NonResponse in Household Expenditure Surveys

by

Thomas Laitila

Statistics Sweden and Örebro University

Presentation at NSM 2022, Reykjavik, August 23

NSM 2022

The NonResponse Problem

- Very problematic case
  - Pr(Responding)=f(**study variable**, auxiliary variables)
  - Data on study variable missing for nonrespondents

- Problematic case
  - Pr(Responding)=f(auxiliary variables)
  - Finding the ”correct” auxiliary variables ?

Prospective Studies

- Retrospective studies
  - Respondents are to respond on past events
  - Participation may be decided on the past events to be reported
  - The very problematic case

- Prospective studies
  - Respondents are to report future events
  - Decision to participate cannot be based on events not yet realized
  - The problematic case ☺ !!

Traditional HES

- Repondents are recruited
  - Those accepting keeps a diary of purcheses for some coming weeks
- HES are prospective
- Adjustment for participation decliners can be made
  using variables explaining the choice
- Variables explaining choice to participate = ?

Variables explaining choice (accept/decline participation)

- They are not known
- A choice made is a behavioural action
- Behavioral theory explains choices made by people
  - A respondent accepts participation if

    Utility(Accept) > Utility(Decline)
- Use choice theory in finding appropriate variables for adjustment of HES data

Economic Choice Theory

- Economic utility function: $U=u(C,L,R)$
  - C = Consumption, L = Leisure time, R = Response indicator (1/0)
- If choice is responding (R=1)
    - Time required for responding, t,
      is drawn from total available time, T,
      leaving less time to allocate between Work and Leisure

- Optimize U with R=1, available time = T-t gives U_1
- Optimize U with R=0, available time = T gives U_0

- Choose to respond (R=1) if $U_1 > U_0$

Dichotomous-Choice (DC) model

- From the theory a DC model is derived for single living households where

$$\Pr(\text{Accept from hh } k) = f(a_k + b_k \cdot z_k)$$

where
z_k = a derived measure of cost of responding
a_k = utility obtained from responding, excl. costs
b_k = valuation of the cost

Table A:
Probit ML estimates of DC model for Pr(Response)

Data from Statistics Sweden 2007 HES. Single living with or without children.

Variables	Estimate	St.Err
Age	.142	.027
Log(Disposable Income)	.284	.102
Log(Age)	-6.17	1.35
D_k	95.8	40.5
D_k Age	-.713	.205
D_kLog(Disposable Income)	-.202	.106
D_k Log(Age)	-65.6	25.9
D_kLog(Disposable Income)^2	13.0	4.64
z_k	108	56.2
z_k Log(N:o persons)	-10.8	2.77
z_k Log(Age)	-29.3	14.8

Figure A: Plot of estimated response probabilities vs Age

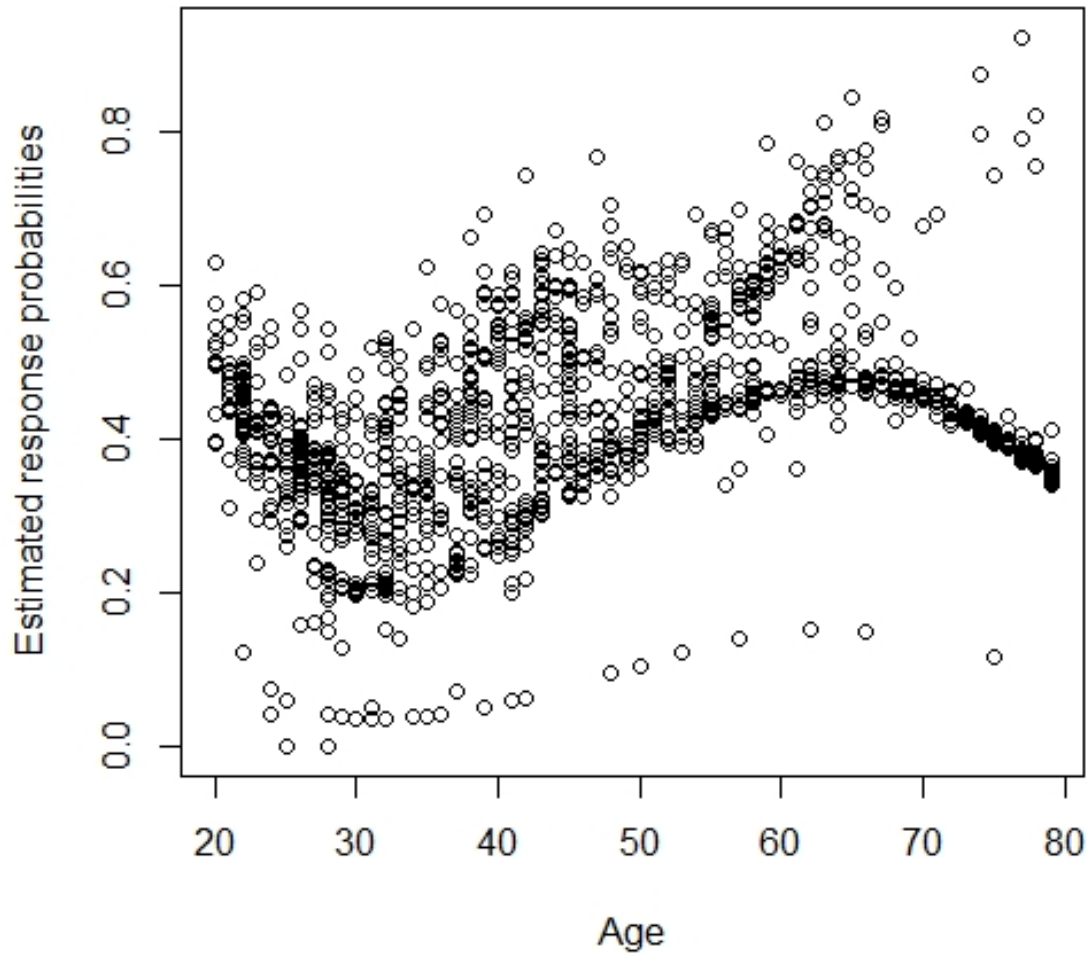


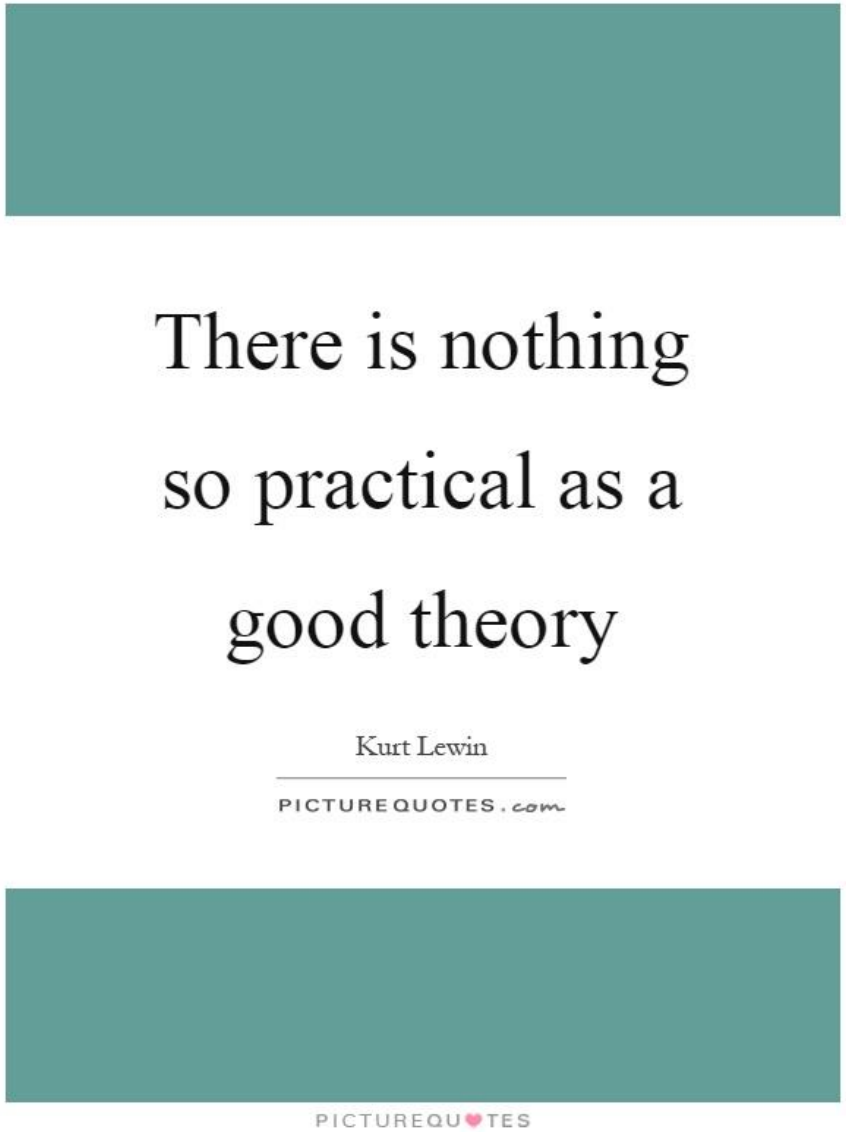
Table B:

	Single with children			Single without children	
Expenditures	DC model	HUT 2007		DC model	HUT 2007
Total	234 535	229 290 ±17 100		168 595	167 540 ±9 910
Food	29 232	28 310 ±2 360		16 508	17 280 ±1 080
Clothes/shoes	11 899	11 590 ±2 640		9 172	8 230 ±1 670
Healthcare	4 140	4 060 ±1 190		4 282	4 020 ±1 270

Benefits of the DC approach

- Estimates rest on sound scientific and theoretical arguments

- Theory provides with guidance on auxiliary variables to include



Design of HES

- Drastically reduce the response burden
  - ”Split questionnaires” with overlaps
  - Shorter measurement periods
  - Simplify what to record in diary
  - etc.

- Revise sampling design
  - Make face-to-face-interviews feasible
  - Kluster sampling
    - Geographical areas with low response rates

Design of HES cont'd

- Use DC approach to handle nonresponse in the recruiting stage
- Handle nonresponse due to drop-outs/attrition with
  - double sampling
  - DC modelling
- Make participation interesting
  - Payments
  - Information feedback

Also in the paper

- Example where errouneous auxiliary variables are introducing bias in estimates

- Range of observations perhaps more important than response rates (to be added)

Thanks for listening

Contact – thomas.laitila@scb.se