Improving Efficiency in Panel Surveys with Paradata

-PRELIMINARY RESULTS-
Paper presented at the International Workshop on Household Survey Nonresponse 2012

by Gerrit Müller, Institute for Employment Research (IAB)
Co-authors: Frauke Kreuter, University of Maryland & IAB/LMU
and Gabriele Durrant, University of Southhampton

August 10, 2012

1 INTRODUCTION

Call scheduling is an important problem for surveys around the world, and years of research on improving call schedules have not shown the expected success (Wagner, 2012). It is generally known that a variation in calling times reduces the number of attempts needed to reach sample units (Cunningham et al., 2003). And it is not hard to imagine that the likelihood to agree to be interviewed is higher if the call happens at a time convenient for the household. The Public Works and Government Services Canada (2007) rightly point out that "calling during the dinner hour may increase the likelihood of reaching people at home, but it may decrease the likelihood of participation because some people find these calls intrusive. If the ideal time to reach people at home is also the same time of day they are the least likely to agree to participate, then consider alternative call times. This strategy could be as simple as beginning calls earlier in the afternoon, perhaps at 3:00 p.m. rather than 4:00 p.m. A good contact rate is insufficient on its own. People must also agree to participate in an interview."

For the general population a few patterns have emerged repeatedly over the years, for example afternoons are more efficient than mornings at least on weekdays (Weeks et al., 1987), whereas weekend mornings are "better" than weekend evenings for the general population. These findings have found their way into advanced call scheduling algorithms (Brick et al., 1996), which often field a case first in the evening, wait a few days before the next call, and ensure that future calls are placed at other times.

The predictions of best times to call can be improved when covariate information about a case is known. Durrant et al. (2011) for example showed across six U.K. household surveys that Saturday evenings are more beneficial when pensioners are present in the household than when they are not, and Wagner (2012) showed for the U.S. National Survey of Family Growth that persons between age 15-45 have the highest contact probabilities when called in the late afternoon on Sundays and Monday evenings. Unfortunately covariate information on a case base level is not available in most cross-sectional telephone surveys (Smith, 2011). Panel surveys are in the fortunate situation to have ample covariates from prior panel waves. These variables can be used to predict best contact times for the subsequent wave. First attempts to use sample characteristics in call scheduling have been made by Wagner (2012).

Panel surveys also have much more direct measures of best calling times: prior waves call history information. Lipps (2012) showed in a post-hoc analysis of Swiss Household Panel data that the probability of contact and more importantly cooperation at first contact increases if respondents are contacted during the call window that had been successful in the prior wave. This effect holds even after controlling for respondent characteristics. Thus in addition to predicting best call windows through

covariates, the direct measure of the "best" window for a given household can be extracted from prior wave paradata. There is some charm to using this information directly, instead of model based call windows or entire call algorithms. Data collectors are often "owner" of the paradata, but do not have easy access to respondent information. Also, some call schedulers are hard to program and specific programming efforts are needed to employ individualized call scheduling algorithms (Wagner, 2012). Thus a simple heuristic – such as "start fielding a case during the successful prior wave window" – might be easier to implement and already provide some efficiency gain. In fact, (Lipps, 2012, 14) suggests that "households should be called more often during the same time (window) at which it was first contacted in the previous wave, especially at first call", because there "seems to be a tendency that household-specific preferable calling and contact times persit across years."

In line with this suggestion we changed fieldwork efforts in a panel study in Germany. Here we report the results from a randomized experiment that used prior wave call information in a subsequent wave of data collection. To our knowledge this kind of information has never been used experimentally as sole intervention in a CATI panel survey. Experimental efforts in the Swedish Labour Force Survey prioritized certain call sequences for certain subgroups of the population, but changed the contact strategies of the experimental group also in several other ways preventing us from singling out the causal effect of using the same window at the next wave. Wagner (2012) built adaptive models based on auxiliary variables and within survey contact information for repeated cross sectional surveys and experimentally assigned best call windows.

2 DATA

The survey data used in this paper come from the German Panel study "Labour Market and Social Security" (PASS). Since 2006 PASS data have been collected annually by the German Institute for Employment Research at the Federal Employment Agency (Trappmann et al., 2010). After wave three PASS changed the data collection agency responsible for the field work (Müller, 2011). For this paper we focus on wave 4 through 6 of PASS data collection to avoid any possible confounders due to the switch in data collection procedures. PASS is a dual-frame mixed-mode (CATI and CAPI) survey, though we focus solely on the CATI portion of PASS, for call times in face-to-face interviews always reflect availabilities and preferences of the interviewers whereas in centralized telephone studies calls are made and assigned to interviewers around the clock.

2.1 Post-hoc Analysis of Wave 4 and Wave 5 Data

Similar to the Lipps analysis, we examined the effect of prior (wave 4) call window on the probability of interview at first contact in wave 5. To avoid small sample sizes we grouped cases into time slots and separated them by weekday and weekend, leading to a distribution for wave 4 shown in the Table 1. Wave 5 showed a similar distribution.

Table 1: Distribution of Wave 4 Calling Times

	1st Call	1st Contact	Interview
Weekday morning 0:00-12:00	25.93	29.14	27.57
Weekday afternoon 12:01-17:00	48.02	37.34	29.49
Weekday evening 17:01-0:00	15.27	13.24	25.73
Weekend morning 0:00-12:00	0.77	6.59	7.62
Weekend afertnoon & evening 12:01-0:00	10.02	13.68	9.59
N	6,000	5,656	4,462

Based on these data, we generated two variables: one indicating that the first contact in wave 5 was in the same time window as the first contact in wave 4, and another one indicating that the first contact

Table 2: Distribution of Wave 5 Calling Times

	1st Call	1st Contact	Interview
Weekday morning 0:00-12:00	23.67	23.44	26.77
Weekday afternoon 12:01-17:00	43.92	36.95	27.07
Weekday evening 17:01-0:00	20.79	19.07	28.56
Weekend morning 0:00-12:00	1.74	5.40	7.04
Weekend afternoon& evening 12:01-0:00	9.88	15.14	10.56
N	5,508	5,039	4,034

in wave 5 was in the same time window as the interview in wave 4. Using these two variables, the effect on cooperation (at first contact) in wave 5 is examined[1].

Table 3: Effect of "Same Time Window" as in Previous Wave on Cooperation (at 1st Contact) in Wave 5

	1st Contact in Wave 4		Interview in Wave 4	
	β	SE	β	SE
Coefficient	0.032	0.02	0.046	0.015
Constant	0.218	0.007	0.211	0.007

Similar to Lipps (2012), Table 4 below shows a positive effect of calling at a successful window from last wave on the probability of gaining cooperation at first contact. However we see a stronger effect using the successful interview window from last wave in Table 5 compared to the successful first contact window (which was used in Lipps (2012)).

Table 4: Cooperation at 1st Contact, Wave 5

	Mean	Time Window = 1st Contact Wave 4	Std. Err.
Pr Interview	.222	+.032**	(.020)

Table 5: Cooperation at 1st Contact, Wave 5

	Mean	Time Window = Interview Wave 4	Std. Err.
Pr Interview	.222	+.046***	(.015)

In preparation for the experimental design we use the latter logic to for the experimental manipulation of the call scheduling algorithm. However, while the same time windows were used, the experiment did not use the rough categorization into weekday and weekend but the actual days of the week.

3 Experimental design

All panel cases in Wave 6 that were also respondents in Wave 5 were eligible for our experiment. Within strata, 80% of the panel cases were randomly assigned to the treatment group and 20% were randomly assigned to the control group. The call scheduler was programmed so that calls to the treatment group were first made at the same weekday and in the same time window as those cases which were interviewed in Wave 5.

[1]The dual-frame mixed mode design used in PASS has even telephone cases to be clustered geographically. We therefore use Taylor-linearized variance estimation to take the clustering into PSUs into account.

Three time windows were specified for each workday, matching the time slots shown earlier. Cases that were interviewed on a Saturday or Sunday in Wave 5 were all assigned to the two Saturday call windows; since the call center was closed on Sundays in Wave 6. If the contact attempt at the pre-specified call window was unsuccessful, the next call was scheduled for the next week at the same day during the same window. After three unsuccessful contact attempts treatment cases were switched back to the standard protocoll and called in similar fashion as the control cases.

The control group algorithm randomly assigned a starting window to a case, if a case can not be contacted during that window, it moves to the next window. The minimum time before a case is called again is 240 Minutes, with the exception of 1) busy cases which were called again within the next 15 minutes, 2) cases in which an appointment was made, and 3) cases that had a terminating status during the call (ineligible, refusal etc.). This algorithm is commonly used in the PASS data collection agency. Fieldwork for Wave 6 started February 14th and is still ongoing at the time of writing.

4 Analysis

In our analysis we focus on two outcome variables Y: The probability to be interviewed at the first contact as a measure of immediate cooperation, and the number of calls until first contact. Both, increased probabilities and reduced numbers of attempts would imply increased efficiency and thus potentially benefit fieldwork agencies.

To assess the effects of the experimental change in the call scheduling algorithm, we first look at these outcomes differentiating simply by assignment status to treatment and control group. In the literature on evaluation of randomized experiments this difference by assignment status is also known as an intention-to-treat (ITT) effect. Formally, the mean difference in terms of the outcome variable Y of those assigned to the treatment group ($Z = 1$) and those assigned to the control group ($Z = 0$):

$$\theta_{ITT} = E(Y|Z = 1) - E(Y|Z = 0) \tag{1}$$

Since in our context "participation" is voluntary among those randomly assigned to receive treatment, there is the issue of compliance with the experimental intervention. That is, what we have manipulated here by concentrating first calls on the same time window as previous wave's interview is just an "offer" of treatment. In many experiments – and so here – it is possible that the treatment assignment is not equal to the actual treatment intended, which would be a successful contact in the designated time window. There are several reasons for this: First, some control group cases might have been called and contacted at the same time window by mere chance. Second, individuals in the treatment group are free to ignore a call or may simply be unavailable at the assigned calling time. That is, of those offered treatment only a – potentially self-selected – subset is successfully contacted (=treated) at the designated time.

That said, the ITT effect is indeed the policy relevant parameter. It tells us the causal effect of the offer of treatment, building in the fact that many of those offered have declined (Angrist and Pischke (2009): 161-166). However, the ITT effect is too small relative to the average causal effect on those who were in fact treated. Therefore an adjustment can be made to the ITT effect that gives us a measure of the effect of treatment on the treated: dividing the θ_{ITT} by the difference in compliance rates between treatment and control groups; with $D = 1$ if the first contact was in fact in the same time window, and $D = 0$ otherwise.

$$\theta_{LATE} = \frac{E(Y|Z = 1) - E(Y|Z = 0)}{E(D|Z = 1) - E(D|Z = 0)} \tag{2}$$

In econometric parlance this is called a local average treatment effect (LATE), which is the (average) treatment effect on the subset of individuals whose treatment status is actually changed by the

treatment offer. It is equivalent to IV estimation using the randomly assigned treatment status as an instrumental variable for treatment received.

5 Results

Table 6: Comparison by Assignment Status to Treatment/Control Group; Intention-to-treat (ITT)

	n	Mean	ITT	Std. Err.
Pr Interview at 1st Contact, overall	3,667	.167	+.007	(.015)
Pr Interview at 1st Contact, 1st - 3rd Call	2,767	.222	+.008	(.019)
No. of Calls until 1st Contact, overall	3,667	3.62	-.320*	(.217)
No. of Calls until 1st Contact, 1st - 3rd Call	2,767	1.52	-.017	(.035)

Table 7: LATE; Instrumental Variable Estimates (IV)

	n	Mean	IV	Std. Err.
Pr Interview at 1st Contact, overall	3,667	.167	+.012	(.025)
Pr Interview at 1st Contact, 1st - 3rd Call	2,767	.222	+.010	(.024)
No. of Calls until 1st Contact, overall	3,667	3.62	-.517**	(.350)
No. of Calls until 1st Contact, 1st - 3rd Call	2,767	1.52	-.021	(.042)

Table 6 in this result section shows that, overall, the probability of cooperation at first contact is not increased significantly. Restricting the analysis to the first three calls – the range over which the experimental manipulation of the call scheduling was in effect in the treatment group – shows a similar result. What we do see though is that if the first contact is achieved at one of the first calls, the cooperation rate is higher than overall (22.2% as compared to 16.7%). That is, cases with a higher number of call attempts until first contact tend to be less cooperative given contact. This finding is not new, but it suggests that the general idea of trying to contact cases early, and in the right time window, might still be a good one. It is just the case that our particular experimental intervention was not successful in that regard.

Turning away from immediate cooperation towards contactability, that is the number of call attempts it took to establish the first contact, we again find no effect over the range of call attempts the experiment was in effect. We do find though a significant difference between treatment and control group cases with respect to the number of call attempts, when we look at the full sample. Analyses not shown here indicate that there is no difference between treatment and control up to about 10 calls. Therefore, the mean difference of .32 calls until first contact (which would, scaled by the mean, amount to a sizeable 10% reduction) appears to be driven by a few cases with a relatively large number of contact attempts and is, most likely, unrelated to the experimental intervention.

In line with our discussion in section 4, the ITT-effects just described are smaller than the local average causal effects on the treated, mainly due to noncompliance with the treatment. With our experimental intervention we could only shift *calls* into designated time windows, but not *contacts*. For example, in the treatment group only 62% had there first contact in the same time window (complied with the "treatment"). And, on the other hand, about 4% of the control group cases happened to be successfully contacted in the same time window just by the usual calling routines. However, table 7 shows that even the effects on those cases actually being treated are not significantly different from zero. Again, with the exception of the number of calls until first contact, in the full sample.

6 Summary and Discussion

What these "null" findings suggest is that there is no causal effect of the "same time window" on the outcomes considered here, and that the significant coefficients in models for cooperation at first contact by Lipps (2012) and in our earlier analyses for PASS might reflect mainly selection effects. That is, individuals who are contacted in that time window are different in terms of unobservable characteristics, which are not – or cannot be – controlled for in such analyses.

References

Angrist, J. and J. Pischke (2009). *Mostly Harmless Econometrics: an Empiricist's Companion*. Princeton University Press.

Brick, J., B. Allen, P. Cunningham, and D. Maklan (1996). Outcomes of a calling protocol in a telephone survey. In *Proceedings of the American Statistical Association, Survey Research Methods Section*.

Cunningham, P., D. Martin, and J. M. Brick (2003). An experiment in call scheduling. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 59–66.

Durrant, G., J. D'Arrigo, and F. Steele (2011). Using field process data to predict best times of contact conditioning on household and interviewer influences. *Journal of the Royal Statistical Society, Series A 174*(4), 1029–1049.

Lipps, O. (2012). A Note on Improving Contact Times in Panel Surveys. *Field Methods 24*(1), 95–111.

Müller, G. (2011). Wechsel des Erhebungsinstituts und Adressweitergabe mit Widerspruchsverfahren im IAB-Haushaltspanel PASS: ein Praxisbericht. *Methoden - Daten - Analysen. Zeitschrift für Empirische Sozialforschung 5*(2), 207–226.

Smith, T. (2011). The Report of the International Workshop on Using Multi-level Data from Sample Frames, Auxiliary Databases, Paradata and Related Sources to Detect and Adjust for Nonresponse Bias in Surveys. *International Journal of Public Opinion Research*, 389–402.

Trappmann, M., S. Gundert, C. Wenzig, and D. Gebhardt (2010). PASS - A household panel survey for research on unemployment and poverty. *Schmollers Jahrbuch (Journal of Applied Social Science Studies) 130*, 609–622.

Wagner, J. (2012). Using paradata-driven models to improve contact rates in telephone and face-to-face surveys. In F. Kreuter (Ed.), *Improving Surveys with Paradata: Making Use of Process Information*. Wiley.

Weeks, M., R. Kulka, and S. Pierson (1987). Optimal Call Scheduling For A Telephone Survey. *Public Opinion Quarterly 51*(4), 540–549.