

Using paradata to inform best times of contact

A multilevel event history modelling approach

Gabriele Durrant, Julia D'Arrigo and Fiona Steele

1 July 2009, ESRA conference, Warsaw

Outline

- Motivation and Background
- Data
- Methodology
- Results
- Summary
- Implications for Survey Practice

3-Year Research Project

- “Hierarchical Analysis of Unit Nonresponse in Sample Surveys”
- Funded by Economic and Social Research Council (ESRC), UK
- Team: University of Southampton, University of Bristol (UK)
- Collaborators: Netherlands, US

Motivation and Background

Motivation

- Response rates falling; implications for data quality
- Establishing contact important part of response process in face-to-face surveys (costly and time consuming)
- Necessary: better understanding of contact process and interviewer calling behaviour

Research Hypotheses

1. What are **best times** of the day and days of the week to establish contact?
2. What are best times to establish contact with certain types of households? (**interactions**)
3. To what extent does establishing contact and the success of the timing of the call depend on **interviewer characteristics**?

Needed: information about each call, household, interviewer, ...

Background

- Collection of field process data (**paradata**)
- Hope for cost savings, increases in response rates, ...
- Examples:
 - Information about each call to the household made by interviewer (date, time, ...)
 - Interviewer observation data about each household
 - Information about interviewers

Background

- Paradata increasingly used:
 - for progress updates, as decision making tools
 - **responsive survey designs** (continuous monitoring of progress during data collection, alterations of design to reduce costs and increase data quality)
 - for adjustment
- Often unclear how best to use paradata
- methodological development needed in specification of models using paradata

Data

Data

- Interviewer call record data (time variant)
 - Date and time of call, time between calls, contact strategy used, outcome of call, ...
- Interviewer observations about each household (time invariant)
 - Type of accommodation, physical barriers, security device, indications if children present, observations about neighbourhood, ...
- for six UK government surveys (LFS, EFS, OMN, NTS, FRS, GHS)

Data

These linked to:

- Information about each household from UK 2001 Census
 - Socio-demographic information; type of household,...
- Information about interviewers
 - Experience, pay grade, qualifications, demographic information, workload and planning, interviewer strategies, attitudes and behaviours, ...
- Area information

The Linked Data

Contact	R=1	Call record data	Inter- viewer Obser- vation	Household Information from Census	Inter- viewer Attitude Survey	Area Infor- mation
	R=0					

17000 hhs 565 interv. 392 areas

Advantages of this study

- Rich paradata available (call records; interviewers,...)
- From six UK surveys
- Rich data available on both responding and nonresponding households
- Models account for hierarchical structure of data (households, interviewers, areas)
- Allow for interviewer influences
- Focus on process leading to contact rather than only final outcome (noncontact)

Methodology

Multilevel discrete-time event history model

- probability of contact at a particular call, given that no contact prior to that call
- HH not contacted by end of data collection period are right-censored

Multilevel event history model

$$\log \left(\frac{\pi_{i(jk)t}}{1 - \pi_{i(jk)t}} \right) = \alpha_t + \beta' \mathbf{x}_{i(jk)t} + u_j + v_k$$

i =households, j =interviewers, k =areas, t =time

$y_{i(jk)t} = 1$ if contact, 0 otherwise

$$\pi_{i(jk)t} = \Pr(y_{i(jk)t} = 1 \mid y_{i(jk)t-1} = 0)$$

$$u_j \sim N(0, \sigma_u^2) \quad v_k \sim N(0, \sigma_v^2)$$

Explanatory variables

- Factors **under control of interviewer** (time varying) (time of call, time between calls, contact strategies); also interviewer characteristics and strategies
- Factors **outside that control** (time invariant) (HH characteristics, proxies for time at home, lifestyle; physical barriers)
- **Interactions** between time varying and time invariant variables: best times of contact for certain types of households
- Informs allocation of time and staff resources

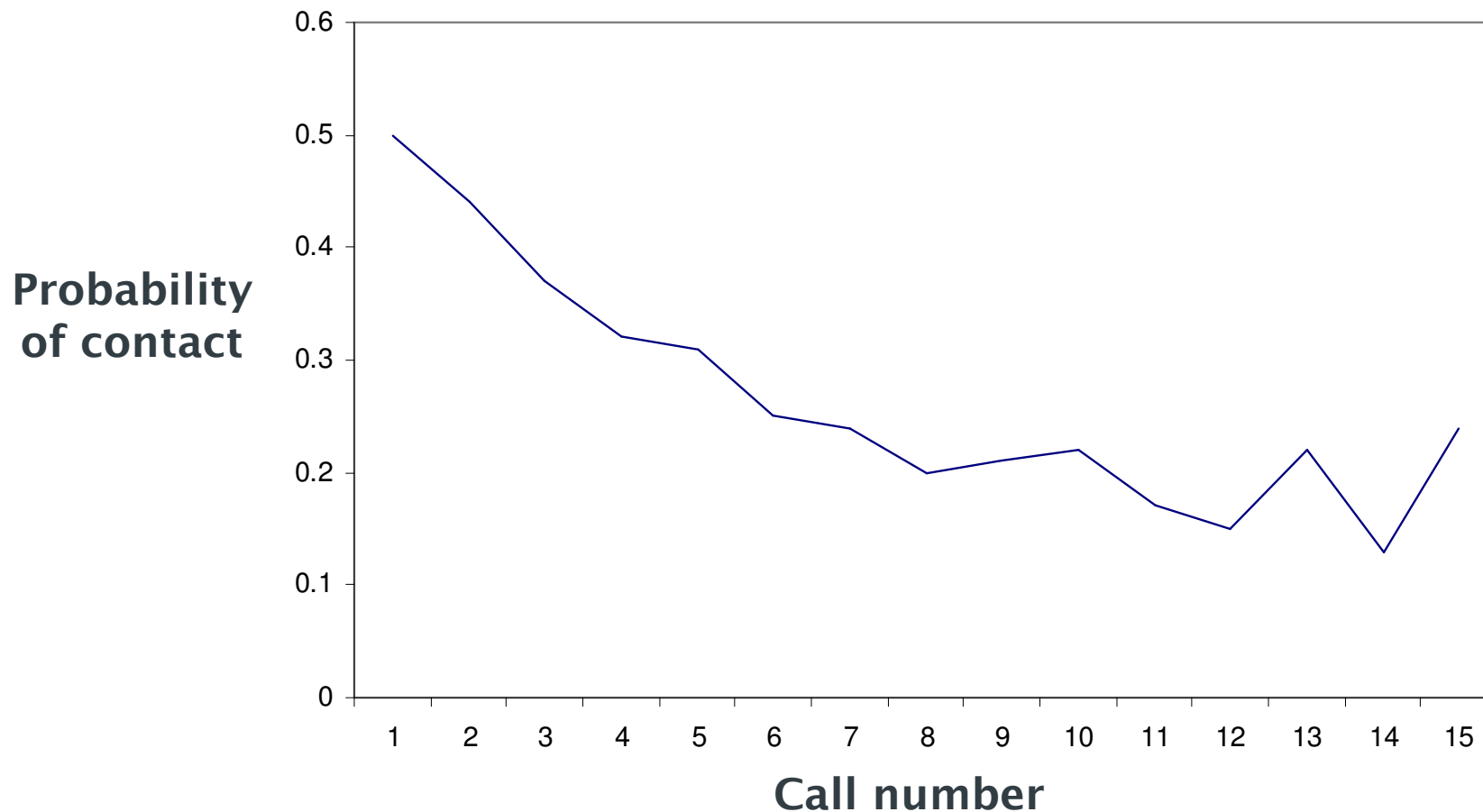
UNIVERSITY OF
Southampton

Southampton Statistical
Sciences Research Institute

Results

Descriptive statistics: hazard rate of contact

- Monotonic decline in contact rate as the number of calls increases



Results:

Variable (0 = Reference category)	Categories	$\hat{\beta}$ ($ste(\hat{\beta})$)
Time variant variables		
Previous call indicator (0 = First call)	1 Call previously made	-0.550 (0.060)***
Number of calls previously made		-0.110 (0.009)***
Day and time of call † (0 = Sun, Mon, Tue, Wed evening)	1 Mon, Tue, Wed morning 2 Mon, Tue, Wed afternoon 3 Thur, Fri morning 4 Thur, Fri afternoon 5 Thur, Fri evening 6 Sat, Sun morning 7 Sat, Sun afternoon 8 Sat evening	-0.278 (0.186) -0.443 (0.109)*** -1.125 (0.283)*** -0.599 (0.143)*** -0.097 (0.147) -0.298 (0.521) -0.348 (0.317) -2.383 (1.399)*
Time of previous call † (0 = Weekday evening)	1 Weekend 2 Weekday morning 3 Weekday afternoon	0.623 (0.140)*** -0.025 (0.103) 0.169 (0.051)***
Number of days between calls (0 = Same day)	1 1-3 days 2 4-8 days 3 9-14 days 4 15+ days	0.097 (0.042)** 0.246 (0.045)*** 0.311 (0.080)*** 0.287 (0.155)*

Household characteristics

Interaction between day and time of call and dependent children in household			
		Dependent children present	
		Present	Not present
Day and time of call	Mon, Tue, Wed morning	55.5	43.6
	Mon, Tue, Wed afternoon	56.9	39.7
	Sun, Mon, Tue, Wed evening	63.2	50.4
	Thu, Fri morning	34.7	25.2
	Thu, Fri afternoon	51.2	36.1
	Thu, Fri evening	57.6	48.0
	Sat, Sun morning	42.3	43.1
	Sat, Sun afternoon	52.6	41.9
	Sat evening	12.4	9.7

Interaction: Household characteristics

Interaction between day and time of call and adults in employment				
		Adults in employment		
		No adult	One adult	Two or more adults
Day and time of call	Mon, Tue, Wed morning	51.2	44.4	36.0
	Mon, Tue, Wed afternoon	49.2	37.7	33.8
	Sun, Mon, Tue, Wed evening	57.9	59.2	63.4
	Thu, Fri morning	32.4	32.0	32.6
	Thu, Fri afternoon	45.4	34.9	29.6
	Thu, Fri evening	55.6	58.1	60.6
	Sat, Sun morning	50.7	45.5	43.2
	Sat, Sun afternoon	49.5	52.4	51.4
	Sat evening	12.8	60.3	71.1

Household characteristics (census and IO)

Variable (0 = Reference category)	Categories	$\hat{\beta}$ ($ste(\hat{\beta})$)
Household-level variables (time invariant)		
Age (HRP) (0 = 16 - 34)	1 35 - 49	0.172 (0.034)***
	2 50 - 64	0.398 (0.038)***
	3 65 - 79	0.454 (0.069)***
	4 80 and older	0.550 (0.080)***
Household type (0 = Single household)	1 Couple household	0.436 (0.030)***
	2 Multiple household	0.404 (0.076)***
Interviewer Observations (time invariant)		
Security device (0 = security device visible)	1 No security device visible	0.193 (0.031)***
Houses in area in good or bad state of repair † (0 = Good)	1 Fair-Bad	-0.166 (0.051)***
House in a better or worse condition than others in area (0 = Better)	1 About the same	-0.069 (0.039)*
	2 Worse	-0.273 (0.055)***

Interviewer and area variance

- Inclusion of interviewer characteristics reduces interviewer variance
- Area variance halved after controlling for household and call level variables
- Area level variables no longer significant

Interviewer characteristics

Interviewer-level Variables (time invariant)		
Pay grade (0 = Merit 1 and 2)	1 Interviewer and advanced interviewer 2 Merit 3 and field manager	0.081 (0.047)* 0.129 (0.056)**
Interviewer qualification (0 = Degree or postgraduate, other higher education)	1 A levels 2 GCSE, qualifications below this level, no qualification	-0.144 (0.058)** -0.129 (0.078)*
Wait to explain survey rather than leaving behind information (0 = Always, frequently, sometimes)	1 Rarely 2 Never	-0.157 (0.055)*** 0.069 (0.071)
Use phone to make appointment (0 = Always, frequently, sometimes)	1 Rarely, never	0.106 (0.041)***

Summary

Summary

- Weekday evenings and weekend daytimes are on average best times to call; however, best times to call depend on household characteristics
- Interviewer observations useful for predicting best times of contact
- Area information useful in absence of other information
- Significant effects of interviewer characteristics
- Some significant effect of interviewer strategies

Implications for Survey Practice

Implications for survey practice

- May inform design of efficient and effective calling behaviours, follow-ups, responsive survey designs
- Predict contact at future calls based on data collected at previous calls, IO data, admin data
- Best times of contact depends on household type
- Survey agencies need to allocate time and staff resources: not all calls in evening/weekends
- Guidance on how best to use paradata (methodological development); what data to collect, ...

Working paper available:

Durrant, G.B., D'Arrigo, J. and Steele, F. (2009): Using field process data to predict best times of contact conditioning on household and interviewer influences.

Any questions or comments:

g.durrant@southampton.ac.uk

Thank you