

# Het gebruik van respons propensities voor nonrespons correctie

Fannie Cobben

*Centraal Bureau voor de Statistiek  
Afdeling Methodologie*



*NPSO lezingendag, 23 september 2008*

## In deze presentatie...

- Respons propensities
- Nonrespons correctie methoden
- Nonrespons correctie & respons propensities
- Toepassing op POLS 2002
- Vergelijking methoden
- Discussie/vragen...



# Notatie

Populatie

$$N, i = 1, 2, \dots, N$$

Selectie indicator

$$\partial_i = \begin{cases} 1, & \text{als element } i \text{ in steekproef} \\ 0, & \text{anders} \end{cases}$$

Steekproef omvang

$$n = \sum_{i=1}^N \partial_i$$

Eerste orde insluit kansen

$$\pi_i = P(\partial_i = 1)$$

Doelvariabele

$$Y_i$$

Gemiddelde doelvariabele

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N Y_i$$

Achtergrondkenmerken

$$X_i = (X_{1i}, X_{2i}, \dots, X_{ji})'$$

Respons indicator

$$R_i = \begin{cases} 1, & \text{respons element } i \\ 0, & \text{nonrespons element } i \end{cases}$$

Respons omvang

$$n_r = \sum_{i=1}^n R_i$$



# Respons propensity

Respons kans

$$\rho_i = P(R_i = 1 | \partial_i = 1)$$

Respons propensity

$$\hat{\rho}_i = \rho(X_i) = P(R_i = 1 | \partial_i = 1, X_i)$$

Te schatten met bijvoorbeeld een logit model

$$\hat{\rho}_i = \rho(X_i) = P(R_i = 1 | \partial_i = 1, X_i) = \frac{\exp(-\beta X_i)}{1 + \exp(-\beta X_i)}$$

of een probit model

$$\hat{\rho}_i = \rho(X_i) = P(R_i = 1 | \partial_i = 1, X_i) = \Phi(\beta X_i)$$

M.a.w. respons propensity is **geschatte respons kans**



# Nonrespons correctie

'Adjustment weighting'

$$\bar{y}_w = \frac{1}{N} \sum_{i=1}^n w_i R_i Y_i = \frac{1}{N} \sum_{i=1}^{n_r} w_i Y_i \quad w_i = \frac{1}{\pi_i} g_i = d_i g_i$$

Horvitz-Thompson schatter

$$g_i = 1 \quad \bar{y}_{ht}^r = \frac{1}{N} \sum_{i=1}^{n_r} d_i Y_i$$

GREG - schatter

$$g_i = 1 + (\bar{X} - \bar{x}_{ht})' \left( \sum_{i=1}^{n_r} d_i X_i X_i' \right)^{-1} X_i \quad \bar{y}_{gr}^r = \frac{1}{N} \sum_{i=1}^{n_r} d_i g_i Y_i$$



# Nonrespons correctie & respons propensities

## Direct

- *wegen*
- *stratificatie*

Géén calibratie eigenschappen gebruikte achtergrondkenmerken!

## Indirect

- *GREG – schatter met aangepaste insluitkansen*



# Direct gebruik respons propensities

Respons propensity **wegen**, Särndal (1981)

$$g_i = \frac{1}{\hat{\rho}} \quad \bar{y}_w^{\hat{\rho}} = \frac{1}{N} \sum_{i=1}^{n_r} d_i g_i Y_i = \frac{1}{N} \sum_{i=1}^{n_r} \frac{Y_i}{\pi_i \hat{\rho}_i}$$

Respons propensity **stratificatie**

Maak  $f = 1, 2, \dots, F$  strata gebaseerd op  $\hat{\rho}_i$

$n_f$  omvang stratum  $f$

$n_{r,f}$  het aantal respondenten in stratum  $f$

$$g_i = \frac{n_f}{n_{r,f}} \quad \bar{y}_{ps}^{\hat{\rho}} = \frac{1}{N} \sum_{i=1}^{n_r} d_i g_i Y_i = \frac{1}{N} \sum_{f=1}^F n_{r,f} \bar{y}_{ht}^{(f)}$$



# Indirect gebruik respons propensities

GREG – schatter met aangepaste insluitkansen

$$g_i = 1 + (\bar{X} - \bar{x}_{ht}^r)' \left( \sum_{i=1}^{n_r} \frac{d_i}{\hat{\rho}_i} X_i X_i' \right)^{-1} X_i$$

$$\bar{y}_{gr}^{\hat{\rho}} = \frac{1}{N} \sum_{i=1}^{n_r} d_i g_i Y_i$$



# POLS 2002

- Permanent Onderzoek LeefSituatie
- Maandelijks worden 3.000 personen geselecteerd
- Vragen over leefsituatie; veiligheid, recht en participatie, gezondheid
- Basis module voor personen > 12 jaar
  
- Dataset geaggregeerd over 2002:  $n = 35.594$  en  $n_r = 20.168$  (57%)
- Doelvariabelen: *werkzaam, opleiding en religie*
- Diverse achtergrondkenmerken bv *leeftijd, regio, WOZ-waarde, uitkering, herkomst, etc.*



# Analyse

## *Doel:*

- Vergelijken verschillende respons propensity methoden
- Vergelijking reguliere GREG-schatter en andere methoden

## *Twee verschillende modellen:*

Weegmodel (relatie Y en R; Schouten, 2004)

$$\text{Leeftijd}_{15} + \text{WOZ}_{14} + \% \text{ allochtonen}_8 + \text{herkomst}_7 + \text{regio}_{15} + \text{type hh}_4 + \text{telefoon}_2 \quad (1)$$

Responsmodel (relatie R; psuedo  $R^2 = 2,2\%$ )

$$\text{Leeftijd}_{15} + \text{WOZ}_{14} + \text{stedelijkheid}_5 + \text{bgstaat}_4 + \text{herkomst}_7 + \text{regio}_{15} + \text{type hh}_4 + \text{telefoon}_2 \quad (2)$$



# Resultaten - Werkzaam

Doelvariabele	Respons gemiddelde	Propensity stratificatie		Propensity wegen		Propensity GREG	
		GREG (1)	(1)	(2)	(1)		(2)
<b>Werkzaam</b>							
>= 12 uur	52,4 (0,35)	53,7 (0,18)	53,5 (0,38)	53,7 (0,38)	53,7 (0,34)	53,8 (0,34)	53,8
werkloos	6,7 (0,18)	6,4 (0,11)	6,4 (0,17)	6,4 (0,17)	6,4 (0,17)	6,4 (0,18)	6,4
< 12 uur	40,9 (0,35)	39,9 (0,17)	40,0 (0,35)	39,9 (0,37)	39,9 (0,34)	39,8 (0,33)	39,9

➤ Geen significante verschillen tussen methoden; wel duidelijk verschil met respons gemiddelde

➤ GREG en propensity wegen met zelfde model: zelfde resultaten

➤ Propensity wegen en propensity GREG hoogste schatting werkzame beroepsbevolking

*NPSO lezingendag, 23 september 2008*



# Resultaten - Opleiding

Propensity stratificatie

Propensity wegen

Doelvariabele	Respons gemiddelde	GREG (1)	(1)	(2)	(1)	(2)	Propensity GREG
<b>Opleiding</b>							
<i>Basischool</i>	7,2 (0,18)	6,3 (0,09)	6,3 (0,20)	6,3 (0,18)	6,3 (0,18)	6,3 (0,20)	6,3
<i>MAVO</i>	12,1 (0,23)	12,0 (0,15)	12,0 (0,23)	12,0 (0,22)	12,0 (0,25)	12,0 (0,24)	12,0
<i>HAVO</i>	19,7 (0,28)	19,9 (0,18)	19,9 (0,27)	19,9 (0,26)	19,9 (0,27)	19,9 (0,29)	19,9
<i>VWO</i>	7,1 (0,18)	7,2 (0,12)	7,2 (0,17)	7,2 (0,19)	7,2 (0,18)	7,2 (0,21)	7,2
<i>MBO</i>	30,8 (0,32)	31,1 (0,21)	31,0 (0,31)	31,1 (0,32)	31,1 (0,32)	31,0 (0,32)	31,0
<i>HBO</i>	16,7 (0,26)	16,9 (0,17)	16,8 (0,26)	16,9 (0,25)	16,9 (0,25)	16,9 (0,26)	16,9
<i>WO</i>	6,3 (0,17)	6,4 (0,11)	6,4 (0,16)	6,5 (0,18)	6,4 (0,19)	6,5 (0,18)	6,5
<i>andere</i>	0,2 (0,03)	0,2 (0,02)	0,2 (0,03)	0,2 (0,03)	0,2 (0,03)	0,2 (0,03)	0,2



# Resultaten - Religie

Doelvariabele	Respons gemiddelde	Propensity stratificatie			Propensity wegen		Propensity GREG
		GREG (1)	(1)	(2)	(1)	(2)	
<b>Religie</b>							
geen	37,7 (0,34)	38,5 (0,21)	38,3 (0,35)	38,4 (0,34)	38,4 (0,37)	38,5 (0,37)	38,5
R.K.	33,5 (0,33)	32,4 (0,19)	32,7 (0,34)	32,6 (0,35)	32,5 (0,35)	32,5 (0,35)	32,5
Protestant	21,0 (0,29)	20,4 (0,17)	20,6 (0,28)	20,5 (0,30)	20,5 (0,35)	20,4 (0,29)	20,4
Islam	2,5 (0,11)	3,3 (0,06)	3,1 (0,18)	3,1 (0,11)	3,2 (0,12)	3,2 (0,12)	3,2
anders	5,2 (0,16)	5,4 (0,01)	5,4 (0,16)	5,3 (0,16)	5,4 (0,15)	5,4 (0,15)	5,4

➤ Meer verschillen tussen methoden

➤ Propensity wegen en propensity GREG hoogste schatting voor geen religie

*NPSO lezingendag, 23 september 2008*



# Conclusies

## *Kanttekeningen:*

- Klein verschil variabelen in weegmodel en respons model
- Lage verklaarde variantie respons model

## *Conclusies:*

1. Onderling verschil methoden klein
2. Geen variantie reductie door directe respons propensity methoden; wel door GREG-schatter
3. Propensity wegen en GREG-schatter met zelfde model geeft zelfde resultaten voor *werkzaam* en *opleiding*



## Discussie

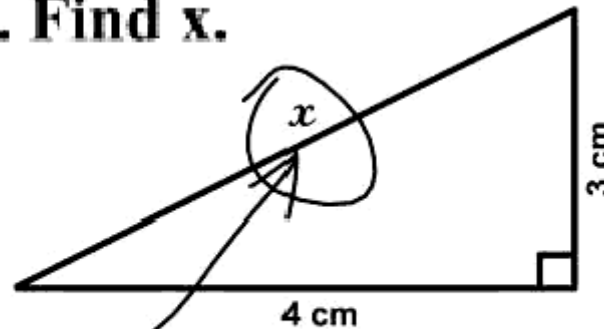
- Calibratie eigenschap GREG-schatter ontbreekt bij directe respons propensity methoden
- Wat voor variabelen in respons model? Ook beschrijven relatie met  $Y$ ?
- ...



Bedankt voor je aandacht!

Vragen?

3. Find  $x$ .

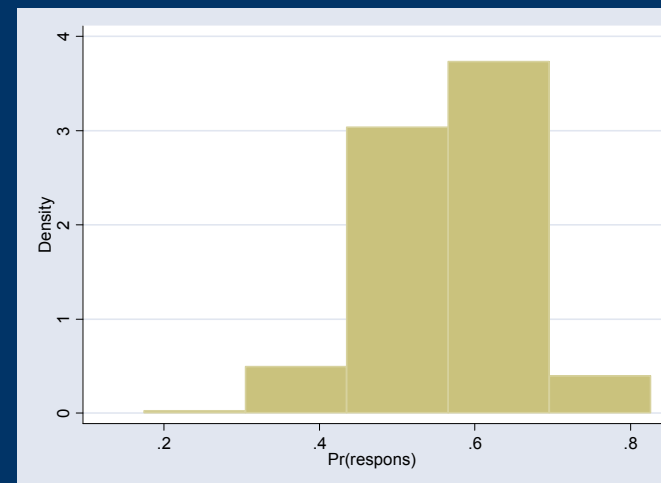
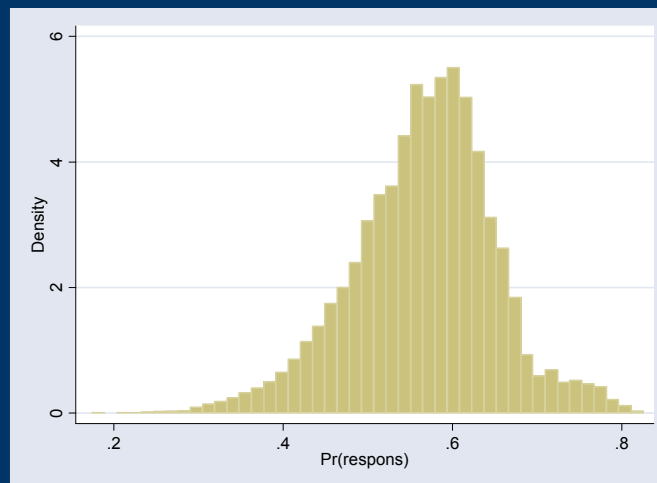


*Here it is*



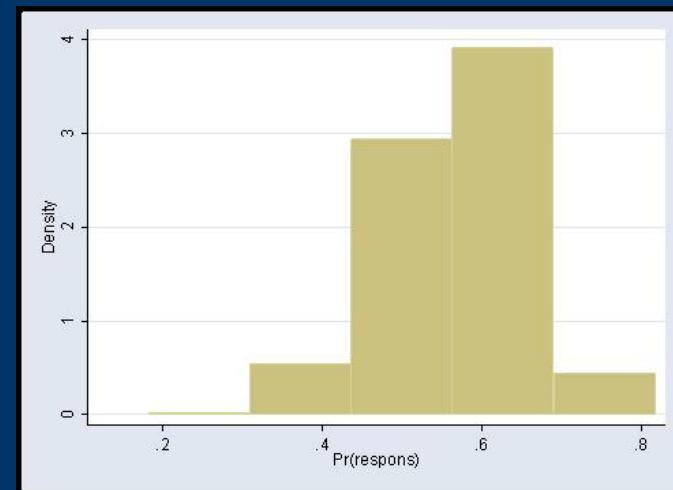
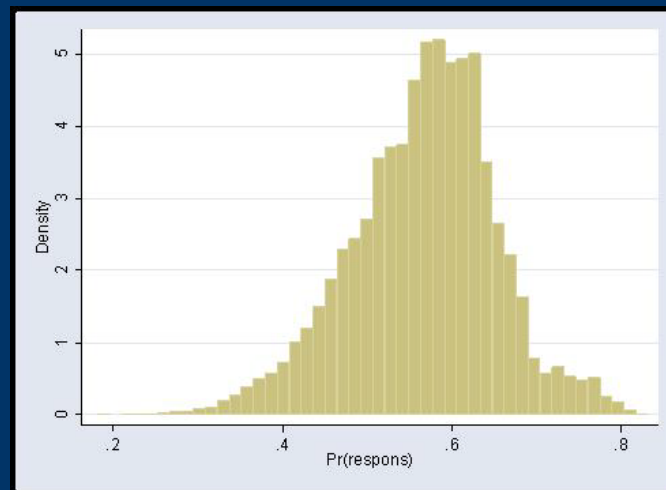
# Stratificatie respons propensities - 1

Model (1)



# Stratificatie respons propensities - 2

Model (2)



# Stratificatie respons propensities - 3

- Respons propensities geschat met logit model
- *Anders:* probit
  
- Stratificatie op basis van respons propensities
- *Anders:* Stratificatie op basis van aantallen respondenten (even grote groepen)



# POLS 2002

<i>Variabele</i>	<i>Categoriën</i>
<i>Achtergrond kenmerken</i>	
Stedelijkheid	Erg sterk, sterk, gemiddeld, nauwelijks, niet
Regio	12 provincies plus 4 grote steden: Utrecht, Den Haag, Rotterdam en Amsterdam
Telefoon	Ja, nee
Leeftijd	12-14, 14-16, ... , 70-74, 74+
Herkomst	Autochtoon, Marokko, Turkijke, Suriname, NL Antillen/Aruba, anders Westers allochtoon, anders niet-Westers allochtoon
Burgerlijke staat	Niet getrouwd, getrouwd/partner, weduwe, gescheiden
AOW	Ja, nee
ABW	Ja, nee
Type huishouden	Alleenstaand, stel zonder kinderen, stel met kinderen, alleenstaande ouder, anders
Gemiddelde WOZ	Missing, 0, 0–50.000, 50.000–75.000, ..., 275.000–300.000, 300.000–350.000, > 350.000 (euro)
% allochtonen	0–5, 5–10, ..., 40–50, 50+
<i>Doelvariabelen</i>	
Werkzaam	>= 12 uur, < 12 uur, werkloos
Opleiding	Basis, MAVO, HAVO, VWO, MBO, HBO, Universiteit, anders
Religie	Geen, Rooms-katholiek, Protestants, Islam, anders

NPSO lezingendag, 23 september 2008

