardappelen en groente	2385	2485	2885	2624	2862	2492	2599
3 fruit	1900	2000	2400	2090	2280	1989	2078
4 suikerhoudende artikelen	2200	2100	1700	1980	1760	2097	1994
5 dranken (niet alcoholisch)	1170	1270	1670	1287	1404	1236	1302
6 dranken (zwak alcoholisch)	1227	1327	1727	1350	1472	1355	1483
7 dranken (sterk alcoholisch)	913	1013	1413	1004	1096	1052	1191
8 oliën en vetten	960	1060	1460	1056	1152	1006	1052
9 vlees, vleeswaren en vis	6900	7000	7400	7590	8280	7182	7464
10 zuivelproducten	4200	4300	4700	4620	5040	4357	4514
11 verteringen buitenshuis	1220	1120	720	1098	976	595	200
12 overige voedingsmiddelen	1045	1145	1545	1150	1254	1218	1391
13 huur of huurwaarde	10700	10600	10200	9630	8560	10165	9510
14 toekomstige kosten n.e.g.	400	300	200	360	320	286	200
15 onderhoud woning	1394	1494	1894	1533	1673	1752	2110
16 nagelaste installaties	307	207	200	276	246	200	200
17 tuin en bloemen	815	915	1315	897	978	929	1043
18 meubelen, stoffering en linengoed	4952	4852	4452	5067	5182	4352	3752
19 huishoudelijke apparaten en gereedschappen	2901	3001	3401	3191	3481	3146	3391
20 verwarming en verlichting	3966	3866	3465	3569	3173	3793	3620
21 herenkleding	1800	1900	2300	1980	2160	1935	2070
22 jongenskleding	800	700	300	880	960	912	1024
23 dameskleding	2500	2600	2000	2750	3000	2660	2820
24 meisjeskleding	900	1000	1400	990	1080	1070	1140
25 overige kleding	700	800	1200	770	840	813	926
26 stoffen en maakloon	800	900	1300	880	960	924	1048
27 schoeisel	1600	1500	1100	1440	1280	1507	1414
28 opschik	630	530	200	567	504	555	480
29 huishoudelijke dienstverlening en reiniging	400	300	200	360	320	275	200
30 reinigingsartikelen	1500	1600	2000	1650	1800	1572	1644
31 lichamelijke verzorging	1500	1600	2000	1650	1800	1608	1716
32 geneeskundige verzorging en ziektekostenverzekering	11100	11000	10600	9990	8880	10747	10394
33 opleiding	1467	1367	967	1370	1174	1230	993
34 schrijfbehoeften en lectuur	1230	1130	730	1107	984	1117	1004
35 sport, spel en kampeeruitrusting	811	711	311	730	649	681	551
36 kamperen en weekendbesteding	700	600	200	770	840	816	517
37 vakanties	1156	1056	656	1040	925	890	624
38 muziek, zang en toneel	200	300	700	220	240	405	610
39 radio, t.v. en grammofoon	2970	3070	3470	3267	3564	3290	3610
40 overige ontspanning	1500	1600	2000	1650	1800	1680	1860
41 speelgoed	689	589	200	620	551	597	505
42 roken	1570	1470	1070	1413	1256	1421	1272
43 openbaar vervoer	580	480	200	522	464	407	234
44 rijtuigen, bromfietsen en motoren	700	600	200	630	560	580	460
45 auto's	4470	4570	4777	4917	5364	5350	6230
46 overige kosten eigen vervoer	2581	2481	2081	2323	2065	2296	2011
47 overige verkeers- en vervoerskosten	800	900	1300	880	960	1009	1218
48 overige bestedingen	911	811	411	820	729	817	723
Totaal	100000	100000	100000	100000	100000	100000	100000
Bovengrens vertekening (prijsindexcijfer 1981=142,8)		0,53	2,43	1,12	2,23	1,01	1,96
Uiterste waarde van de index		142,3	140,4	141,7	140,6	141,8	140,8

Tabel 4. Budgetaandelen $\times 100000$ van de respons en de berekende budgetaandelen $\times 100000$ van de non-respons die leiden tot een ondergrens in de vertekening in het onderzochte prijsindexcijfer van 1981 voor zes alternatieven

artikelen-groep	budgetaandeel respons Y_i	berekende budgetaandelen non-respons					
		$M_i = 100$ $\times 10^{-5}$	$M_i = 500$ $\times 10^{-5}$	$M_i = 0,1 Y_i$	$M_i = 0,2 Y_i$	$M_i = 0,18 Y_i$	$M_i = 0,25 Y_i$
1 brood, gebak en grutterwaren	3880	3980	4380	4268	4656	4016	4152
2 aardappelen en groente	2385	2285	1885	2147	1908	2278	2171
3 fruit	1900	1800	1400	1710	1520	1811	1722
4 suikerhoudende artikelen	2200	2300	2700	2420	2640	2300	2400
5 dranken (niet alcoholisch)	1170	1070	669	1053	936	1104	1039
6 dranken (zwak alcoholisch)	1227	1127	727	1104	982	1099	971
7 dranken (sterk alcoholisch)	913	813	413	822	730	774	635
8 oliën en vetten	960	860	460	864	768	1006	868
9 vlees, vleeswaren en vis	6900	6800	6400	6210	5520	6618	6336
10 zuivelproducten	4200	4100	3700	3780	3360	4043	3886
11 verteringen buitenshuis	1220	1320	1720	1342	1464	1845	2470
12 overige voedingsmiddelen	1045	945	545	940	816	872	699
13 huur of huurwaarde	10700	10800	11200	11770	12840	11295	11890
14 bijkomende kosten n.e.g.	400	500	900	440	480	514	628
15 onderhoud woning	1394	1294	894	1255	1115	1036	678
16 nagelveste installaties	307	407	807	338	368	625	943
17 tuyn en bloemen	815	715	315	733	652	701	587
18 meubelen, stoffering en linengoed	4952	5052	5452	4817	4682	5238	5588
19 huishoudelijke apparaten en gereedschappen	2901	2801	2401	2611	2321	2656	2411
20 verwarming en verlichting	3966	4066	4466	4363	4759	4139	4312
21 herenkleding	1800	1700	1300	1620	1440	1665	1530
22 jongenskleding	800	700	801	720	640	688	576
23 dameskleding	2500	2400	2000	2250	2000	2340	2180
24 meisjeskleding	900	800	400	810	720	780	660
25 overige kleding	700	600	200	630	560	587	474
26 stoffen en maakloon	800	700	300	720	640	676	552
27 schoeisel	1600	1700	2100	1760	1920	1693	1786
28 opschik	630	730	1130	693	756	705	780
29 huishoudelijke dienstverlening en reiniging	400	500	900	440	480	525	650
30 reinigingsartikelen	1500	1400	1000	1350	1200	1428	1356
31 lichamelijke verzorging	1500	1400	1000	1350	1200	1392	1284
32 geneeskundige verzorging en ziektekostenverzekering	11100	11200	11600	12210	13320	11453	11806
33 opleiding	1467	1567	1967	1614	1760	1704	1941
34 schrijfbehoeften en lectuur	1230	1330	1730	1353	1476	1343	1456
35 sport, spel en kampeeruitrusting	811	911	1311	892	973	941	1071
36 kamperen en weekendbesteding	700	800	1200	630	560	390	200
37 vakanties	1156	1256	1656	1272	1387	1422	1688
38 muziek, zang en toneel	200	300	200	200	200	200	200
39 radio, t.v. en grammofoon	2970	2870	2470	2673	2376	2650	2330
40 overige ontspanning	1500	1400	1000	1350	1200	1320	1140
41 speelgoed	689	789	1189	758	827	781	873
42 roken	1570	1670	2070	1727	1884	1719	1868
43 openbaar vervoer	580	680	1080	638	696	753	926
44 rijwielen, bromfietsen en motoren	700	800	1200	770	840	820	940
45 auto's	4470	4370	3970	4023	3576	3590	2710
46 overige kosten eigen vervoer	2581	2681	3081	2839	3097	2866	3151
47 overige verkeers- en vervoerskosten	800	700	300	720	640	591	382
48 overige bestedingen	911	1011	1411	1002	1093	1005	1099
Totaal	100000	100000	100000	100000	100000	100000	100000
Ondergrens van de vertekening (prijsindex 1981=142,8)		-0,50	-2,56	-1,11	-2,22	-1,01	-2,07
Uiterste waarde van de index		143,3	145,4	143,9	145,0	143,8	144,9

dat de prijsindex voor 1981 ligt tussen 142,3 en 143,3. Staan we meer variatie toe ($M_i = 500 \times 10^{-5}$) dan vinden we een interval van (140,4 - 145,4). De twee methoden waarbij de breedte van het interval waarbinnen het budgetaandeel van de non-respondenten mag variëren evenredig is met het budgetaandeel van de respondenten leidt tot intervallen (141,7 - 143,9) en (140,6 - 145,0). Het eerste interval is bij 10% variatie, het tweede interval bij 20% variatie. Bij de laatste twee alternatieven, tenslotte, is verondersteld dat de breedte van het interval respectievelijk $0,1 \underline{s}_i$ en $0,2 \underline{s}_i$ is, waarbij $\underline{s}_i$ de geschatte standaardafwijking van budgetaandeel i van de respondenten is. (Om de gedachten te bepalen: $\underline{s}_i$ is gemiddeld 2100×10^{-5}). Dit leidt tot de intervallen (141,8 - 143,8), respectievelijk (140,8 - 144,9). De gevoeligheid van deze prijsindex voor de invloed van non-respons blijkt relatief gezien gering te zijn. De berekende waarden geven namelijk de meest ongunstige situatie weer. We hebben immers als doelfunctie voor het lineaire programmeringsprobleem de vertekening in het prijsindexcijfer genomen.

De invloed van de non-respons bij een budgetonderzoek op een prijsindex hangt behalve van het behaalde percentage non-respons en het contrast, ook af van de enkelvoudige prijsindices. Formule (7) laat zien hoe deze indices doorwerken in de totale vertekening. Naarmate de verschillen tussen de enkelvoudige prijsindices groter zijn, zal de totale vertekening groter kunnen zijn. In tabel 5 worden de onder- en bovengrenzen van de onderzochte prijsindexcijfers van de jaren 1976 t/m 1981 gegeven. Deze grenzen zijn be-

Tabel 5. Intervallen op grond van de berekende budgetaandelen van de non-respondenten die leiden tot een ondergrens, respectievelijk bovengrens van de vertekening in het jaarcijfer 1981 ($M_i = 0,2 \underline{s}_i$)

jaar	prijsindex op grond van de respons	ondergrens van deze index	bovengrens van deze index
1975	100	100	100
1976	108,8	108,5	109,1
1977	115,8	115,2	116,3
1978	120,5	119,7	121,4
1979	125,6	124,5	126,8
1980	133,8	132,3	135,4
1981	142,8	140,8	144,9

rekend op grond van de berekende budgetaandelen van de non-respondenten die leiden tot de boven- of ondergrens van de vertekening in het indexcijfer van 1981 bij $M_i=0,2s_i$. De toename van de afstand van de grenzen tot de index berekend op grond van de respondenten in de loop der jaren is duidelijk zichtbaar.

De invloed van de dispersie van de enkelvoudige prijsindices op de vertekening kan ertoe leiden dat de boven- en ondergrenzen van de vertekening in de maandindices groter kunnen zijn dan in de jaarindices. De verschillen tussen de enkelvoudige prijsindices zullen namelijk in de loop van het jaar groter kunnen zijn dan de verschillen tussen de enkelvoudige prijsindices op jaarbasis, welke een soort gemiddelde zijn.

4. De mogelijke invloed van een ondervertegenwoordiging van hoge inkomens

Volgens CBS (1982) zijn bij het Budgetonderzoek 1978 de hoogste en de laagste inkomensgroep ondervertegenwoordigd. Gezien het verband tussen het inkomen en de wijze waarop men het inkomen verbruikt is deze non-respons voor het Budgetonderzoek van belang. We houden ons in deze paragraaf bezig met de vraag in hoeverre een ondervertegenwoordiging van de hogere inkomens doorwerkt in een prijsindexcijfer. We veronderstellen hierbij impliciet dat de invloed van de ondervertegenwoordiging van de hogere inkomens van meer belang is dan die van de ondervertegenwoordiging van de lagere inkomens, mede gezien de wijze waarop een budgetaandeel wordt berekend. Uit onderzoek van het consumentengedrag zijn resultaten beschikbaar van verbruikselasticiteiten, zie hiervoor o.a. Van Driel (1979). Voor de verschillende artikelen zijn door Van Driel verbruikselasticiteiten geschat. We gebruiken de verbruikselasticiteiten om het interval waarbinnen het budgetaandeel voor het non-respons stratum mag variëren, nader te concretiseren. Dit interval is voor artikelengroep i:

$$(Y_{r_i} - M_i, Y_{r_i} + M_i) .$$

Wanneer de verbruikselasticiteit van artikelengroep i groter dan 1 is, betekent dat dat de hogere inkomens in het algemeen relatief gezien meer aan deze artikelengroep zullen uitgeven. Het op grond van de respons (waar een ondervertegenwoordiging van hogere inkomens is) berekende budgetaandeel is

Tabel 6. Budgetaandelen $\times 100000$ van de respons en de berekende budgetaandelen $\times 100000$ van de non-respons die leiden tot een onder- dan wel bovengrens van de vertekening in het onderzochte prijsindexcijfer, onder veronderstellingen over de invloed van de ondervertegenwoordiging van hogere inkomens

artikelen groep	budgetaandeel respons	berekende budgetaandelen van de non-respons	
		die leiden tot de ondergrens van de vertekening	die leiden tot de bovengrens van de vertekening
1 brood, gebak en grutterwaren	3880	3608	3608
2 aardappelen en groente	2385	2171	2171
3 fruit	1900	1722	1722
4 suikerhoudende artikelen	2200	1994	1994
5 dranken (niet alcoholisch)	1170	1038	1170
6 dranken (zwak alcoholisch)	1227	1227	1483
7 dranken (sterk alcoholisch)	913	913	1191
8 oliën en vetten	960	868	960
9 vlees, vleeswaren en vis	6900	6336	6900
10 zuivelproducten	4200	3886	4200
11 verteringen buitenshuis	1220	2364	1220
12 overige voedingsmiddelen	1045	699	1045
13 huur of huurwaarde	10700	11890	10700
14 bijkomende kosten n.e.g.	400	400	400
15 onderhoud woning	1394	1394	2110
16 nagelaste installaties	307	943	307
17 tuin en bloemen	815	815	1043
18 meubelen, stoffering en linnengoed	4952	4952	4952
19 huishoudelijke apparaten en gereedschappen	2901	2901	3391
20 verwarming en verlichting	3966	3966	3620
21 herenkleding	1800	1530	1800
22 jongenskleding	800	576	576
23 dameskleding	2500	2500	2500
24 meisjeskleding	900	660	900
25 overige kleding	700	474	690
26 stoffen en maakloon	800	552	800
27 schoeisel	1600	1600	1414
28 opschik	630	780	630
29 huishoudelijke dienstverlening en reiniging	400	400	400
30 reinigingsartikelen	1500	1356	1500
31 lichamelijke verzorging	1500	1500	1716
32 geneeskundige verzorging en ziektekostenverzekering	11100	11100	10394
33 opleiding	1467	1941	1467
34 schrijfbehoeften en lectuur	1230	1230	1230
35 sport, spel en kampeeruitrusting	811	811	811
36 kamperen en weekendbesteding	700	700	700
37 vakanties	1156	1688	1156
38 muziek, zang en toneel	200	200	610
39 radio, t.v. en grammofoon	2970	2330	2970
40 overige ontspanning	1500	1140	1500
41 speelgoed	689	505	505
42 roken	1570	1570	1272
43 openbaar vervoer	580	926	580
44 rijwielen, bromfietsen en motoren	700	700	700
45 auto's	4470	4470	4470
46 overige kosten eigen vervoer	2581	3151	2581
47 overige verkeers- en vervoerskosten	800	800	1218
48 overige bestedingen	911	723	723
Totaal	100000	100000	100000
Grenzen voor de vertekening in de totale index		-1,24	0,73
Grenzen voor de index (1981=142,8)		144,0	142,1

derhalve te laag. Het interval waarbinnen het budgetaandeel van het non-responsstratum mag variëren wordt nu gesteld op

$$(Y_{r_i}, Y_{r_i} + M_i) .$$

Het omgekeerde doet zich voor bij een verbruikselasticiteit die kleiner dan 1 is. Passen we deze theorie toe op het jaarcijfer 1981 van de door ons onderzochte prijsindex, met $M_i = 0,2s_i$ voor artikelengroep i , dan wordt met behulp van lineaire programmering een bovengrens voor de vertekening van 0,73 en een ondergrens van -1,24 gevonden. Dit betekent dat het op grond van de respons berekende prijsindexcijfer, onder deze veronderstellingen maximaal 1,24 punten te laag is en maximaal 0,73 punten te hoog. In tabel 6 worden de berekende budgetaandelen voor de non-respondenten gegeven die leiden tot de onder- en bovengrens van deze vertekening, alsmede de budgetaandelen voor de respondenten. Deze uitkomsten lijken er op te wijzen dat indien er een vertekening aanwezig is het op grond van de respons berekende prijsindexcijfer eerder wat aan de lage kant dan aan de hoge kant is. Vanzelfsprekend geldt dat de door ons gehanteerde veronderstellingen en randvoorwaarden voor de budgetaandelen van de non-respondenten, niet noodzakelijk realistisch zijn.

5. De invloed van non-respons op verschillen in indexcijfers

Voor het bepalen van de mogelijke invloed van non-respons bij een budgetonderzoek op het verschil in prijsindexcijfers kan men op 2 manieren te werk gaan. De eerste manier is het bepalen van een boven- en ondergrens van de vertekening in het verschil tussen twee indexcijfers. Zij P_t het totale prijsindexcijfer voor jaar t en zij P_{t+1} het totale prijsindexcijfer voor jaar $t+1$. De vertekening in $\Delta P_{t+1} = P_{t+1} - P_t$ wordt gegeven door

$$B(\Delta P_{t+1}) = B(P_{t+1}) - B(P_t) = Q \sum_i (Y_{ir} - Y_{inr})(P_{it+1} - P_{it}) , \quad (11)$$

waarbij P_{it+1} het partiële prijsindexcijfer voor jaar $t+1$ en artikelen-groep i

P_{it} het partiële prijsindexcijfer voor jaar t en artikelen-groep i .

Nemen we (11) als doelfunctie voor het lineaire programmeringsprobleem en bepalen we de ondergrens en de bovengrens van (11) dan worden budgetaandelen voor de non-respondenten gevonden. Deze berekende budgetaandelen behoeven echter niet noodzakelijkerwijs gelijk te zijn aan de berekende budgetaandelen wanneer we twee (andere) jaren bekijken. Aangezien de indexcijfers alle gebaseerd zijn op hetzelfde budgetonderzoek is het beter om uitgaande van eenmaal berekende budgetaandelen van de non-respondenten de verschillen tussen de indexcijfers op grond van deze berekende budgetaandelen te bepalen. In tabel 7 worden de procentuele verschillen tussen de indexcijfers van jaar op jaar gegeven. Als uitgangspunt zijn de berekende budgetaandelen van de non-respondenten genomen die leiden tot een onder- of bovengrens van de vertekening in het prijsindexcijfer van 1981 (zie hiervoor ook tabellen 3 en 4, met $M_1=0,2s_1$). Duidelijk is te zien dat we ook voor het procentuele verschil tussen indexcijfers een interval kunnen geven en dat dit interval nagenoeg symmetrisch is ten opzichte van de op grond van de respondenten berekende prijsindex.

Tabel 7. Procentuele verschillen in het prijsindexcijfer van jaar op jaar op basis van de respondenten en op basis van de budgetaandelen van de respondenten en berekende budgetaandelen van de non-respondenten gezamenlijk ($M_1=0,2s_1$)

jaar	op basis van de respondenten	op basis van de ondergrens van het interval	op basis van de bovengrens van het interval
1976	8,8	8,5	9,1
1977	6,4	6,2	6,6
1978	4,1	3,9	4,4
1979	4,2	4,0	4,4
1980	6,5	6,3	6,8
1981	6,7	6,5	7,0

6. Conclusie

De non-respons bij een budgetonderzoek kan leiden tot vertekening in de schattingen van de budgetaandelen. Een prijsindex die gebruik maakt van de bij het budgetonderzoek berekende schattingen, kan een mogelijke vertekening 'meenemen'. De gevoeligheid voor non-respons van een prijsindex kan

onder bepaalde veronderstellingen eenvoudig worden bepaald. We hebben onderzocht wat in het meest ongunstige geval de boven- en ondergrens van de vertekening in de door ons beschouwde prijsindex is. Door gebruik te maken van geschatte verbruikselasticiteiten kan ook de mogelijke invloed van een eventuele ondervertegenwoordiging van huishoudens met een hoog inkomen op de door ons onderzochte index worden bepaald. Voor de index die onderzocht is en de veronderstellingen die wij hebben gemaakt, geldt dat indien er van invloed van non-respons sprake is, de op grond van de respons berekende index eerder wat aan de lage kant zou kunnen zijn dan aan de hoge kant. De maximaal mogelijke vertekening ten gevolge van non-respons in de onderzochte prijsindex, blijkt gering te zijn.

Literatuur

CBS (Centraal Bureau voor de Statistiek), 1979, Werknemersbudgetonderzoek 1974/'75 (Staatsuitgeverij, 's-Gravenhage).

CBS (Centraal Bureau voor de Statistiek), 1982, Budgetonderzoek 1978, deel 1, methodologie: enquêtedocumenten (Staatsuitgeverij, 's-Gravenhage).

Driel, J. van, 1979, Verbruikselasticiteiten en budgetaandelen voor werknemersgezinnen, geschat uit het Budgetonderzoek 1974/'75. Intern CBS-rapport (Centraal Bureau voor de Statistiek, Voorburg).

Kersten, H.M.P. en J.G. Bethlehem, 1981, Het vaste responsmodel. Intern CBS-rapport (Centraal Bureau voor de Statistiek, Voorburg).

Ontvangen: 19-9-1983