The distribution of survey contact and participation in the United ...
Bickart, Barbara;Schmittlein, David
JMR, Journal of Marketing Research; May 1999; 36, 2; ProQuest Central
pg. 286

BARBARA BICKART and DAVID SCHMITTLEIN*

In this article, the authors show how survey responses to the question "How often have you participated in a survey in the past [specified time period]?" can be used to construct a meaningful estimate of average survey contact and response rates in the population at large, as well as of the distribution of survey responses in the population. That is, the authors show how the nonresponse bias associated with asking people about their prior survey participation can be removed. They propose and validate a simple yet powerful statistical model of survey contact across the U.S. population and of survey response among those contacted. The authors find that survey contact and participation rates are lower than previously assumed and that a small percentage of the population is completing the majority of surveys.

The Distribution of Survey Contact and Participation in the United States: Constructing a Survey-Based Estimate

> Deluged by a growing number of telephone solicitations and increasingly jealous of their time, more and more Americans are refusing to participate in market-research surveys.
>
> —*The New York Times*, October 5, 1990, p. 1

> As more and more Americans feel greater pressures and demands for their time, we as an industry are finding it increasingly difficult to reach representative samples of respondents, economically.
>
> —*Walker: Industry Image Study*, 9th ed., 1990

To judge by the preceding popular press articles, we might conclude that Americans are overwhelmed by the volume of survey requests and, as a result, have become unwilling to assist decision makers with their collective input through survey research. Furthermore, there appears to be a concern that some Americans' views may be overrepresented in surveys. How often are Americans surveyed? Perhaps more important, what proportion of survey responses are accounted for by the subpopulation most likely to be both contacted and responsive? Both of these questions can be answered by estimating the distribution of survey responses and responsiveness in the population.

The approach used most frequently involves contacting U.S. adults and asking their frequency of survey participation during some prior time (e.g., Groeneman 1994; National Research Council 1979; O'Neill 1996; Roper 1986; Schleifer 1986). Although this approach is simple to implement, there is one major disadvantage; namely, not all people are willing to respond to a survey request. It appears that survey response propensity is a personal characteristic that endures through time and varies substantially across the population (Brennan and Hoek 1992; Goyder 1987). Thus, the respondents to the survey asking about survey frequency are themselves unrepresentative of the population by the very nature of their having answered the survey (cf. Goyder 1986).

Although many of the survey-based studies previously cited have noted this problem, none has attempted to address this lack of representativeness. Therefore, these studies cannot offer an answer to two general questions of interest: (1) On average, how often are Americans surveyed? and (2) To what extent does a small fraction of Americans account for the lion's share of survey participation? We show how survey responses to the question "How often have you participated in a survey in the past [specified time period]?" can be used to estimate survey contact and response rates, as well as the distribution of survey responses in the population. That is, we show how the nonresponse bias described in the preceding paragraph can be removed. We use empirical survey results collected by The Council for Marketing and Opinion Research (CMOR) and

*Barbara Bickart is Assistant Professor of Marketing, School of Business-Camden, Rutgers–The State University of New Jersey (e-mail: bickart@crab.rutgers.edu). David Schmittlein is the Ira A. Lipman Professor and Professor of Marketing, The Wharton School, University of Pennsylvania (e-mail: schmittlein@wharton.upenn edu). The authors appreciate the helpful comments of Seymour Sudman and the assistance of Harry W. O'Neill of Roper Starch and Sid Groeneman of Market Facts in providing some of the data analyzed. To interact with colleagues on specific articles in this issue, see "Feedback" on the *JMR* Web site at www.ama.org/pubs/jmr.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

Market Facts Inc. to calculate how often Americans are contacted to participate in a survey, how responsive they are to those surveys, and how often survey research results are based on answers from the same small fraction of the population. We find that unadjusted estimates of these quantities from survey responses (i.e., the figures most often cited in industry and popular press reports) are severely biased.

In the next section, we offer a discussion of the main issues that have arisen regarding the frequency of survey contact and survey response. Subsequently, we review the empirical evidence brought to bear on these issues, namely, the claimed frequency of survey participation among respondents to a telephone survey. We propose and validate a model envisioning heterogeneity in survey contact rates among U.S. adults, as well as heterogeneous survey response probabilities across these adults. The model is used to assess the true population distribution of survey contact and participation frequencies and to calculate various survey participation indices of interest.

ISSUES REGARDING SURVEY CONTACT AND SURVEY PARTICIPATION

Understanding survey participation rates in the United States is of interest for two reasons. First, there is concern that, with the growth of survey research, the number of contacts per person has increased, and subsequently, a larger proportion of the population has been asked to participate in surveys. Groves, Cialdini, and Couper (1992) suggest that this type of oversurveying could result in lowered response rates for the following two reasons: (1) With increasing contacts, respondents' overall attitudes toward the survey industry may become less favorable, and (2) as people are contacted more often, the sense that the opportunity to provide their opinions in a survey is a "rare" and, therefore, valuable experience diminishes, resulting in lower compliance.

Second, the extent to which surveys are repeatedly relying on a small fraction of the population is another important issue. It has been suggested that some segments of the population are easier to contact and more willing to cooperate in surveys (Goyder 1986; Schleifer 1986); in other words, there is heterogeneity in contact and participation rates across the population. The concern here centers on two issues. First, survey responders and nonresponders differ on demographic and attitudinal variables (see Groves 1989; Struebbe, Kernan, and Grogan 1986). Less educated and older people are less likely to participate (Groves 1989). Second, people who participate frequently in surveys are likely to be enthusiastic and involved respondents, regardless of the topic, resulting in response biases such as yea-saying and telescoping (Roper 1986). Because of their high involvement level, these respondents may think through their answers more carefully than would less interested respondents. Thus, their answers are likely to be more elaborate than not only the survey responses of the less involved nonrespondents, but also the real-world behavior of those less involved nonrespondents, which leads to biased survey results. Understanding the seriousness of these biases requires an ability to obtain an unbiased estimate of survey participation.

CLAIMED SURVEY PARTICIPATION BY SURVEY RESPONDENTS

Evidence regarding the (over)surveying of the United States generally is based on a telephone survey measuring claimed survey participation. The question asked is typically of the form, "How often have you participated in a survey in the past [specified time period]?" Results are reported as a histogram of survey participation frequency. In 1994, Market Facts Inc. conducted a random digit dial (RDD) telephone survey of U.S. adults, asking their number of survey participations during the previous six months (see Groeneman 1994). More recently, ten market research firms acting under the aegis of the CMOR asked the frequency of survey participation during the previous year (CMOR 1995; O'Neill 1996). Our analyses will be based on (and compare) the results from these two surveys. To illustrate the relevant issues, however, the remainder of this section uses only CMOR's 1995 survey.

The most fundamental finding regarding survey participation among U.S. adults is the distribution we show in Table 1 for the survey participation frequency within the year prior to the survey, as claimed by respondents to the CMOR survey. That 45.3% of the survey respondents claimed participation in the past year and more than 20% claimed to have had three or more participations seems to indicate a potential for survey response fatigue in the United States. Projecting to the estimated 194 million adults in the United States as of 1995 (U.S. Bureau of the Census 1995), these figures translate to 88 million adults surveyed every year and 43 million surveyed four or more times in each such period.

Overall, the Market Facts and CMOR surveys appear consistent with the best current practice in the industry, and the RDD approach provides some reassurance regarding representativeness of the individuals contacted. One factor, however, limits the projectability of this survey's (and similar surveys') participation frequency distribution. The CMOR 1995 telephone survey had, in total, 4800 appropriate contacts. Of these, 40% (1920) actually participated. Although the 4800 people contacted may represent the desired universe, the 1920 respondents almost certainly do not; by their willingness to participate in this survey, the 1920 respondents have demonstrated that they are particularly accessible and willing to participate in surveys generally, to the extent that differences exist in such a propensity.

Consider the two archetypical interpretations of the claimed participation frequency distribution. In the first, all adults are imagined to have the same survey participation propensity. Using the 40% figure from the 1995 study, this would mean that each contacted person has a probability of .4 of participating in each and any survey. In this case, the 1920 respondents are representative of all 4800 individuals contacted because, by assumption, all people have the same participation probability.

The alternative to the preceding complete homogeneity assumption would be complete heterogeneity; that is, some fraction of U.S. adults are willing to respond to surveys,

Table 1
CLAIMED PARTICIPATION AMONG CMOR RESPONDENTS

	Number of Surveys in Past Year				
	0	*1*	*2*	*3–5*	*6+*
Number of respondents	1020	166	270	279	130
Percentage of respondents	54.7	8.9	14.5	15.0	7.0

Source: O'Neill 1996.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

and the remainder are completely unwilling. So using the 1995 results, we would view 40% of U.S. adults as always being willing to participate in any survey and 60% as never willing. In this extreme case of heterogeneity, the survey participations reported by these 1920 respondents would represent all survey participations by the entire set of 4800 contacted (because the 4800 – 1920 nonrespondents never participate in surveys). Projections to the adult population are still feasible but are much different than the values previously listed. Instead of projecting that 45.3% of U.S. adults participate in a survey each year, we would conclude that 45.3% of the 40% who participate in surveys participated during this period, 54.7% of this 40% will participate if asked but did not have a chance during this period, and the remaining 60% never participate. With complete heterogeneity assumed, estimated survey participation during a typical year drops to (.453)(.40) = .181 of U.S. adults.

A comparison of simple summary statistics for survey participation in the complete homogeneity and complete heterogeneity scenarios appears in Table 2. This distinction shows that the claimed participation frequency distribution cannot be used directly to make statements about the concentration in survey responses, for example, that the degree of failure of the claimed participation distribution to represent accurately the role of heavy survey participators is related directly to the actual differential importance of these highly active respondents. If they really differ from other U.S. adults, those differences have not been captured correctly in the claimed participation distribution.

The truth about survey participation propensity lies somewhere between these two extreme scenarios. People surely differ somewhat in their general propensity to be contacted for surveys and their willingness to participate when contacted. In light of the wide gap between inferences in the two scenarios, simply leaving them stand as upper and lower bounds hardly would be satisfying. Instead, a simple model of survey contact and response that allows for both our extreme scenarios, and for the intermediate scenarios more likely, will be calibrated (and validated) with the claimed participation frequency tables available.

Table 2

PROJECTION OF ONE-YEAR SURVEY PARTICIPATION FREQUENCY AMONG U.S. ADULTS (1995)

	Complete Homogeneity	*Complete Heterogeneity*
Percentage participating at least one time	45.3	18.1
Percentage participating at least three times	22.0	8.8
Number of U.S. adults surveyed (millions)	88	36
Number surveyed at least three times (millions)	43	17
Total number of survey participations (millions)	349	140

ACCOUNTING FOR DIFFERENTIAL CONTACT RATES AND DIFFERENTIAL RESPONSE PROPENSITY: THE BB/NBD MODEL

We begin with the assumption that each person's survey contact and response experience is stationary and random in time, as follows:

Assumption 1: Survey contacts arrive randomly in time for a given person, with (constant) rate λ. (Poisson process for contacts.)

Assumption 2: For a given person, when contacted, the probability p that this person will agree to participate in a survey remains constant through time and is not affected by previous decisions regarding participation. (Bernoulli process for participation given contact.)

Assumption 1 is supported because survey contacts in the aggregate among U.S. adults arise as the superposition of many survey firms' individual, nonsynchronized decisions. This superposition is known to lead to randomly timed contacts in the aggregate, even when the component processes are nonrandom (Karlin and Taylor 1975, pp. 221–28). Assumption 2 also holds promise as a representation for an adult's survey participation decisions. Although a person's inclination to participate may vary from occasion to occasion depending on the type of survey (mail, telephone, mall, and so on) and time of day, the superposition of different survey types, times of day, and so forth will make the aggregate survey participation phenomenon appear similar to a simple Bernoulli process (zero order, constant participation probability) for any single person (Kahn, Morrison, and Wright 1986).

In Assumptions 3 and 4 we allow both the contact rate λ and the participation probability p to vary from person to person among U.S. adults. By specifying a probability distribution to capture the heterogeneity of each of λ and p across persons, we can derive the expected distribution of reported survey participation among the subpopulation that happens to agree to participate in some randomly arriving survey. More important, this expectation can be inverted; that is, from the distribution of reported participation frequency, it is possible to infer the distribution for both survey contacts and participations in the general U.S. adult population that is most consistent with these observations.

Assumption 3: A natural heterogeneity distribution for λ is the gamma, because it is the conjugate prior for the Poisson (see Morrison and Schmittlein 1988).

The probability density function for this distribution of survey contact rates λ across persons is as follows:

$$f(\lambda) = \frac{\alpha}{\Gamma(r)}(\alpha\lambda)^{r-1}e^{-\alpha\lambda};\ \lambda > 0; \tag{1}$$

with mean and variance $E[\lambda] = r/\alpha$ and $Var[\lambda] = r/\alpha^2$.

Assumption 4: The beta distribution is used to capture heterogeneity in participation probabilities across adults. It is the conjugate prior for our binomially distributed survey participation frequency (conditioned on N survey contacts) and has a long and successful history in representing heterogeneous Bernoulli processes (Greene 1982;

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

Johnson and Kotz 1970, pp. 42–43; Schmittlein 1989).

The beta's density function is as follows:

$$(2) \quad g(p) = \left[\frac{1}{B(a,b)}\right] p^{a-1}(1-p)^{b-1}; \; 0 < p < 1,$$

where B(a,b) is the beta function. The distribution has mean and variance $E[p] = a/(a+b)$ and $Var[p] = ab/[(a+b)^2(a+b+1)]$.

Interpreting the amount of heterogeneity in contact rates and participation probabilities across adults is of particular interest to us. For the gamma distribution of survey contact rates, note that the coefficient of variation (standard deviation/mean) = $r^{1/2}$, so the parameter r is an indicator of homogeneity in these rates. As r becomes large, the variation in survey contact rates diminishes to 0, and conversely, as r approaches 0, the distribution's mass concentrates in the tails (contact rates either very high or near 0). Similarly, the sum of the two beta distribution parameters (a + b) indicates the degree to which survey participation probabilities are homogeneous across U.S. adults. For example, (a = 1, b = 1; a + b = 2) corresponds to the uniform distribution. When both a and b are greater than 1, the beta distribution is bell-shaped with an interior mode. As a + b increases to infinity, the probability mass condenses on this mode. Conversely, if both a and b are small (<1), the distribution of participation probabilities is U-shaped with a segment of adults that is very willing to participate (high p values) and another segment only rarely willing to participate (low p values).

One final assumption is needed to complete our model of survey contact and participation:

Assumption 5: The survey contact rate λ and survey participation probability p are distributed independently across U.S. adults.

This assumption is useful in simplifying the required calculations and necessary in creating a model sufficiently parsimonious to be estimated from the claimed participation frequency distribution. On one hand, the inclination of survey firms to pursue affluent, successful, busy Americans (who are less likely to be willing to participate) would tend to induce a negative correlation between λ and p. On the other hand, high willingness to participate often is used to create groups of people to contact, as in the large mail panels operated by market research firms. Although we cannot be sure that these two effects will counteract perfectly, independence between contact rates λ and participation probabilities p appears a reasonable working assumption.

Because a randomly chosen person's survey participation follows the Beta Binomial distribution and survey contact frequency follows the Negative Binomial Distribution, we call the survey contact/participation process the BB/NBD model. This process arises naturally in some other social science contexts, and the needed properties of the model first were derived in Schmittlein, Bemmaor, and Morrison's (1985) work and appear in the Appendix.

As we noted previously, we fit the BB/NBD model to the observed distribution of claimed survey participation among respondents to the telephone survey. We already indicated that, with the BB/NBD model's assumptions, these respondents are not representative of the population at large. The distribution of their collective contact rates λ should be representative of the contact rate distribution $f(\lambda)$ in the U.S. adult population because contacts in the telephone survey were generated through a stratified random sample. But the distribution of participation probabilities $g^*(p)$ among those surveyed is not the same as the participation probability distribution $g(p)$ among U.S. adults at large, because those participating in this survey are likely to be persons with relatively high p values (willingness to participate) from the general population.

In short, if survey contact/participation follows the BB/NBD model with parameters (r, α, a, b) reported survey participation among those who participated in the telephone survey will not follow the BB/NBD (r, α, a, b) distribution. To fit the BB/NBD model to the reported participation histogram, we must know how $g^*(p)$ differs from the beta (a, b) distribution $g(p)$.

Letting X denote the number of survey participations for a person receiving N contacts, $g^*(p)$ is the updated distribution for p conditioned on the event (X = 1; N = 1). That is, those reporting their survey participation frequency have been contacted once (N = 1) and agreed to participate (X = 1). This distribution is, in general:

$$(3) \quad g^*(p \mid N=1;\, X=1) = \frac{\Pr[X = 1 \mid N+1]\; g(p)}{\int_0^1 \Pr[X = 1 \mid N = 1;\, p]\; g(p)\, dp},$$

where $g(p)$ is again the beta (a,b) distribution p.d.f. given in Equation 2 and $\Pr[X = 1 \mid N + 1; p]$ is the binomial distribution probability. Substituting these quantities in Equation 3, $g^*(p)$ becomes

$$(4) \quad g^*(p \mid N=1;\, X=1) = \left[\frac{1}{B(a+1,\, b)}\right] p^{a}\,(1-p)^{b-1}.$$

Notice from Equation 2 that this posterior distribution $g^*(p)$ is again a beta distribution, with new updated parameters ([a + 1], b).

This result allows us to estimate the parameters (r, α, a, b) of the BB/NBD process for the U.S. adult population from the telephone survey respondents. Although reported participation among these respondents (assumed to follow the BB/NBD process) does not follow the U.S. adult population's BB/NBD (r, α, a, b) distribution, it follows a BB/NBD distribution with the parameters (r, α, a + 1, b). Thus, the BB/NBD distribution probabilities given in the Appendix can be applied to reported participation among respondents, with the substitution of a + 1 for a. Doing so, the (r, α, a, b) parameters estimated (in our case by maximum likelihood) can be applied to project survey contact and participation among U.S. adults through direct substitution into the Appendix's formulas.

Table 3 shows the BB/NBD parameter estimates for survey contact and participation among the U.S. adult population and the fit of the BB/NBD for the two data sets. Recall that the BB/NBD parameters used to calculate these

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

Table 3
FIT OF THE POISSON AND BB/NBD MODELS TO THE DISTRIBUTION OF CLAIMED SURVEY PARTICIPATION FREQUENCY

	1995 CMOR Data			*1994 Market Facts RDD Data*		
Survey Participation Claimed Frequency:	*Actual: One Year*	*Poisson*	*BB/NBD*	*Six Months*	*Actual: Poisson*	*BB/NBD*
0	1020	502	990	422	260	411
1	166	659	311	161	320	202
2	270	432	173	144	196	113
3				72	80	65
4	279	268	239	36	25	39
5				54	8	59
6+	130	4	152			
Mean Absolute Deviation		262	67		73	16
χ^2		4582	132		484	19
Degrees of Freedom		4	1		5	2
BB/NBD Parameters (annualized)						
r			.364			.832
α			.182			.296
a			.107			.0255
b			.161			.0725
Average Annual Contact Rate r/α			2.00			2.80
Average Participation Probability $a/(a + b)$			.400			.260
Average Annual Participation Rate $ra/[\alpha(a + b)]$			.800			.728

expected frequencies were the values listed in Table 3, except with a + 1 substituted for a (e.g., 1.107 instead of .107 for the CMOR data) to account for the willingness to participate bias. With our assumptions, as previously listed, the BB/NBD should be able to fit the frequency counts for both a one-year (CMOR) and a six-month (Market Facts) period. For the sake of comparability, Table 3 reports BB/NBD parameters for each data set that would characterize a one-year time period. (As is evident in the Appendix, modifying the Market Facts data set parameters to replicate the six-month histogram simply involves multiplying the one-year α value by 2.)

The fit of the BB/NBD is reasonably good. It does a much better job than the homogeneous Poisson in capturing the fat-tailed distribution of reported frequency among respondents. (The Poisson is a special case of the BB/NBD in which contact and participation rates do not vary across adults.) The reduction in chi-squared goodness of fit is dramatic, and the BB/NBD's mean absolute deviation is only 25% that of the Poisson. Notice also that the CMOR and Market Facts RDD surveys focus on all U.S. adults, and so, in spite of their coming from surveys covering different recall periods (one year and six months, respectively), they generate remarkably similar (annualized) BB/NBD parameters and, most important, similar estimates of survey contact and participation rates, as we show in the bottom three rows of Table 3. We conclude that BB/NBD is an effective model for representing survey contact and participation. Because the many studies conducted to date focus on reporting the (biased) participation frequency distribution among respondents, we begin our use of the BB/NBD by projecting the (unbiased) distribution of annual survey participation among U.S. adults.

THE DISTRIBUTION FOR SURVEY CONTACTS AND SURVEY PARTICIPATION AMONG U.S. ADULTS

With the BB/NBD model parameter values estimated in the last section for the CMOR data, the distribution for survey contact frequency among U.S. adults in 1995 is the negative binomial (with parameters $r = .364$, $\alpha = .182$). According to the BB/NBD model, the CMOR report finds that 50.6% of U.S. adults were not contacted, 15.6% were contacted once, 9.0% were contacted two times, 13.0% were contacted between three and five times, and 11.8% were contacted six or more times in one year.

Similarly, the distribution for survey participation frequency, using the BB\NBD formula (Equation A6) in the Appendix, appears in Table 4. For comparison, we also list the distribution of reported participation frequency among respondents to the 1995 CMOR survey. Notice the magnitude of the nonresponse bias in the CMOR statistics; we project that the percentage of U.S. adults who remain participation free in a year is more than 40% greater than a direct examination of the CMOR results would indicate (i.e., 77% versus 54.7%). Furthermore, the percentage of U.S. adults surveyed heavily (six or more times) in a year is esti-

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

mated to be 40% less than the CMOR statistic (i.e., 4.2% versus 7%). So, using the BB/NBD model to account for nonresponse bias changes substantially the picture of U.S. survey participation. Table 5 provides a summary of the discrepancy between apparent participation statistics (based on the survey's respondents) and the corresponding BB/NBD estimate, which removes the participation bias. We next use the model to address the two main questions of this article: "Are U.S. adults overwhelmed by a tidal wave of surveys at present?" and "How concentrated are survey responses among a small fraction of U.S. adults?"

ARE AMERICANS OVERWHELMED BY THE VOLUME OF SURVEYS?

We began this article with claims that U.S. adults have been deluged with surveys and that this volume of pulse-taking has been a factor in declining survey participation rates. On the basis of our model results, we have included some indicators of survey volume and distribution among people in Table 6. When a statistically sound method for estimating survey participation is used, the pattern of contact and participation provides a clear view of this issue. The overall volume of surveys simply does not support the survey deluge hypothesis. The BB/NBD model's estimates of survey contact and participation are listed in the top half of Table 6.

These results show that more than three-quarters of U.S. adults do not participate in any surveys in a year. No more than 10% participate three or more times during a year. The number of survey contacts per adult in a year is between two and three, and the number of survey participations per adult in the same period is less than one (i.e., .728 in the Markets Facts study and .800 in the CMOR results). These figures are substantially lower than the raw CMOR survey result statistics indicated. In short, allocating the current total volume of surveys to U.S. adults "fairly," each adult would participate in a survey every 15 months, which hardly is overwhelming to him or her.

Survey contacts and participations do not occur evenly across all sectors of society. There are three reasons that survey participation "clumps" in any given year:

1. Random chance in selecting contacts, whereby some persons get selected more often (and others less often) than the average;
2. Skewing contacts toward certain population groups (e.g., based on demographics or geographics); and
3. Skewed willingness to participate among certain population groups.

Table 4

FREQUENCY OF SURVEY PARTICIPATION IN ONE YEAR

	0	*1*	*2*	*3–5*	*6+*
U.S. Adult Population BB/NBD model estimate, percentage	77.0	8.7	4.3	5.8	4.2
1995 CMOR survey respondents, percentage	54.7	8.9	14.5	15.0	7.0

Table 5

DISCREPANCY BETWEEN APPARENT SURVEY PARTICIPATION (BASED ON SURVEY RESPONDENTS) AND ESTIMATED PARTICIPATION THROUGH MODEL-BASED PROJECTION TO U.S. ADULTS

	Apparent:Survey (Respondents)	*Model-Based Estimate: (BB/NBD)*	*Percentage Discrepancy: 100 × {[(a) – (b)]/(b)}*
Percentage participating at least once			
In a year (CMOR 1995)	45.3	23.0	+ 97.0
In six months (Market Facts RDD 1994)	52.2	15.4	+240.9
Percentage participating three or more times			
In a year (CMOR 1995)	21.9	9.9	+121.2
In six months (Market Facts RDD 1994)	18.2	5.0	+264.0
Number of participations per year per U.S. adult			
CMOR 1995	1.80	.80	+125.0
RDD 1994	2.62*	.73	+263.9

*Estimated using the BB/NBD model.

Table 6

ESTIMATED SURVEY CONTACT AND PARTICIPATION STATISTICS FOR U.S. ADULTS

	CMOR 1995	*Market Facts RDD 1994*
Number of participations per person per year	.800	.728
Number of contacts per person per year	2.00	2.80
Percentage participating at least once in a year	23.0	20.0
Percentage participating at least three times in a year	9.9	10.0
Top 20% of Survey Participants		
Percentage of total U.S. adult population that they represent	4.6	4.0
Percentage of total U.S. surveys for which they account	56.5	50.0
Average number of participations per year for these people	9.8	9.0

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

The BB/NBD's estimated parameter values provide some initial evidence about this clumping of participation. Specifically, using the CMOR study's parameters, among those people who are contacted at least once in a year, the average number of survey contacts is 4.05. Similarly, among those who participate in a survey at least once during a year, the average number of survey participations is 3.48. Instead of a deluge, then, survey participation in America resembles heavily scattered showers; that is, the burden of response falls on a small fraction of the population. In the next section, we show just how great this reliance is.

THE CONCENTRATION IN SURVEY PARTICIPATION

Many repetitive human activities have been observed to follow an 80/20-type law; that is, a small fraction of the population (e.g., 20%) accounts for a large fraction (e.g., 80%) of the total volume of activity. The BB/NBD's projection of survey contact and participation across the population enables us to examine this kind of concentration. With the parameter values reported in Table 3 and the results in the top half of Table 6, we find that the top 20% of U.S. adults in a year (i.e., the 20% with the most survey participations) account for essentially all surveys completed in that year. (Recall that only 20% to 23% of U.S. adults participate at all in a year; accordingly, this group accounts for all participations.)

This kind of concentration in participation is striking but perhaps not surprising. When some fraction of the population does not participate at all (e.g., because of demographic profile or inability to contact for survey participation), researchers often summarize concentration among participators as an alternate indicator of overreliance on a small segment of the population (e.g., Schmittlein, Cooper, and Morrison 1993). These results are provided in the bottom half of Table 6. Among survey participants in a year, the top 20% of those participants account for between 50 and 56% of all survey responses. Furthermore, we note that this top 20% of participants, who represent approximately half of all U.S. surveys completed, are only 4% to 5% of the total U.S. population. Finally, this small group participates, on average, between nine and ten times a year, as per the bottom row of Table 6. It would not be surprising if this 4% to 5% of the population felt overwhelmed. They are being surveyed almost once a month and are providing most of the country's survey data.

Such extreme concentration cannot be occurring through random chance in contacts. It also, of course, cannot occur simply because disparate survey researchers each are interested in some small segment of the U.S. population. Rather, such researchers would have to share an interest in the same small segment(s). Accordingly, we ask whether this high level of survey participation concentration is mandated by the extent to which importance or relevance for certain research issues is itself concentrated among certain groups. For example, marketing researchers may focus product design research only on those with the wherewithal to purchase. We note, however, that the distribution for survey participation shows substantially more concentration than even the distribution of U.S. income. Specifically, the top 5% of U.S. households account for approximately 20% of total U.S. income (U.S. Bureau of the Census 1996, p. 467), whereas the top 5% of households in terms of survey participation account for nearly 60% of all surveys. Put another way, if survey participation (or contact) went up proportionately with income, the resulting concentration would be approximately one-third of that actually observed here. If the observed concentration stemmed from this kind of demographic indicator alone, surveying propensity would increase much more than proportionately with the person's economic status.

Other sources of concentration are businesses' increasingly popular "relationship marketing" programs, which establish customer lists as a ready basis for survey initiatives; the aggressive marketing of media subscription lists and catalog customer lists for rental; and the impact of general list providers. In this vein, a major source of concentration is presumably the maintenance of consumer panels by major survey research firms. Such panel members get recycled frequently for contact because they are readily available to the firm and known to have a relatively high likelihood of responding (because they volunteered for inclusion in the panel).

CONCLUSION

We have shown how surveys regarding claimed/recalled survey participation must be adjusted to account for the influence of the distinctiveness of individual U.S. adults in their propensity to be contacted and willingness to participate. We have proposed a relatively simple probability mixture model that is consistent with the observed pattern of survey responses. Using this stochastic process, the degree of adjustment needed in accounting for nonresponse bias is quite large. For example, key summary statistics regarding survey participation in the U.S. adult population differ by 40% or (much) more from the raw statistics calculated from survey respondents. Although there may be substantial response errors in the data we have used to construct our estimates, we obtain similar results across two surveys that use different question wording and recall periods. The important point is that we estimate much lower participation rates than previous authors who have used the same data.

The model's projection of survey contact and participation leads to two conclusions. First, in contrast to suggestions by the popular press, U.S. adults as a whole simply are not being overwhelmed by the current level of survey activity. Second, U.S. survey participations are extraordinarily concentrated among a small subsegment of the total adult population. In a year, between 20 and 23% of adults are accounting for all survey responses, and a tiny 4% to 5% of adults account for more than half the survey responses. This concentration is much greater than would be expected from the randomness of survey contacts and an inclination to make survey contact proportional to income or other demographic factors of interest. Thus, the research industry may be burning out the small fraction of heavy responders. Future efforts should be directed toward expanding the scope of survey participation so as not to overwork and overwhelm this 5%.

APPENDIX

Homogeneous Models

Using Assumption 1, the number of survey contacts N in some time period of duration t is distributed Poisson with rate λ for each person, as follows:

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com
المنارة للاستشارات

(A1) $$P[N = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}; \; n = 0, 1, 2, \ldots,$$

with mean E[N] = λt and variance Var[N] = λt. Assumption 2 ensures that the distribution for the number of participations, X, for any person receiving N survey contacts is binomial:

(A2) $$P[X = x \mid N] = \binom{N}{x} p^x (1 - p)^{N-x}; \; x = 0, 1, \ldots, N,$$

with mean and variance E[X|N]= *p*N and Var[X|N] = *p*(1 – *p*)N. Finally, unconditional on the number of contacts, the number of survey participations in time t is distributed Poisson for any person with participation rate *p*λ:

(A3) $$P[X = x] = \frac{(p\lambda t)^x}{x!} e^{-p\lambda t}; \; x = 0, 1, 2, \ldots,$$

with mean and variance E[X] = pλt and Var[X] = *p*λt.

The BB/NBD Model

Survey contacts. As a result of Assumptions 1 and 3 (i.e., Equations A1 and 1), the number of survey contacts N encountered by a randomly chosen U.S. adult during a time period of length t (in years) follows the negative binomial distribution, as follows:

(A4) $$P_{NBD}[N = n \mid r, \alpha, t] = \binom{r + n - 1}{n} \left[\frac{\alpha}{(\alpha + t)}\right]^r \left[\frac{t}{(\alpha + t)}\right]^n; \; n = 0, 1, \ldots,$$

with mean and variance E[N] = rt/α and Var[N] = (rt/α) + (rt^2/α^2).

Participations, conditional on contacts. A person contacted N times with a survey request elects to participate X times, where X can range from 0, 1, ..., N. Conditional on our randomly chosen U.S. adult receiving N contacts, Assumptions 2 and 4 (Equations A2 and 2) ensure that the number of surveys in which this person participates follows the beta binomial distribution:

(A5) $$p_{BB}[X = x \mid N, a, b] = \binom{N}{x} \left[\frac{(a)_x (b)_{N-x}}{(a + b)_N}\right]; \; x = 0, 1, \ldots, N,$$

where $(a)_x$ is the increasing x-term factorial a(a + 1) ... (a + x – 1). The mean and variance of X conditional on N are E[N|X] = aN/(a + b) and

$$Var[X \mid N] = \left[\frac{Nab}{(a + b)^2}\right] + \left[\frac{N(N - 1)\, ab}{(a + b)^2 (a + b + 1)}\right].$$

Survey participation frequency. Of primary interest is the distribution for survey participation frequency unconditioned on the frequency of contact. This distribution for X unconditioned on N can be obtained by integrating the individual-level probability in Equation A3 over both the survey contact distribution (Equation 1) for λ and the survey participation propensity distribution (Equation 2) for *p*. The result, provided by Schmittlein, Bemmaor, and Morrison (1985), gives the distribution for survey participation frequency among randomly chosen U.S. adults, as follows:

(A6) $$P_{BBNBD}[X = x \mid r, \alpha, a, b, t] = \binom{r + x - 1}{x} \left[\frac{\alpha}{(\alpha + t)}\right]^r \left[\frac{t}{(\alpha + t)}\right]^x \left[\frac{(a)_x}{(a + b)_x}\right] {}_2F_1\left(x + r, b; x + a + b; \frac{t}{a + t}\right), \; x = 0, 1, 2, \ldots,$$

where $(a)_x$ = a(a + 1) … (a + x – 1), and ${}_2F_1(\,)$ is the Gauss hypergeometric function (Abramowitz and Stegun 1972, p. 558):

(A7) $${}_2F_1(a, b; c; z) = \frac{\Gamma(c)}{\Gamma(b)\Gamma(c - b)} \int_0^1 t^{b-1}(1 - t)^{c-b-1}(1 - zt)^{-a}\, dt, \; c > b.$$

This hypergeometric function can be evaluated using the algorithms in Luke's (1977) work. The mean and variance of X are as follows:

(A8) $$E[X \mid r, \alpha, a, b, t] = \left[\frac{rat}{\alpha\,(a + b)}\right], \text{ and}$$

(A9) $$Var[X \mid r, \alpha, a, b, t] = \left[\frac{rat}{\alpha(a + b)}\right] \left\{\frac{t}{\alpha\,(a + b + 1)} \left[\left(\frac{rb}{a + b}\right) + a + 1\right] + 1\right\}.$$

We call the distribution in Equation A6 the BB/NBD distribution.

One other quantity will prove useful, namely, the expected number of survey participations X* during some future time period for a person who has engaged in X surveys during the current period. (Both periods are assumed to be of length t.) Using Robbins's (1977) result, this is as follows:

(A10) $$E[X^* \mid X = x, r, \alpha, a, b, t] = \frac{(x + 1)\; P[x + 1 \mid r, \alpha, a, b, t]}{P[x \mid r, \alpha, a, b, t]}, \; x + 0, 1, 2, \ldots,$$

where P[x|r, α, a, b, t] are the BB/NBD probabilities listed in Equation A6.

REFERENCES

Abramowitz, M. and I.A. Stegun, eds. (1972), *Handbook of Mathematical Functions.* New York: Dover Publications Inc.

Brennan, Mike and Janet Hoek (1992), "The Behavior of Respondents, Nonrespondents, and Refusers Across Mail Surveys," *Public Opinion Quarterly*, 56 (Winter), 530–35.

CMOR (1995), *Respondent Cooperation Survey.* Port Jefferson, NY: Council of American Survey Research Organization (August).

Goyder, John (1986), "Surveys on Surveys: Limitations and Potentialities," *Public Opinion Quarterly*, 50 (Spring), 27–41.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com

——— (1987), *The Silent Minority: Nonrespondents on Sample Surveys*. Boulder, CO: Westview.

Greene, Jerome D. (1982), *Consumer Behavior Models for Non-Statisticians*. New York: Praeger.

Groeneman, Sid (1994), "Multi-Purpose Household Panels and General Samples: How Similar and How Different?" paper presented at the annual meeting of the American Association for Public Opinion Research, Danvers, MA (May 12–15).

Groves, Robert M. (1989), *Survey Errors and Survey Costs*. New York: John Wiley & Sons.

———, Robert B. Cialdini, and Mick P. Couper (1992), "Understanding the Decision to Participate in a Survey," *Public Opinion Quarterly*, 56, 475–95.

Johnson, Norman L. and Samuel Kotz (1970), *Continuous Univariate Distributions–2*. New York: John Wiley & Sons.

Kahn, Barbara E., Donald G. Morrison, and Gordon P. Wright (1986), "Aggregating Individual Purchases to the Household Level," *Marketing Science*, 5 (Summer), 260–68.

Karlin, Samuel and Howard M. Taylor (1975), *A First Course in Stochastic Processes*, 2d ed. New York: Academic Press, 221–28.

Luke, Yudell L. (1977), *Algorithms for the Computation of Mathematical Functions*. New York: Academic Press.

Morrison, Donald G. and David C. Schmittlein (1988), "Generalizing the NBD Model for Customer Purchases: What Are the Implications and Is It Worth the Effort?" *Journal of Business and Economic Statistics*, 6 (April), 145–66.

National Research Council (1979), *Privacy and Confidentiality as Factors in Survey Response*. Washington DC: National Academy of Sciences.

O'Neill, Harry W. (1996), "CMOR Refusal Rates and Industry Image Survey," paper presented at the annual meeting of the American Association for Public Opinion Research, Salt Lake City, UT (May 18).

Robbins, Herbert (1977), "Prediction and Estimation for the Compound Poisson Distribution," *Proceedings of the National Academy of Science, USA*, 74, 2670–71.

Roper, Burns W. (1986), "Evaluating Polls with Poll Data," *Public Opinion Quarterly*, 50 (Spring), 10–16.

Schleifer, Stephen (1986), "Trends in Attitudes Toward Participation in Survey Research," *Public Opinion Quarterly*, 50 (Spring), 17–26.

Schmittlein, David C. (1989), "Surprising Inferences from Unsurprising Observations: Do Conditional Expectations Really Regress to the Mean?" *The American Statistician*, 43, 176–83.

———, Albert C. Bemmaor, and Donald G. Morrison (1985), "Why Does the NBD Model Work? Robustness in Representing Product Purchases, Brand Purchases and Imperfectly Recorded Purchases," *Marketing Science*, 4 (Summer), 255–66.

———, Lee G. Cooper, and Donald G. Morrison (1993), "Truth in Concentration in the Land of (80/20) Laws," *Marketing Science*, 12 (Spring), 167–83.

Struebbe, Jolene M., Jerome B. Kernan, and Thomas J. Grogan (1986), "The Refusal Problem in Telephone Surveys," *Journal of Advertising Research*, 26 (3), 29–37.

U.S. Bureau of the Census (1995), (July), [available at http://www.census.gov/population/estimates/nation).

——— (1996), *Statistical Abstract of the United States*, 116th ed. Washington, DC: U.S. Bureau of the Census.

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission. www.manaraa.com