Cam Rungie[1],

Gilles Laurent[2],

Francesca Dall'Olmo Riley[3],

Donald G. Morrison[4],

Tirthankar Roy[5]

Measuring and Modeling the (Limited) Consistency of Free Choice Attitude Questions

The authors thank Albert Bemmaor and A.S.C. Ehrenberg for their comments on a previous version.

[1] School of Marketing, University of South Australia, North Terrace, Adelaide, SA, Australia. Telephone 61 (0)8 8302 0449, fax 61 (0)8 8302 0765, cam.rungie@unisa.edu.au
[2] Carrefour Professor of Marketing, HEC School of Management, 78350 Jouy-en-Josas, France. Telephone: 33 (0)1 3967 7480, fax: 33 (0)1 3967 7087, laurent@hec.fr
[3] School of Marketing, Kingston University Business School, Kingston Hill, Kingston-upon-Thames KT2 7LB, UK. Telephone : 44 (0) 20 8547 2000, fax: 44 (0) 20 8547 7026, f.d.riley@kingston.ac.uk
[4] The Anderson School at UCLA, 110 Westwood Plaza, Suite B515, Los Angeles, CA 90095. Telephone: (310)825 7706, fax: (310)206 7422, donald.morrison@anderson.ucla.edu.
[5] University of Michigan Business School, 701 Tappan Street, Ann Arbor, MI 48109-1234. Telephone: (734)763 6821 fax: (734)936 8715, tirroy@umich.edu

Measuring and Modeling the (Limited) Consistency of Free Choice Attitude Questions

ABSTRACT

On average, respondents who give a positive answer to a binary free choice attitude question are NOT more likely, if surveyed again, to respond positively than to response negatively. However, stronger brands obtain more repeated positive answers. Our model shows why these two effects have to happen, even though all brands in a category benefit from the same reliability.

INTRODUCTION

In free choice attitude questions, an often used format in market research, respondents are typically presented with a list of about ten brands, and are asked, for each of about ten attitude items, to which brands the item applies. Answers are recorded in binary form for each respondent: Is item i applicable to brand j?

Such questions have multiple uses, given the development of one-to-one Marketing:

- To measure the impact of a campaign (Is a consumer recently exposed to a specific communication more likely to give a positive answer to item i?).
- To enter in a multivariate data analysis (Which component of brand attitude is the best predictor of brand choice? Conversely, are buyers at normal price more likely to give a positive answer than buyers in promotion? How can we cluster consumers on the basis of their attitudes?).
- To target an action, such as, e.g., sending a sample of the new product form, or a brochure, to those respondents who say that the brand is "old-fashioned;" or sending a coupon to those who find it "expensive."

These analyses take at face value answers to free choice attitude questions. However, when repeated observations are available in which answers are recorded twice in separate interviews of the same consumers, then, on average, a positive answer by a respondent on a first interview is as likely to be followed, in a second independent interview, by a negative answer as it is to be followed by a repeated positive answer.

We propose, and validate, a model which explains why such disappointing results are not the result of some extraneous noise, but rather have to be expected, given the stochastic structure of the phenomenon; and which shows why, as the response level RL_{ij} (the proportion of respondents giving a positive response) increases, so does the repeat rate RR_{ij} (the proportion of those respondents giving a positive answer on the first interview who also give a positive answer on the second interview). In the process, we analyze measures of consistency and reliability for these questions. This leads to a series of warnings and recommendations regarding their use in Marketing practice.

CONSISTENCY: A MEASURE OF RELIABILITY

The reliability of a measure can be formally defined as the ratio of the variance of the construct being measured to the variance of the measure. Reliability is operationalized in this manner in structural equation methods (Bollen 1989). The reliability of a binary question can be measured following this principle.

We assume the process to be zero order and stationary. Each respondent n has a probability p_{ijn} of giving a positive answer on the applicability of item i to brand j. The respondent's observed answer is r_{ijn} where a positive answer is coded as $r_{ijn}=1$ and a negative one as $r_{ijn}=0$. The probability p_{ijn} is a realization of a latent random variable P_{ij} which has a distribution over the population. The response r_{ijn} is a realization of a manifest random variable R_{ij} which follows a random Bernoulli process mixed by $P_{ij.}$ The expected value of P_{ij} gives the probability π_{ij} of receiving a positive answer at the first interview from a consumer taken at random:

$$E\left[P_{ij}\right] = E\left[R_{ij}\right] = p_{ij}$$

The variance of the observed binary variable R_{ij} is:

$$Var\left(R_{ij}\right) = p_{ij}\left(1 - p_{ij}\right)$$

while the variance of the latent random variable P_{ij} is by definition:

$$Var\left[P_{ij}\right] = E\left[\left(P_{ij} - E\left[P_{ij}\right]\right)^2\right]$$

$$Var\left[P_{ij}\right] = E\left[P_{ij}^2\right] - \left(E\left[P_{ij}\right]\right)^2$$

If we view the manifest binary variable as a measure of the latent probability variable then, by definition, the ratio c_{ij} of these two variances of the two variables is the reliability of the measure.

$$c_{ij} = \frac{Var\left(P_{ij}\right)}{Var\left(R_{ij}\right)} = \frac{Var\left(P_{ij}\right)}{p_{ij}\left(1 - p_{ij}\right)}$$

We shall refer to c_{ij} as "consistency," as it directly quantifies the extent to which respondent answers are consistent. A consistency of 1 suggests the response is the “true” one, with no random variation in the binary response across successive interviews from the same respondent. A consistency of 0 indicates that respondents cannot be relied upon, that it is impossible to predict their personal answer next time on the basis of their previous answer. This would be the case, for example, if they tossed a coin each time to choose their response. The higher the consistency, the higher the reliability of the binary question and of the interview process. Market researchers obviously would like reliability to be as high as possible.

An alternative justification of c_{ij} as the "consistency" of the binary answer can be derived from considering the extreme values of the variance of P_{ij}. Its lower bound is zero. This corresponds to the case where all respondents have the same probability of giving a positive answer:

$$p_{ijn} = E[P_{ij}] = p_{ij} \qquad \text{n=1,N}$$

The variance of P_{ij} is obviously zero. There is no underlying difference between respondents. We cannot predict a respondent's second answer on the basis of the first answer. Respondents are *inconsistent*, as their answers are totally random, following a Bernoulli process. In contrast, the upper bound of the variance of P_{ij} corresponds to the case when a proportion p_{ij} of the population has a probability p_{ijn} of 100% of giving a positive answer, and the rest of the population (a proportion equal to 1 - p_{ij}) has a probability of 0% of such an answer. In that case, respondents are totally *consistent* with themselves. Every respondent strongly favors one of the two answers. We can predict perfectly their answer at the second interview on the basis of their answer at the first interview.

The variance here becomes:

$$Var(P_{ij}) = p_{ij}(1 - p_{ij})$$

We propose to standardize the variance of P_{ij} by dividing it by its maximum possible value, given the mean. This maximum is $(\pi_{ij}\ (1-\pi_{ij}))$. This produces an index of *consistency*, c_{ij}

$$c_{ij} = \frac{Var(P_{ij})}{\text{Maximum Variance}} = \frac{Var(P_{ij})}{p_{ij}(1 - p_{ij})}$$

which is always comprised between 0 and 1. It takes value 1 when respondents are maximally consistent, i.e. when each respondent repeatedly produces the same answer. It takes value 0 when a respondent’s answer in one survey gives no hint which helps estimate his or her response in the next survey.

Of course, this is identical to the definition of c_{ij} given earlier as a measure of reliability. Consistency can be justified in two manners: as the reliability of the binary measure, as the standardized variance of P_{ij} across respondents.

REPEAT RATES

The response level RL_{ij} is the proportion π_{ij} of positive answers in a single interview, i.e. the proportion $\pi_{ij}=E[P_{ij}]$ of respondents who indicate that item i applies to brand j. Let the proportion who give positive answers on both interviews be $\rho_{ij}=E[P_{ij}^2]$. Let the “repeat rate” (RR) be the proportion, φ_{ij}, of those respondents who gave a positive response on the first interview who again give a positive response on the second interview. Thus $\varphi_{ij} = \rho_{ij} / \pi_{ij}$.

There is a useful relationship, for each brand-item pair *ij*, between the consistency, the response level and the repeat rate. As:

$$c_{ij} = \frac{\mathrm{Var}\left(P_{ij}\right)}{\pi_{ij}\left(1-\pi_{ij}\right)} = \frac{E\left[P_{ij}^2\right]-\left(E\left[P_{ij}\right]\right)^2}{\pi_{ij}\left(1-\pi_{ij}\right)} = \frac{\rho_{ij}-\pi_{ij}^2}{\pi_{ij}\left(1-\pi_{ij}\right)}$$

$$\rho_{ij} = c_{ij}\pi_{ij} + (1-c_{ij})\pi_{ij}^2$$

and

$$\varphi_{ij} = c_{ij} + (1-c_{ij})\pi_{ij}$$

Of course, this is a logical identity only for a given ij pair. However, if consistency were identical for all brands i and items j in a given category, i.e. if c_{ij} took the same value, call it c, for all ij, then we would have a simple relationship:

$$j_{ij} = c + (1-c)p_{ij}$$

or:

$$RR = c + (1 - c)\, RL$$

We shall test this model later.

EMPIRICAL DATA SHOW ONLY A MODERATE CONSISTENCY

We summarize briefly previous empirical results on respondent consistency (Castleberry et al. 1994; Dall'Olmo Riley et al. 1997; Dall'Olmo Riley, Ehrenberg and Barnard 1998). For eight product categories, in two countries (Table 1), free choice attitude questions were asked twice from each respondent (the interval between the two surveys varying from one month to eighteen months). Over all brands and items, 28% of respondents gave a positive answer in the first survey (average RL). Average results per category are presented in Table 1. The repeat rates (RR) show a moderate consistency: Over all brands and all items, the average RR is 49 %. Furthermore, repeat rates vary markedly across items and brands, and increase, as could be expected[i], with the response level. Brand-items with a high RL, say 60%, have a high RR, around 60% or 80%. Brand-Items with a low RL, say 20%, have a low RR, between 20% and 60%.

TABLE 1 ABOUT HERE

TWO COMPETING MODELS

Several previous papers (Castleberry et al. 1994; Dall'Olmo Riley et al. 1997; Dall'Olmo Riley, Ehrenberg and Barnard 1998; Rungie and Dall'Olmo Riley 2000) proposed a simple model of the relationship between the answers in the first and second surveys:

Model 1 $$RR = RL + 20\,\%$$

While this model incorporates the qualitative relationship, is concise, and useful as a first approximation, it has three major drawbacks. First, it is still far from a perfect prediction, with marked variations across the straight line of the model. Second, some predicted values are not logically consistent: When RL is 90%, the predicted RR is 110%. Third, it is purely empirical, not being based on an assumed underlying stochastic model. While previous papers identified an empirical regularity (Dall'Olmo Riley et al., 1997, Ehrenberg, 1995), we wish to propose a better probabilistic model of that empirical regularity. "Better" means that the model structure should not lead to inconsistent predictions, such as probabilities above 100%, and that it should provide a better statistical fit.

We therefore propose an alternate model, namely that consistency c_{ij} takes the same value c for all ij pairs in a product category. This is a sensible hypothesis, as questions on the different brands and items are asked at about the same time, using the same format, from the same respondents, and therefore can be reasonably hypothesized to have the same reliability. According to this new model:

Model 2 $$RR = c + (1 - c)\, RL$$

This ensures that the predicted RR can never be above 100%.

Model 1 and Model 2 are not compatible. We shall evaluate empirically which one fits best. Before that, we show that, although neither model assumes a stochastic distribution, both are compatible with the assumption of an underlying beta distribution.

ASSUMING A BETA DISTRIBUTION

Assume that P_{ij} follows a beta distribution with parameters α_{ij} and S_{ij} (Johnson et al. 1994a; Johnson et al. 1994b; Johnson et al. 1993; Lilien, Kotler and Moorthy 1992). Then, for a randomly selected respondent, the probability of getting, at the first interview, the positive answer that item i applies to brand j is:

$$p_{ij} = E[P_{ij}] = \frac{a_{ij}}{S_{ij}}$$

And the probability of getting a positive answer on both interviews is:

$$r_{ij} = E[P_{ij}^2] = \frac{a_{ij}(a_{ij}+1)}{S_{ij}(S_{ij}+1)}$$

Thus, the variance of P_{ij} is:

$$Var(P_{ij}) = E[P_{ij}^2] - (E[P_{ij}])^2$$

$$Var(P_{ij}) = \frac{a_{ij}(a_{ij}+1)}{S_{ij}(S_{ij}+1)} - \left(\frac{a_{ij}}{S_{ij}}\right)^2$$

$$Var(P_{ij}) = \frac{a_{ij}(S_{ij}-a_{ij})}{S_{ij}^2(S_{ij}+1)}$$

$$Var(P_{ij}) = p_{ij}(1-p_{ij})\frac{1}{S_{ij}+1}$$

Hence, in this special case where P_{ij} follows a beta distribution, the reliability c_{ij} of the binary question in the interview process is:

$$c_{ij} = \frac{Var(P_{ij})}{p_{ij}(1-p_{ij})} = \frac{1}{S_{ij}+1}$$

The statistic $1/(1+S_{ij})$ is an important characteristic of the beta distribution. It is often called the "coefficient of polarization" (Sabavala and Morrison 1977, Kalwani and Morrison 1980). Remember that c_{ij} had been defined without reference to a specific distribution. In the special case where P_{ij} follows a beta distribution, then c_{ij} takes as its value the coefficient of polarization.

Note that Kalwani and Silk (1982), following Morrison (1979), have shown that the reliability of a beta-binomial process, where one models the sum of n independent drawings from a binary variable, where the underlying probabilities are distributed beta over the population, is given by:

$$\frac{n}{a+b+n}$$

where $\alpha + \beta$ is the notation corresponding to our notation S_{ij}. When there is only one answer recorded (n = 1), this reduces to:

$$\frac{1}{S_{ij}+1}$$

i.e. to the definition of consistency.

In a previous paper, Morrison and Roy (1996) remarked that the simple model ("Model 1" above) proposed in previous papers (RR = RL + 20%), combined with a beta distribution, leads to a very specific hypothesis regarding S_{ij}. If the process follows a beta distribution, then the expected response level p_{ij} is given by α_{ij} / S_{ij}, and the repeat rate f_{ij} is equal to $(1 + \alpha_{ij}) / (1+S_{ij})$. A simple manipulation leads to:

p_{ij}

$$j_{ij} = p_{ij} + (1 - p_{ij}) \frac{1}{S_{ij} + 1}$$

This is very similar to the empirical result (RR = RL + 20%). If indeed the last element $(1 - p_{ij}) \frac{1}{S_{ij} + 1}$ in the equation is a constant (20%), then it follows that when the response level p_{ij} is larger, S_{ij} has to be smaller. To quote Morrison and Roy (1996), "strong attributes are more polarized." Different brand-items should have different values of S_{ij} and a brand attribute ij that gets a high score RL_{ij} at the first interview should have a smaller value of S_{ij}.

In contrast "Model 2" assumes that c_{ij} takes the same value for all pairs ij. In the framework of a beta distribution, since $c_{ij} = 1 / (1 + S_{ij})$, this leads to a second model where all ij pairs should have the same value S for the parameter S_{ij}. The two models can be presented as hypotheses regarding the coefficients of the beta distribution. We assess now, by an empirical analysis, which one of the two hypotheses stands better to the data.

VALIDATING THE STOCHASTIC MODEL

As indicated above, we have eight data sets, with about one hundred observations per set (answers to about ten items for about ten brands). Each observation comprises two values, p_{ij} and r_{ij}, the empirical observations corresponding to two theoretical values π_{ij} and ρ_{ij}. π_{ij} is the probability of getting, at the first interview, the positive answer that item i applies to brand j. ρ_{ij} is the probability of getting this positive answer on both interviews. The associated empirical data are:

p_{ij} the empirical frequency of positive answers to item *i* about brand *j*

r_{ij} the empirical freque ncy of a double positive answer to item *i* about brand *j*

We use these empirical data to check the validity of the two models proposed above.

According to Model 1, we should have $RR_{ij} = RL_{ij} + 0.20$ and $\varphi_{ij} = \pi_{ij} +.20$. As $\varphi_{ij} = \rho_{ij} / \pi_{ij}$:

Model 1 $$\rho_{ij} = \pi_{ij}^{2} + 0.20\, \pi_{ij}$$

Thus, if Model 1 holds, the probability ρ_{ij} of observing twice a positive answer should be a quadratic function of π_{ij}, the probability of observing a positive answer at the first interview. Furthermore, the coefficient of the first-degree term should be 0.20, the coefficient of the second-degree term should be 1, and there should be no constant term in the equation.

Model 2 assumes the same c for all pairs ij:

Model 2 $$r_{ij} = (1-c)p_{ij}^{2} + cp_{ij}$$

Thus, if Model 2 holds, ρ_{ij} , the probability of observing twice a positive answer, should be a quadratic function of π_{ij}, the probability of observing a positive answer at the first interview. Furthermore, the coefficients of the first-degree term and of the second-degree term should add to one, and there should be no constant term in the equation.

In summary, the two models lead to the following predictions.

	Model 1	Model 2
Constant term	0	0
First-degree term	0.20	c (between 0 and 1)
Second-degree term	1	1 - First-degree term

In the statistical analysis, we replace the theoretical terms ρ_{ij} and π_{ij} by the corresponding empirical observations, r_{ij} and p_{ij}. Rather than constraining the coefficients on the basis of the theoretical predictions, we use an ordinary regression model, in order to check whether the empirical results confirm the theoretical predictions. We run a separate regression for each of the eight data sets, as well as a global regression on the pooled data. Figure 1 displays the data, and fitted regressions, on the eight data sets. Detailed results are presented in Table 2.

TABLE 2 AND FIGURE 1 ABOUT HERE

Overall, the fit is excellent, as is visible in Figure 1. The R^2, for each data set, is above 0.982, with an average value of 0.989. Even when pooling the data, we obtain excellent results, with an R^2 of 0.984. The constant terms are close to zero, which is the result predicted by both models. The totals of the first-order and second-order coefficients are close to one, which is in conformity with the second model. The first degree coefficient varies from 0.322 to .452, and is therefore very significantly above the value of 0.200 predicted by the first model. The second degree coefficient varies from 0.554 to 0.771, and is therefore very significantly below the value of 1.000 predicted by the first model. These results support the validity of the beta distribution. They fit well with Model 2, not with Model 1. In the remaining of the paper, we therefore use Model 2. Respondent answers can be modeled as a zero-order beta process, with a constant S (and c) over all brands i's and items j's in a category.

POOLED OR CONSTRAINED ESTIMATION?

We formally test our model by performing a series of F tests, along two axes. First, we can compute a single estimate over all our pooled data, or compute separate estimates for each one of the eight data sets. Second, we can either estimate a three parameter model:

$$r_{ij} = \mathrm{b}_0 + \mathrm{b}_1 \, p_{ij} + \mathrm{b}_2 \, p_{ij}^2$$

or a constrained equation, corresponding to the second model, in which the constant is constrained to be zero, and the total of the two other coefficients is constrained to be one:

$$r_{ij} = \mathrm{c}\ p_{ij} + (1 - c)\ p_{ij}^2$$

Figure 2 presents the results. It gives, for each of the four estimation methods, the Sum of Squared Residuals (SSR) and R^2. It also gives the results of the four formal F tests (Fisher, 1970) comparing these estimations. All four tests indicate significant differences. However, the increases in fit obtained when passing from pooled estimates to separate estimates for each data set are much larger (F tests of 60.8 and 36.4) than the increases obtained when passing from a constrained one-parameter estimations to free three-parameter estimations (F of 15.3 and 7.58).

From this we derive two conclusions. First, while the results obtained in different categories are similar, they are significantly different. We recommend estimating different coefficients for different product categories. The pooled estimates are mostly useful to give a first order approximation, an order of magnitude of the reliability of free choice attitude data, as illustrated later. Second, while three parameter models indeed fit significantly better than one parameter models, the improvement is very modest. Taking into account the very high R^2 of one parameter models (0.984 for pooled data, 0.989 when making separate estimates for each data set), we conclude that our proposed model (for all brands and all items within a category, reliability is constant, as too is the S statistic for the beta distribution, across respondents, of the probability of giving a positive answer) is empirically supported.

TABLE 3 ABOUT HERE

Table 3 provides the estimates of c, the consistency, obtained on the basis of each of the data sets, and on the basis of pooled data, using the constrained model. As indicated above, values of consistency are similar, mostly between 0.25 and 0.35, while category 6 seems to constitute an exception. These figures show clearly that c is much below one, i.e. that free choice attitude questions are very far from being reliable.

DISCUSSION

Our results can be compared to the simple model suggested in previous papers (RR = RL + 20 %). Our analysis leads to a somewhat different equation. On the basis of the pooled data, we obtain an estimate for c equal to 0.325. This leads to:

$$RR = 0.675\ RL + 32.5\ \%$$

The repeat rate RR varies in a consistent, non-brand specific manner. Since the coefficient is very significantly positive (0.675), the repeat rate is indeed higher for brands with a higher initial response level RL. However, the increase in repeat rate is smaller than the increase in initial response level (0.675 is significantly below 1). Of course, in absolute terms, the difference between the increase in initial response level (1), and the increase in repeat rates (0.675) is mostly visible for brands with high initial response levels. Overall, we have here one more example of a size effect of a "double jeopardy" kind, with smaller response levels RL also being repeated less often (McPhee 1963; Ehrenberg 1988).

In many ways the more interesting question now is "Why does consistency vary between categories?" rather than "Why is it constant within categories?" Presumably reliability is a function of consumers relationship with the whole category rather than with the individual brands (or attributes) within the category.

A COUNTER-INTUITIVE ASPECT OF CONSISTENCY AS A FORM OF RELIABILITY

Finally, it is useful to make a distinction between two closely related concepts: Reliability (the ratio of true variance divided by observed variance, as indicated above), and the probability of obtaining similar values from repeated measures on the same individual. We contrast the case of binary questions, such as the ones we study in this paper, against the case of an interval scale measure.

Consider an interval scale measure, an answer X_{ijn} given by respondent n about brand j and item i. It is often described as the sum of a true general average μ_{ij} (over all respondents) plus a true deviation τ_{ijn} of respondent n from that general average (a constant for respondent n, a random variable when one draws a respondent at random from the population), plus a random component ε_{ijn} that may vary, for the same respondent, on successive interviews:

$$X_{ijn} = m_{ij} + t_{ijn} + e_{ijn}$$

Common assumptions are that the τ_{ij}'s have the same variance across respondents for all ij pairs, that the ε_{ijn}'s have the same variance for all i, j and n's, and that the τ_{ij}'s and ε_{ijn}'s are uncorrelated. This leads to two consequences. First, the observed population means have the same variance for all ij pairs, whether the specific ij pair has a high or a low expected score. Second, for any specific ijn case, the variance is the same, measured by indices such as the standard deviation of the observed score X_{ijn} around the true score $\mu_{ij} + \tau_{ijn}$, or the distribution of the difference between repeated measures on the same person. Thus, whether an individual has a low or a high true score $\mu_{ij} + \tau_{ij}$, the variance created by the random component ε_{ijn} is the same. For a given level of reliability (a given ratio of τ and ε variance over μ variance), all respondents have the same level of individual variance.

For a binary question like the ones we study in this paper, the situation is different. If we assume that, within a product category, consistency is constant across brand-item pairs (an hypothesis supported by our empirical data), this nevertheless produces contrasted results. All ij pairs have the same consistency c, i.e., the same ratio of observed variance divided by the maximum possible variance. However, a brand-item pair ij with a high probability of a positive response at the first interview also has a high probability of the positive answer being repeated. And a brand-item pair ij with a low probability of a positive response at the first interview has, in addition, a low probability of the positive answer being repeated. The probability of repeating a positive answer is better for stronger brands, and worse for weaker ones. As indicated above, this is a case of double jeopardy.

CONSUMER HETEROGENEITY AND RESPONSE DISCRIMINATION: ORDERS OF MAGNITUDE

The estimate based on pooled data (c=0.325) provides a useful and reliable estimate of the order of magnitude of the phenomenon. Table 4 gives, for different values of the initial response level, the expected frequency of a double positive answer. This illustrates the typically moderate consistency of answers to these free choice attitude questions. When RL is 50 %, RR is only 66 %. And only 33 % of the respondents will give twice a positive answer (versus, again, 50 % who give a positive answer on the first interview and 50% who give a positive answer on the second interview). This is not a problem if one is interested mostly in the aggregate percentage. But it may become a major drawback if one wants to use individual answers for further statistical analyses (e.g. for a cluster analysis or for predicting overall preference on the basis of perceived brand attributes) or to target marketing actions (such as targeting a promotion to respondents as a function of their answers).

TABLE 4 ABOUT HERE

Figure 3 illustrates this order of magnitude by displaying the distribution of the values of P_{ijn} across respondents, for the value of c derived from the pooled data (c=0.325) and an hypothetical value of π_{ij} (0.5). It shows that the average response level of 50% indeed corresponds to a modal value of P at 0.5. However, the density hardly declines when P values move away from 0.5. The density becomes somewhat lower only for values of P around 0 (respondents who are very unlikely to say that the item applies to the brand) and, symmetrically, for values of P around 1 (respondents who are very likely to say that the item applies to the brand). Thus there are some respondents who would be very consistent in their answers over time and there are some who would not be. But the number of very consistent respondents is smaller than the number of people in the middle, which are the ones who lower the reliability of binary data.

It is interesting to contrast results for this "typical" value (c = 0.325) derived from pooled data with those from the most consistent category (soups, c = 0.484). We have plotted both density curves, at the same scale, on Figure 3a, for an hypothetical brand-item with a high average response level (π_{ij} = .5). The increased reliability in the soup data is generated by the presence of higher numbers of very consistent respondents (half of them very consistent in giving a positive answer, half of them very consistent in giving a negative answer). Compared to the curve derived from pooled data, there are many more consistent respondents with P values close to zero or one and quite fewer inconsistent respondents with P values "in the middle." Figure 3b offers a similar plot, but for an hypothetical brand-item with a low average response level (π_{ij} = .1). Here, the two density curves have similar shapes (inverted J's). The higher consistency level for soups indeed leads to a higher density of high-probability respondents, compared to the curve derived from pooled data. However, both densities are, in absolute terms, very small. This shows an important aspect of the "double jeopardy" that handicaps weak brands: There are very few respondents with a high (close to 1) probability of giving a positive answer. So, when a respondent by chance gives a positive answer, he or she does it typically IN SPITE OF a low personal probability. And he or she is therefore NOT LIKELY to repeat the positive answer, if asked again the same question. In contrast, for a strong brand, when a respondent gives a positive answer, he or she does it typically BECAUSE of a high personal probability. And he or she is therefore LIKELY to repeat the positive answer, if asked again the same question.

This leads to an important question: How discriminating is a positive answer? In other words, if an analyst observes such an answer in an interview, how sure can she be that she would have observed the same answer in another interview? Those consumers who give a positive answer are a selective sample, but how selective? P represents the probability of a positive response from a consumer randomly selected from the population of all consumers. Let the random variable Q represent the probability of another positive response from a consumer chosen randomly from the selective sub-population of consumers who have given a positive answer in the first interview. If P has a beta distribution with parameters α and S-α, then Q also has a beta distribution but with parameters α+1 and S-α. This is in accordance with Model 2 above and E[Q]=c+(1-c)p   This conditional distribution is useful in evaluating the use of dichotomous questions for screening, such as in disaggregate analysis of market research surveys and in direct marketing. Figure 4a is based on the consistency value derived from pooled data (c=0.325), assuming a high average response level (π_{ij} = .5). It plots the density curves for the population (identical to the density in Figure 3a), for the probability of consumers giving a positive answer in a single interview. It also plots the density curve for the probability of consumers giving a positive response in a second interview, given they gave a positive response at the first interview. Figure 4b displays a similar plot, for the same consistency value (c=0.325), assuming a low average response level (π_{ij} = .1). There is a striking contrast. For a strong brand (π_{ij} = .5), consumers who have given a positive answer constitute a very specific segment with a high probability of giving a positive response at the second interview. This repeat rate is much above 50% on average. For a weak brand (π_{ij} = .1), respondents who have given a positive answer do not constitute a specific segment, they have probabilities spread over all levels. Overall, consumers who have given a positive answer at the first interview have a probability lower than 50% of giving another positive answer at the second interview. This is another aspect of double jeopardy: For a weak brand, a positive answer does not say much about the respondent who gave it, as it is most likely to come from a consumer who had a small probability of giving it. For a strong brand, a positive answer is a good, discriminating signal, as it is very likely to come from a consumer with a high probability of giving it.

CONCLUSION

Our results indicate that the consistency of free-choice attribute questions is constant within a product category, but not between product categories. The implications of this result are substantial to market research. First, positive answers recorded about a weak brand provide a much less precise respondent characterization than identical answers recorded about a strong brand. Second, it is possible for market researchers to anticipate the consistency of their attitudinal questions without taking repeated observations on the full sample. We strongly suggest a systematic assessment of the magnitude of the phenomenon, using a double interview with a small sample. Once an order of magnitude is established, it can be used to assess, before a survey is completed, to what extent each individual answer can be relied upon.

This paper illustrates the usefulness of consistency as a measure of reliability for binary variables. While our results strictly apply to the data we analyze (free choice attitude questions), they suggest that, similarly, reliability could be moderate for other types of binary attitude questions, such as stated preferences, or self-reported purchases. Also, Ehrenberg and his co-workers (Goodhardt, Ehrenberg and Chatfield 1984; Ehrenberg 1988; Uncles, Ehrenberg and Hammond 1995) have shown repeatedly that the beta distribution and its multivariate extensions describe well actual purchase behavior observed in a consumer panel.

Finally, useful research could be undertaken to explain why consistency changes between categories.

Table 1

AVERAGE REPEAT RATES ARE LOW

Data set	Product	Country	N of brands	N of items	Sample item	N of observations	**Average repeat rate**
1	Cereal	UK	8	12	Crispy in milk	8x12=96	**45 %**
2	Cereal	US	11	11	Low in sugar	11x11=121	**38 %**
3	Fast food	US	7	12	Wide menu	7x12=84	**60 %**
4	TV news networks	US	7	13	Fast coverage	7x13=91	**59 %**
5	Washing powder	US	9	11	Gentle to clothes	9x11=99	**44 %**
6	Soups	UK	8	12	Rich and thick	8x12=96	**50 %**
7	Toothpaste	UK	8	12	Refreshing taste	8x12=96	**51 %**
8	Washing powder	UK	9	13	Gets stains out	9x13=117	**47 %**
Pooled	*	*	*	*	*	800	**49 %**

Table 2
EMPIRICAL OLS ESTIMATES OF THE QUADRATIC RELATIONSHIP BETWEEN ρ AND π

Data set	N	R^2	F	Constant (t)	First Degree Coefficient (t)	Second Degree Coefficient (t)	Total of two previous columns
1	96	0,992	5662	-0,0094 (t=-2,93)	0,348 (t=13,69)	0,727 (t=17,66)	1,075
2	121	0,985	3933	-0,0071 (t=-2,30)	0,322 (t=12,80)	0,654 (t=13,38)	0,976
3	84	0,992	4765	-0,0292 (t=-2,39)	0,403 (t=7,45)	0,614 (t=11,49)	1,017
4	91	0,986	3199	-0,0170 (t=-1,73)	0,331 (t=7,08)	0,707 (t=13,11)	1,038
5	99	0,989	4370	-0,0080 (t=-1,76)	0,301 (t=9,68)	0,709 (t=16,35)	1,010
6	96	0,997	13601	-0,0048 (t=-2,14)	0,407 (t=20,22)	0,771 (t=25,18)	1,171
7	96	0,982	2534	-0,0209 (t=-2,48)	0,452 (t=7,53)	0,554 (t=5,62)	1,006
8	117	0,988	4619	-0,0074 (t=-2,71)	0,394 (t=14,86)	0,743 (t=13,66)	1,137
Pooled data	800	0,984	24319	-0,0062 (t=-3,39)	0,353 (t=28,94)	0,652 (t=38,66)	1,005

Table 3

ESTIMATES OF CONSISTENCY BASED ON EACH OF THE DATA SETS, AND ON POOLED DATA

Data set	Estimate of consistency *c*	Standard error
1	.343	.008
2	.269	.010
3	.287	.009
4	.293	.008
5	.264	.008
6	.484	.013
7	.349	.006
8	.403	.007
Pooled data	0.325	.004

Table 4
ORDERS OF MAGNITUDE:
PROBABILITIES OF A DOUBLE POSITIVE ANSWER,
AND OF REPEATING A POSITIVE ANSWER,
AS A FUNCTION OF THE INITIAL RESPONSE LEVEL
(ESTIMATES BASED ON POOLED DATA)

Initial response level	Probability of repeat	Probability of a double positive answer
0%	33%	0%
10%	39%	4%
20%	46%	9%
30%	53%	16%
40%	60%	24%
50%	66%	33%
60%	73%	44%
70%	80%	56%
80%	87%	69%
90%	93%	84%
100%	100%	100%

Figure 1
EMPIRICAL RELATIONSHIPS BETWEEN THE RESPONSE LEVEL P_{IJ} AND THE PROBABILITY R_{IJ} OF A REPEATED POSITIVE ANSWER

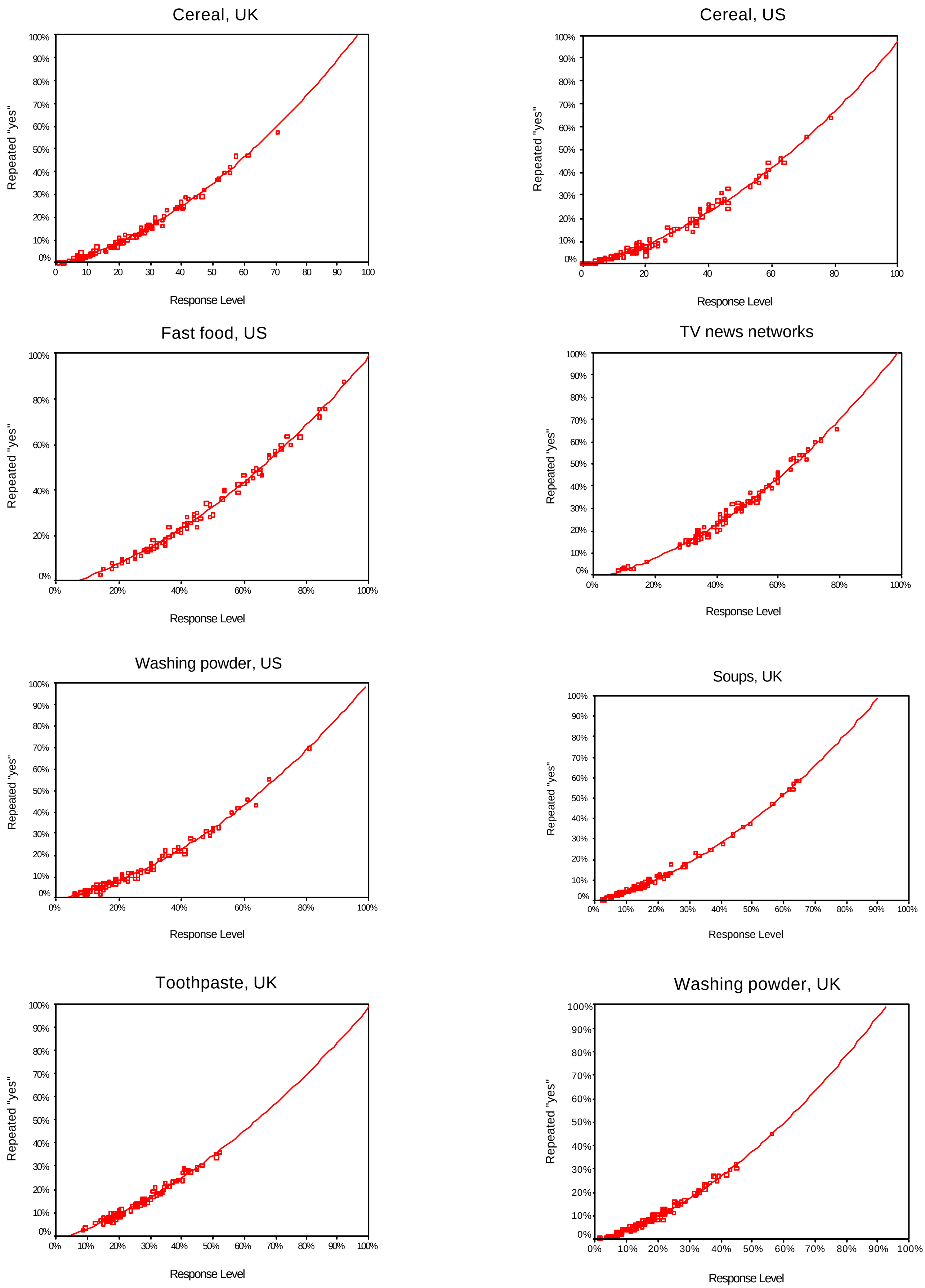


Figure 2

COMPARING FOUR SETS OF ESTIMATES

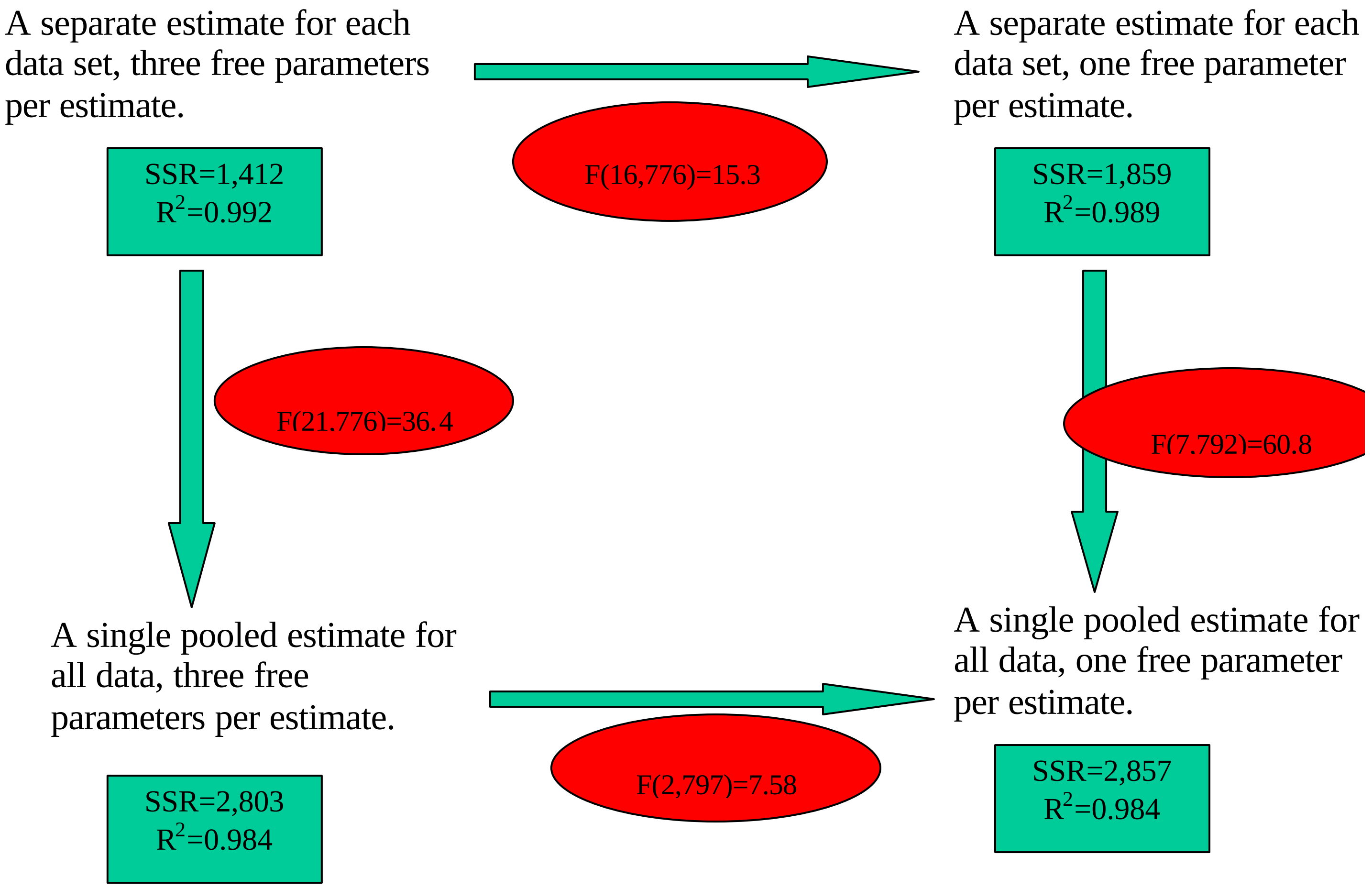


Figure 3a

VISUALIZING THE DISTRIBUTION OF P_{IJN} OVER RESPONDENTS
FOR A HIGH AVERAGE RESPONSE LEVEL (π_{IJ}=0.5)
FOR A TYPICAL CONSISTENCY VALUE BASED ON POOLED DATA
(c=0.325)
AND FOR THE DATA SET WITH HIGHEST CONSISTENCY
(c=0.484)

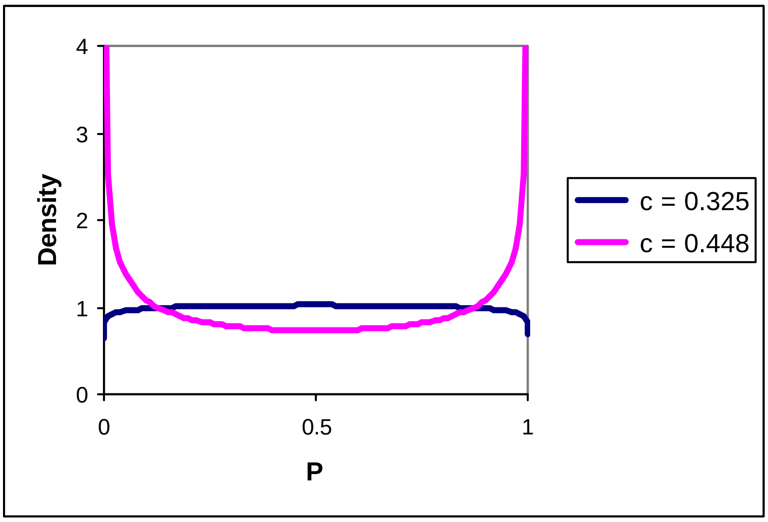


Figure 3b

VISUALIZING THE DISTRIBUTION OF P_{IJN} OVER RESPONDENTS
FOR A LOW AVERAGE RESPONSE LEVEL (π_{IJ}=0.1)
FOR A TYPICAL CONSISTENCY VALUE BASED ON POOLED DATA
(c=0.325)
AND FOR THE DATA SET WITH HIGHEST CONSISTENCY
(c=0.484)

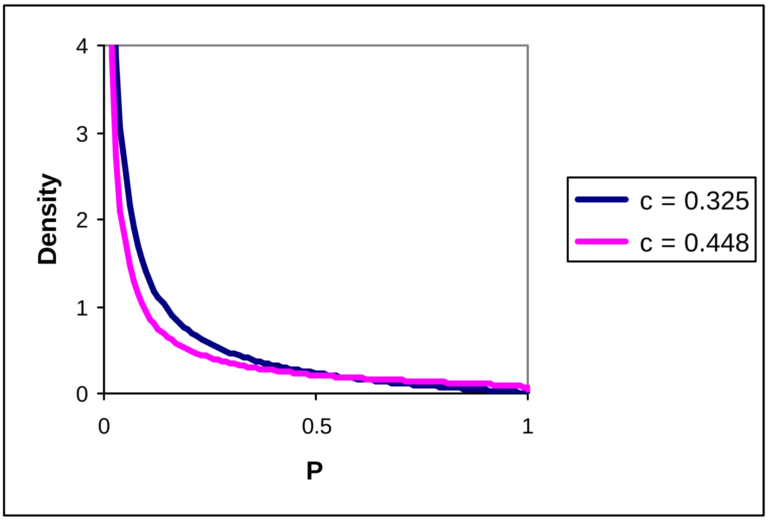


FIGURE 4a

DISTRIBUTION OF P FOR THE POPULATION,

FOR CONSUMERS GIVING A POSITIVE ANSWER AT A SINGLE INTERVIEW,

AND FOR CONSUMERS GIVING A POSITIVE ANSWERS AT A SECOND INTERVIEW,

CONDITIONAL ON HAVING GIVEN A POSITIVE ANSWER AT THE FIRST INTERVIEW

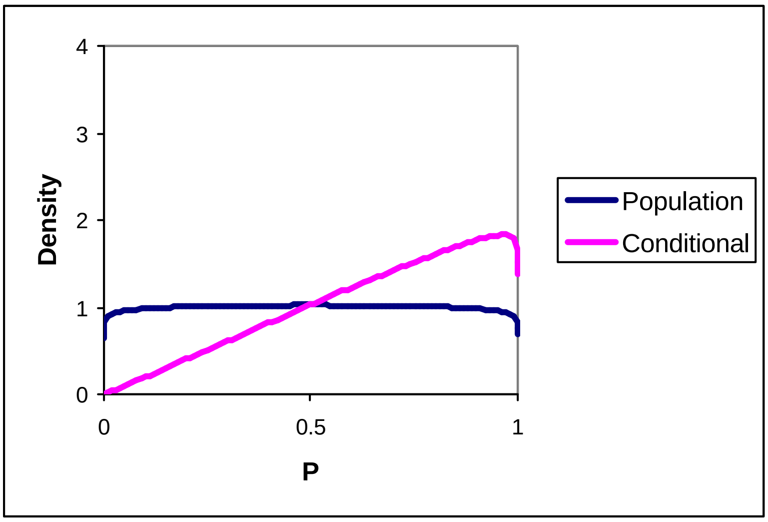


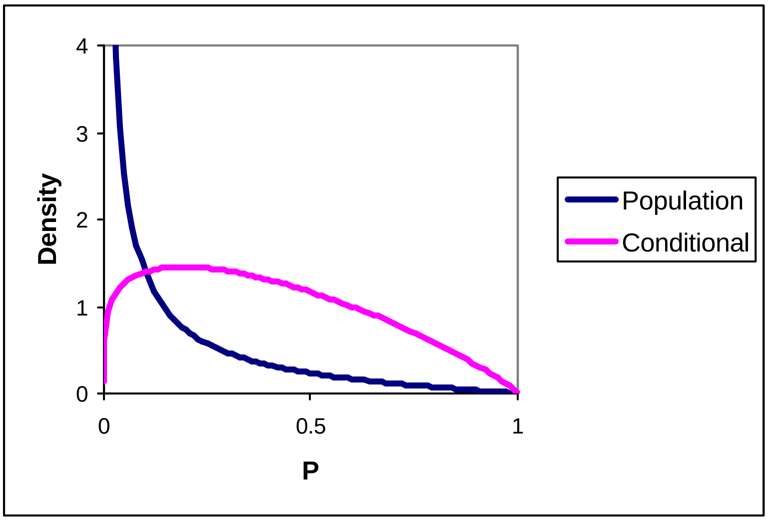
FIGURE 4b

DISTRIBUTION OF P FOR THE POPULATION,

FOR CONSUMERS GIVING A POSITIVE ANSWER AT A SINGLE INTERVIEW,

AND FOR CONSUMERS GIVING A POSITIVE ANSWERS AT A SECOND INTERVIEW,

CONDITIONAL ON HAVING GIVEN A POSITIVE ANSWER AT THE FIRST INTERVIEW



REFERENCES

Bass, Frank M., Abel Jeuland, and Gordon P. Wright (1976), “Equilibrium Stochastic Choice and Market Penetration Theories: Derivations and Comparisons,” Management Science, 22 (10), 1051-63.

Castleberry, S. B., Neil R. Barnard, Andrew S.C. Ehrenberg and Francesca Dall'Olmo Riley (1994), “Individual Attitude Variation Over Time,” Journal of Marketing Management, 10, 153-62.

Cronbach, Lee J. (1951), "Coefficient Alpha and the Internal Structure of Tests," Psychometrika, 16, 297-334.

Dall'Olmo Riley, Francesca, Andrew S. C. Ehrenberg and Neil R. Barnard (1998), "Changes in Attitudes and Behavior," 27th European Marketing Academy Conference Proceedings, Stockholm, Stockholm School of Economics.

Dall'Olmo Riley, Francesca, Andrew S. C. Ehrenberg, S.B. Castleberry, T. Patrick Barwise and Neil R. Barnard (1997), “The Variability of Attitudinal Repeat-Rates,” International Journal of Research in Marketing, 14: 437-450.

Ehrenberg, Andrew S. C. (1988), Repeat Buying: Facts, Theory and Applications, 2d ed. New York: Oxford University Press.

Ehrenberg, Andrew S.C. (1995), "Empirical Generalizations, Theory, and Method," Marketing Science, 14(3, Part 2 of 2), G20-G28.

Fisher, Frank M. (1970),"Tests of Equality Between Sets of Coefficients in Two Linear Regressions: An Expository Note," Econometrica, 38, 361-65.

Goodhardt, Gerald J., Andrew S. C. Ehrenberg and Christopher Chatfield (1984), “The Dirichlet: A Comprehensive Model of Buying Behaviour,” Journal of the Royal Statistical Society, Series A, 147(Part 5), 621-55.

Kalwani, Manohar U. and Donald G. Morrison (1977). “A Parsimonious Description of the Hendry System,” Management Science, 23(5), 467-77.

Kalwani, Manohar U. and Donald G. Morrison (1980), “Sample Size Requirements for Zero-Order Models,” Journal of Marketing Research, 17, 221-27.

Kalwani, Manohar U. and Alvin J. Silk (1982), "On the Reliability and predictive Validity of Purchase Intention Measures," Marketing Science, 1, 243-86.

Lilien, Gary L., Philip Kotler and K. Sridhar Moorthy (1992), Marketing Models. Englewood Cliffs, NJ: Prentice Hall.

McPhee, William N. (1963), Formal Theories of Mass Behavior. New York: The Free Press of Glencoe.

Morrison, Donald G. (1979), "Purchase Intentions and Purchase Behavior," Journal of Marketing, 43 (Spring), 65-74.

Rungie, Campbell and Francesca Dall'Olmo Riley (2000), "Response Patterns for Attributes," ANZMAC Conference, Gold Coast, Queensland, Australia, Griffith University.

Sabavala, Darius and Morrison, Donald G. (1977), "A Model of TV Show Loyalty," Journal of Advertising Research, 17(6), 35-43.

Uncles, Mark, Andrew Ehrenberg, and Kathy Hammond (1995), "Patterns of Buyer Behavior: Regularities, Models and Extensions," Marketing Science, 14 (3 Part 2), G71-G78.

ENDNOTES

[i] If respondents are heterogeneous, then, for any level of heterogeneity, the conditional probability of a positive answer on the second interview, given a positive answer to the first interview, is always higher than the unconditional probability of a positive answer. So, when the unconditional probability increases, we expect the conditional probability to also increase.