Adaptive Contact Strategies in a Telephone Survey

James Wagner*

University of Michigan, Survey Research Center
jameswag@isr.umich.edu

Introduction

An important step in the process of completing interviews using the telephone is establishing contact. There is a body of literature that focuses on methods for accomplishing this task in more efficient ways. Much of this literature focuses on determining the average best times to call; or the sequence of calls that have, on average, the highest contact rates.

Greenberg and Stokes (1990) looked at methods that used the history of previous calls to determine the best time to place the next call. Brick et al. (1996) considered a similar approach that used logistic regression models to identify the best time of day, day of week and lag time between calls. Predictors in the model included contextual data as well as information about the results of previous attempts. Others have looked at the impact of various sequences of calls (Weeks et al., 1980; Weeks et al., 1987; Cunningham et al., 2003).

An alternative approach to this problem would be to develop household-specific estimates of the best time to attempt contacting each household. Such an estimate might provide the basis for a more efficient contact strategy. In addition, these estimates would allow us to “tailor” the contact strategy to the household.

The problem is that for most households, we have very little data. Most cases are resolved within the first few calls. Most cases are called in one or a few time slots or “windows.” Very few cases are called in all call windows. In some -- or even most cases -- we will base our judgments about the best time to call on information from other households. This problem seems well-suited for multi-level modeling that allows us to borrow strength from other households when estimating parameters for households for which we have very little data. These estimates become more specific to the household as we gather data from that household.

In this paper, I will outline such a strategy and present evidence of its efficacy from a randomized experiment. The strategy is described as “adaptive” since it learns from sequentially gathered data while also directing how those data are gathered. It develops an estimate of the current best calling time using the current call history data. The prescribed strategy is then attempted, the result is added to the data, and a new estimate is derived from this supplemented data. Results from the field of reinforcement learning will be used to suggest future directions for this research.

Background

Efficient call-scheduling algorithms have long been a subject of research for survey methodologists. Weeks at al. (1980) looked at the best times to place a call using data from an in-person survey. This research was extended by Weeks et al. (1987) to a telephone survey and the timing of the first three calls was considered. Greenberg and Stokes (1990) employed a Markov Decision Process that used the history of previous calls to as well as the frame

---

* Thanks to Richard Curtin for allowing this experiment to be conducted. Thanks also to Joe Matuzak, Dave Dybicki, and Rebecca McBee for working on implementing this experimental design.

data to determine the best time to place the next call. The transition probabilities were estimated using logistic regression models. Brick et al. (1996) considered a similar approach that used logistic regression models to identify the best time of day, day of week and lag time between calls. Predictors in the model included contextual data as well as information about the results of previous attempts. Other research has looked at calling patterns for the first few calls (Kulka et al., 1988; Massey et al., 1996; Cunningham et al., 2003).

The concept I want to employ for exploring the best calling algorithm is a learning model. Over time, as we accrue more data on any particular household, we are learning about their patterns for being at-home and willing to answer the telephone. Successfully contacting a household at one time of day or day of the week (i.e. in a particular calling "window") increases our estimate of the chance of success in that window for that household. Unsuccessful attempts to contact a household in another window decrease our estimate of the probability of achieving contact in that window.

For many households, contact and interview results relatively early in the process -- within the first two or three calls. This means that for many households we have little or no data with which to estimate the probability of being home and willing to accept a telephone call within any window. For these households, we will need to "borrow strength" from the data generated by telephone calls to other households. This can be done using multi-level models (Gelman and Hill, 2007). In those households where we have no data, we are essentially using conditional means where we condition on the frame data that are available for all households (see below).

Research in other fields has addressed similar problems in different contexts. In the area of marketing research, Rossi, McCulloch, and Allenby (1996) consider a similar problem. Their goal was to customize or tailor the face value of a coupon to a specific household. They attempted to estimate household-level parameters using demographic and purchase history data. They used multi-level models to do so. Bollapragada and Nair (2001) considered the problem of improving "right party contact" rates at credit card collection calling centers. They conceived of the problem in a manner similar to that of Rossi and his colleagues. In other words, their goal was to estimate contact probabilities for each household the call center is attempting to reach. They developed an algorithm that assigns the overall average contact rate to each household and adjusts these starting values for each household upward when a call attempt is successful and downward when the attempt fails. In both these papers, the authors have attempted to estimate characteristics of households using historical data from households.

In my application, I will also attempt to estimate household characteristics -- the probability of being contactable (i.e. at home and willing to answer a telephone call). I will use multi-level models where the household is a grouping factor. The fixed effects are frame variables available for all cases. The household-level estimates will be used to decide which cases have their highest probability of contact (not necessarily the highest of all cases) in the current window. Those cases those cases that have their highest probability of contact in the current window are prioritized in the current window. The models are re-estimated daily and the sample is re-prioritized at the beginning of each call window.

Data and Methods

The data come from an RDD telephone survey that is conducted on a monthly basis -- the Survey of Consumer Attitudes (SCA). The survey collects approximately 300 RDD interviews per month. The sample is prepared by a vendor that attaches contextual data to the sample file. The ZIP code of each telephone number is estimated using listed numbers from the same 100-bank. Census data for the associated ZIP Code Tabulation Area (ZCTA) are then attached to each telephone number. Table 1 lists several of the key context variables that are available. Of course, given the estimated geography of the case, any data that are reported for particular geographies can be attached to the sample in a similar manner.

Previous research in suggests that the urbanicity and median income of the estimated geographic area (Dennis et al., 1999; Brick et al., 1996) are predictive of contact rates. As part of the model fitting exercise, different transformations on some of these variables were tried. The natural logarithm of the median income sometimes produced a better fit. Brick et al. (1996) reported using a similar strategy. Other research has reported that the proportion of the population that is Black, the proportion Hispanic, and the median years of education of the estimated geography of the telephone number are predictive of contact rates as well (Brick et al., 1996).

Table 1. Contact Propensity Predictor Variables ($\mathbf{X}_{ij}$)

Context Variables
Listed/Letter Sent
% Exchange Listed
Log(% Exchange Listed)
Sin(% Exchange Listed)
Total Households
Household Density (households per 1000Sq ft.)
Median Yrs Education
Median Income
Log(Median Income)
Census Region
% 18-24
% 25-34
% 35-44
% 45-54
% 55-64
% 65+
% White
% Black
% Hispanic
% Owner Occupied

The data from the telephone survey include the records of every call. The time of each call (its window) and the result (contact/no contact) were recorded for each call. This initial set of calls was reduced for various reasons. The conceptual approach of the estimation is that each data element is an independent, random draw from a Bernoulli distribution of the probability that the household could be contacted. These "draws" allow us to estimate probabilities of contact within windows for each household. In order to make this assumption more plausible, any calls that were set as appointments were deleted. Since it is assumed that these were independent trials, the call number did not enter the models as a predictor. This assumption enables us to estimate the probability of being at home for any call in that window. Estimating the probability of being at home after eight calls of a particular sequence, for example, was not the goal.

Table 2. Call Window Definitions

Window	**Definition**
1	Sat-Sun-Mon 4pm-9pm
2	Tues-Fri 5pm-9pm
3	Sat-Sun 9am-4pm
4	Mon 9am-4pm, Tues-Fri 9am-5pm

In addition, for operational reasons related to the sample management software, refusal conversion and Spanish language calls were eliminated from the analysis.

At the beginning of the field period, there are no call histories for the current sample. Therefore, we use data from prior months. Specifically we used the call records from the same month in the prior year (in order to capture any seasonal effects in the data) and the month prior to the current. Data from the current month are analyzed daily. The models are re-estimated daily, and the results are updated daily with call records from the first day through the prior day included.

Multi-level models are fit daily with the household being a grouping factor. The models provide household-specific estimates of the probability of contact for each of the "call windows" (see Table 2). The predictor variables in this model are the context variables described in Table 1. Let $\mathbf{X_{ij}}$ denote a k_j x 1 vector of demographic variables for the i^{th} person and j^{th} call. The data records are calls. There may be zero, one, or multiple calls to a household in each window. The outcome variable is an indicator for whether contact was achieved on the call. This contact indicator is denoted R_{ijl} for the i^{th} person on the j^{th} call to the l^{th} window. Then for each of the four call windows denoted l, a separate model is fit where each household is assumed to have its own intercept which is from a $N(0,\sigma^2)$ distribution. The model is estimated:

$$\Pr(R_{il} = 1) = \operatorname{logit}^{-1}(\beta_{0l} + \beta_{0il} + \sum_{j=1}^{p} \beta_{jl} X_{ijl})$$

The next step is to compare the estimated contact probabilities within a household and find the window with the highest probability of contact for that household. During that window, the case -- along with all other cases that meet this criterion -- will be sorted to the top of list by the call scheduling algorithm. Each case had a window with the second highest estimated probability of contact. During that window, the case would be sorted on the list after the cases for which that window had the highest probability. In this way, all active cases were available for calling in every window.

Under this sorting approach, a case with a low probability of contact could be sorted to the top of the list in any given call window, as long as the estimated probability of contact was highest for the case within that window. Once this group had been identified, they were sorted randomly. Future research aims to address what method of sorting within this group may work best.

The experimental design required frequent sorting of the list as the call windows were specific to the time zone. For example, on a Tuesday, the list was sorted first thing in the morning, at 5pm EST, 6pm EST, 7pm EST, and at 8pm EST as the various time zones crossed the call window boundary. The experimental design required that the experimental and control groups be sorted in an interleaving fashion. The past practice had been to sort at the beginning of the day. The control group sort was based on an algorithm that assigned weights to various factors and then sorted based on the sum of these weights. There were two weighting schemes. The first was used for the first part of the month. It prioritized cases by time zone, those that had fewer than 5 calls, and those that were not called already on that day. Later in the month, the weighting scheme included the following factors: time zones, number of calls, whether the case had already been called that day, whether contact had previously been made with the number, and whether a household listing had been taken.

Results

The protocol appears to improve contact rates. Table 3 presents the overall contact rates for the experimental and control groups for the "eligible" calls by month and combined. The χ^2 test reported here uses the Rao-Scott approach to account for the clustering of the observations within households. In addition to these differences, we found that approximately 30% of all calls in the experimental group resulted in a "Ring-no-answer" (RNA) while 39% of the calls in the control group produced this result.

Table 3. Calls, Contacts, and Contact Rates by Experimental Group and Month

	Control			Experiment			
Month	**Calls**	**Contacts**	**Contact Rate**	**Calls**	**Contacts**	**Contact Rate**	**Pr> χ^2**
August	4,467	470	0.105	4,238	517	0.122	0.179
September	5,418	507	0.094	5,025	596	0.119	0.016
Combined	9,885	977	0.099	9,263	1,113	0.120	0.008

Although contact rates and efficiency were improved by the experimental method, this method was not applied to refusal conversions or Spanish language cases. The efficiency gains for the calls governed by the algorithm were lost later in the process. This was because refusal conversions in the experimental group required more calls than those cases in the control group. Table 4 presents the total number of calls and interviews by those calls that were governed by the experimental method ("In-Algorithm") and total calls. The inefficiency of refusal conversions in the experimental group can be seen in total calls column. The total calls are nearly equal to those for the control group, despite the relative efficiency of the in-algorithm calls.

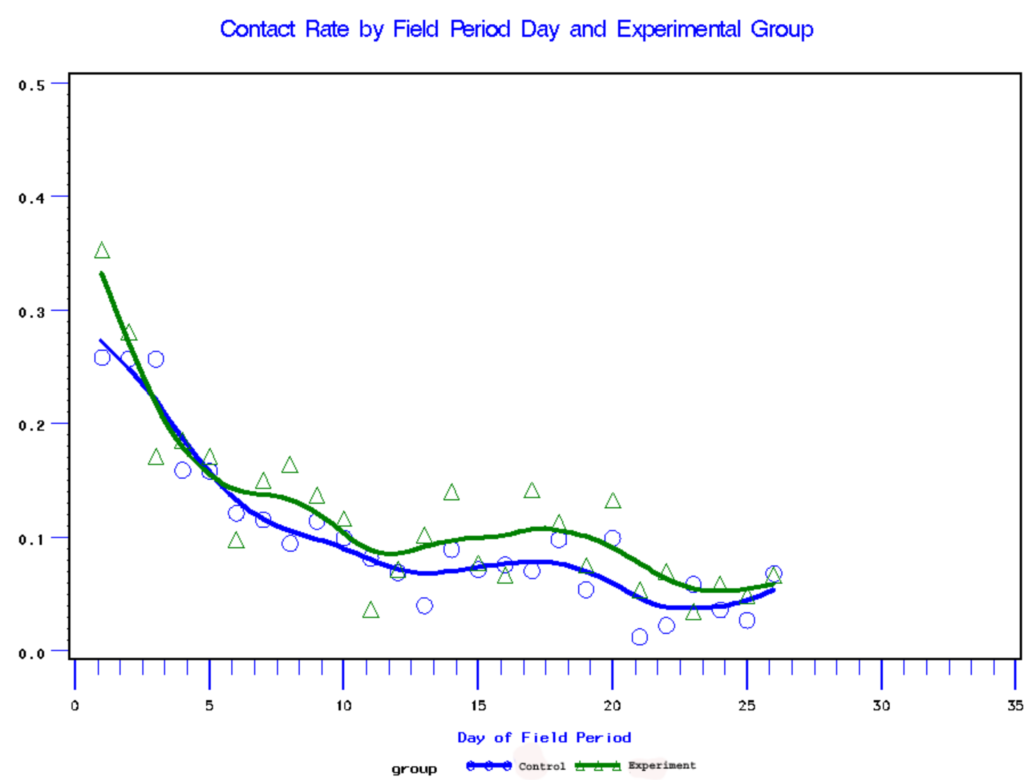
Table 4. In-Algorithm and Total Calls and Interviews by Experimental Group and Month

		In-Algorithm			Total		
Month	**Group**	**Call s**	**Interviews**	**Calls Per Interview**	**Total Calls**	**Interviews**	**Calls Per Interview**
August	CON	4,467	84	53.2	6,510	150	43.4
	EXP	4,238	97	43.7	6,264	156	40.2
September	CON	5,418	96	56.4	6,961	146	47.7
	EXP	5,025	98	51.3	7,094	148	47.9
Combined	CON	9,885	180	54.9	13,471	296	45.5
	EXP	9,263	195	47.5	13,358	301	44.4

This result seems to imply that there is some interaction between what happens in the early attempts at contact and the later phase of refusal conversion. It may be that the experimental method leads to contacts at times that are inconvenient for persons in the household. Future experiments will aim to control the refusal conversion calls to see if the efficiency of these calls can be improved.

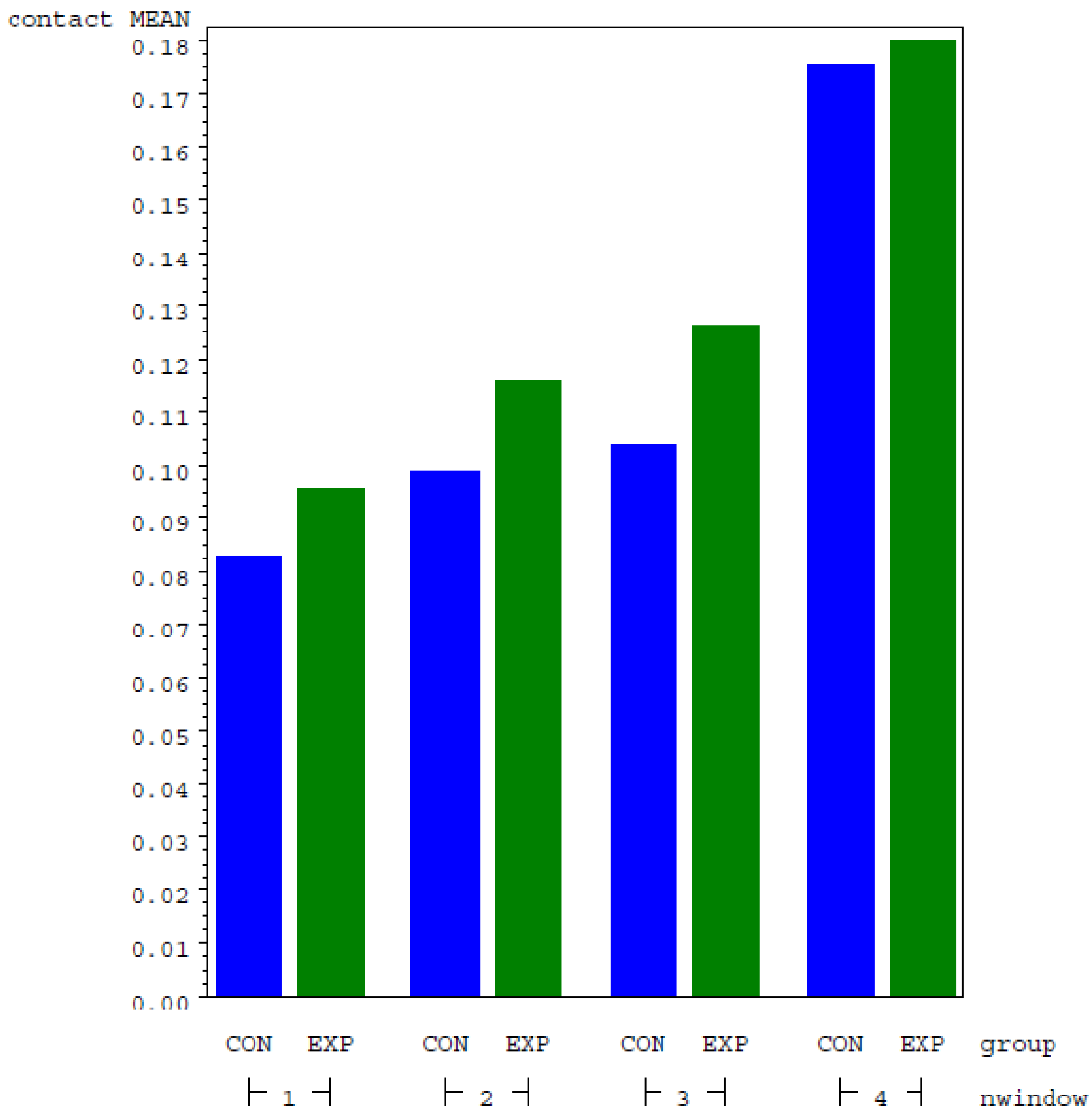
There were some other interesting results from the experiment. Figure 1 presents the contact rates for the experimental and control groups by day of the field period. The results for the experimental group relative to the control group appear to follow the trajectory of a learning model that explores the data. This exploration leads to early, relative inefficiencies, but then produces efficiency gains late. The trend line is a smoothed spline regression line.

Figure 1. September Contact Rate by Field Period Day and Experimental Group



Given that the experiment was defined using four call windows, it is useful to look at how the experiment and control groups fared over each of these call windows. Figure 2 presents these results for August. The experimental group had an edge in each of the windows.

Figure 2. August Contact Rates by Experimental Group and Call Window.



Conclusion

The ability to establish contact with sampled telephone numbers is still an important part of the data collection process. Early evidence from an algorithm proposed here suggests that improvements in contact strategies can be made. These results may be limited by the particular survey context – a monthly survey with a fixed field period. The approach is adaptive in that it changes its proposed action after the results of prior actions are known.

Future research will attempt to resolve the issue with the relative inefficiency of refusal conversions in the experimental group. In addition, more "exploratory" strategies will be attempted. In the field of reinforcement learning (Sutton and Barto, 1998), the balance between "exploration vs exploitation" is a key defining element of any particular learning strategy. For any given problem, the learning algorithm must decide whether to maximize its expected gain from the current decision or to explore less than maximal options in order to see if there are suboptimal strategies for the current decision that lead to a higher payoff in the long run. I hope to extend the results of this experiment by employing more exploratory strategies that prioritize the cases about which we have more uncertainty to see if this increases efficiency or helps us to contact difficult cases. In addition, the method will be employed on other surveys to see if it is dependent on the context of the survey used here.

References

Bollapragada, S. and S. Nair (2001). Improving right party contact rates at outbound call centers.

Brick, J. M., B. Allen, P. Cunningham, and D. Maklan (1996). Outcomes of a calling protocol in a telephone survey. *Proceedings of the Survey Research Methods Section of the American Statistical Association*, 142-149.

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.

Dennis, J. M., C. Saulsberry, M. P. Battaglia, A.-S. Roden, D. C. Hoaglin, M. Frankel, N. A. Mathiowetz, P. J. Smith, and R. A. Wright (1999). Analysis of call patterns in a large random-digit-dialing survey: The national immunization survey. *Conference website of the International Conference on Survey Nonresponse 1999*, 1-23.

Gelman, A. and J. Hill (2007). *Data analysis using regression and multi-level/hierarchical models*. Cambridge ; New York: Cambridge University Press.

Greenberg, B. S. and S. L. Stokes (1990). Developing an optimal call scheduling strategy for a telephone survey. *Journal of Official Statistics 6* (4), 421-435.

Kulka, R. A. and M. F. Weeks (1988). Toward the development of optimal calling protocols for telephone surveys: a conditional probabilities approach. *Journal of Official Statistics 4* (4), 319-332.

Massey, J. T., C. Wolter, S. C. Wan, and K. Liu (1996). Optimum calling patterns for random digit dialed telephone surveys. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, 485-490.

Rossi, P. E., R. E. McCulloch, and G. M. Allenby (1996). The value of purchase history data in target marketing. *Marketing Science 15* (4), 321-340.

Sutton, R. S. and A. G. Barto (1998). Reinforcement Learning: An Introduction. Cambridge, Mass., MIT Press.

Weeks, M. F., B. L. Jones, J. Folsom, R. E., and C. H. Benrud (1980). Optimal times to contact sample households. *Public Opinion Quarterly 44* (1), 101-114.

Weeks, M. F., R. A. Kulka, and S. A. Pierson (1987). Optimal call scheduling for a telephone survey. *Public Opinion Quarterly 51* (4), 540-549.