Improving process efficiency in panel surveys with paradata

Frauke Kreuter

July 4, 2012

Acknowledgement

- Co-authors
  - Gerrit Müller, Institute for Employment Research (IAB), Germany
  - Gabriele Durrant, University of Southampton
- Implementation of experiment in collaboration with
  Mark Trappmann, IAB Germany, (PASS Program Director)
  Infas, Germany (Data Collector)
- Research support: ESRC Research Grant RES-062-23-2997
  "The Use of Paradata (Field Process Data) in Cross-Sectional
  and Longitudinal Surveys" (Durrant, Kreuter & Smith)
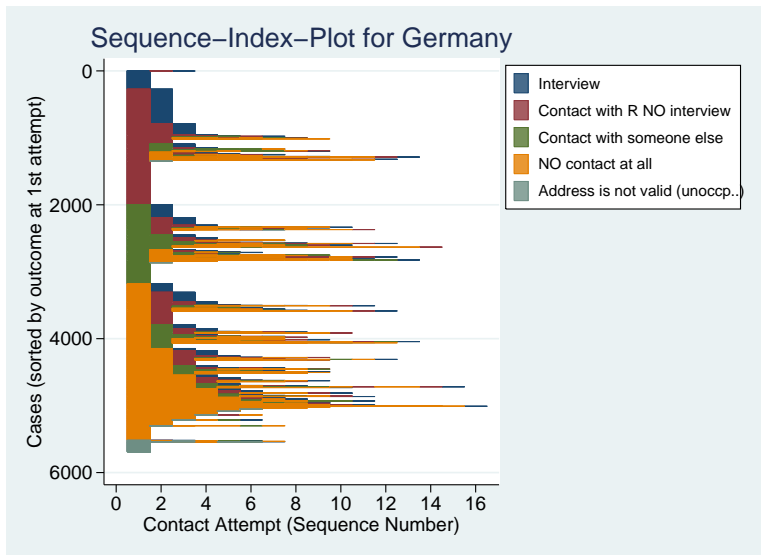
Outline

Contact Protocol Data: Example from ESS

VISIT RECORD *(Visit = every attempt made to reach the respondent/ household)*

Visit No	1. Date dd/mm	2. Day of the week	3. Time 24 hr clock	4. Mode of visit 1 = personal visit 2 = telephone 3 =personal visit, but only intercom 4 = info through survey organisation	5. RESULTS of the visit 1= Completed interview 2= Partial Interview 3 = Contact with someone, Target Respondent not yet selected 4 = Contact with Target Respondent but NO interview 5 = Contact with somebody other than Target Respondent 6 = No contact at all 7 = Address is not valid (unoccupied, demolished, institutional,…)
1	/		:		
2	/		:		
3	/		:		

Problem: Unproductive "Calls" Kreuter & Kohler (2009)

Best Call Windows Wagner (2012)

- Example from the U.S. National Survey of Family Growth
- Heatmap reflecting best times to contact all and eligible cases (age 14-45)

hour	Sunday		Monday		Tuesday		Wednesday		Thursday		Friday		Saturday	
	All	Elig	All	Elig	All	Elig	All	Elig	All	Elig	All	Elig	All	Elig
9	0.30	0.34	0.25	0.35	0.30	0.32	0.24	0.31	0.23	0.30	0.27	0.33	0.30	0.35
10	0.32	0.40	0.31	0.38	0.28	0.33	0.29	0.34	0.30	0.36	0.27	0.34	0.31	0.39
11	0.36	0.43	0.30	0.38	0.31	0.38	0.31	0.39	0.31	0.39	0.32	0.40	0.35	0.43
12	0.37	0.44	0.32	0.42	0.32	0.38	0.32	0.40	0.30	0.37	0.31	0.38	0.34	0.42
13	0.37	0.45	0.32	0.42	0.24	0.31	0.29	0.38	0.30	0.38	0.32	0.39	0.34	0.43
14	0.38	0.46	0.34	0.44	0.33	0.40	0.32	0.40	0.32	0.39	0.33	0.40	0.35	0.43
15	0.39	0.48	0.35	0.44	0.32	0.40	0.33	0.42	0.33	0.41	0.33	0.41	0.36	0.46
16	0.39	0.49	0.36	0.45	0.37	0.46	0.36	0.45	0.35	0.43	0.34	0.42	0.35	0.45
17	0.39	0.49	0.40	0.49	0.38	0.46	0.38	0.47	0.36	0.46	0.34	0.43	0.33	0.43
18	0.37	0.44	0.38	0.47	0.39	0.48	0.37	0.47	0.36	0.45	0.33	0.42	0.35	0.44
19	0.37	0.44	0.39	0.47	0.37	0.45	0.37	0.46	0.35	0.44	0.31	0.42	0.35	0.43
20	0.40	0.44	0.38	0.45	0.39	0.45	0.38	0.46	0.37	0.45	0.32	0.40	0.36	0.44

State of the Art

- "Best call times" vary by subgroups

  e.g. Durrant, D'Arrigo & Steele 2012

- Covariate information should be used

  e.g. Wagner 2012

- Panel surveys
  - have a variety of covariates from prior waves but also
  - paradata about effective call times in prior waves

    e.g. Lundquist 2009, Lipps 2012

Outline

Opportunity in Panel Surveys

- Can we identify a simple predictor of "best call" times in panel surveys?

- Does efficiency (time to first contact) increase if cases are called at the "best time"?

- Are cases called at "best times" more likely to participate?

Outline

1. Background: Paradata from Contact Protocols

2. Research Questions

3. Data and Methods

4. Experimental Results

PASS - Panel 'Labour Market and Social Security'

- Since 2006 annual household survey conducted by the Institute for Employment Research (IAB)
- Multi-frame survey
  - Administrative data frame of benefit recipients
  - Register sample for general population
- Sequential mixed-mode design
  - CATI-CAPI
- Here: observational data from
  Wave 4 (n=6000) and
  Wave 5 (n=5508)
- Experiment done in Wave 6 (n=4060)

Distribution of calls across windows

- Call times misaligned with best interview times
- Distributions very similar across waves
- Conveniently reached vs. convenient interview time

Call Window	1st Calls	1st Contact	Interview
Weekday 0:00-12:00	25.93	29.14	27.57
Weekday 12:01-17:00	48.02	37.34	29.49
Weekday 17:01-0:00	15.27	13.24	25.73
Weekend 0:00-12:00	0.77	6.59	7.62
Weekend 12:01-0:00	10.02	13.68	9.59

"Best window" from prior wave

- Linear probability model (for ease of interpretation)
- Taylor-linearized variance estimation
- Y: Probability of successful interview at 1st attempt
- X: Same window as . . .

| | 1st contact in wave $t-1$ | | interview in wave $t-1$ | |
	β	SE	β	SE
Coefficient	0.032	0.02	0.046	0.015
Constant	0.218	0.007	0.211	0.007

Experimental Design

- Wave 5 interview time $\Rightarrow$ Wave 6 first three attempts
- Three windows specified for each day
- Wave 6 panel cases
  80% treatment assignment ($Z = 1$) and 20% control ($Z = 0$)
- Analysis:

  Intention to Treat: $\theta_{ITT} = E(Y|Z = 1) - E(Y|Z = 0)$

  Local Average Treatment Effect: $\theta_{LATE} = \frac{E(Y|Z=1)-E(Y|Z=0)}{E(D|Z=1)-E(D|Z=0)}$

Outline

Positive Effect on Efficiency

- Number of contact attempts to first contact
  Average = 3.693
  ITT -0.3323 (0.1972)
  LATE -0.5852 (0.3409)

- Probability – interview at first contact
  Average = 0.193 (or 19.3%)
  ITT 0.0078 (0.0156)
  LATE 0.0139 (0.0275)

Summary and Challenges

- Interview date at t-1 "better window" than contact at t-1
- Efficiency (time to first contact) does increase when cases are first called at the "best time"
- No gain in response rates through experimental design
- However cost savings can scale up
  (think 10 Euros per call and 10.000 cases)

- Models ignore interviewer assignments
- Models ignore covariate interaction
- Call schedulers often do not allow flexible programming