Prediction beyond the Survey Sample: Correcting for Survey Effects on Consumer Decisions*

Christiaan Heij† and Philip Hans Franses‡
Econometric Institute
Erasmus University Rotterdam

Econometric Institute Report EI 2006 - 48

November 20, 2006

Abstract

Direct extrapolation of survey results on purchase intentions may give a biased view on actual consumer behavior. This is because the purchase intentions of consumers may be affected by the survey itself. On the positive side, such effects can be incorporated in econometric models to get reliable estimates of actual behavior of non-surveyed consumers, which often is the ultimate purpose of survey studies. This paper proposes a reasonably simple methodology to correct for such possible survey effects and to get consistent predictions beyond the survey sample. The potential merits of the method are illustrated by a supermarket survey on easy-to-prepare food products and related health issues. This indicates that the required corrections can be quite substantial and that predictions that neglect survey effects can be seriously biased indeed.

Keywords consumer behavior, survey effects, bias correction, purchase prediction, econometric models

*The authors thank Mr. De Man and Mr. Vermaas, of supermarket 'Super De Boer' in Dordrecht - Stadspolders and Hendrik Ido Ambacht, for permitting us to perform the survey study in their shops. Special thanks are for our students Jolanda Collard, Marieke de Groot, Nadine Plaisier, Irene van Setten, and Jacqueline van der Weijde, for their assistance in starting up this project and for collecting the data.

†Corresponding author, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, email: heij@few.eur.nl, fax: +31-10-4089162, tel: +31-10-4081277.

‡Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, email: franses@few.eur.nl, fax: +31-10-4089162, tel: +31-10-4081278.

INTRODUCTION

Consumers may adjust their purchase decisions in response to survey questions. In general, the purpose of surveys is to get an idea of the decision process of consumers who do not participate in the survey, and hence corrections for the survey effects are needed. There is a substantial literature on the effects of surveys on subsequent behavior, and a recent summary of the relevant marketing literature is given in Morwitz (2005). The main result is that measuring intentions (of purchasing products), or satisfaction (with services), makes people in the experimental group to change their behavior. For instance, if products are reviewed in a positive way then subsequent purchase levels tend to be higher in the experimental group. Even if the survey is neutral, in the sense that no reviews of previous opinions on the product or service are given, the mere fact that the consumer is asked for satisfaction levels tends to increase their satisfaction. Dholakia and Morwitz (2002) document that such effects can be persistent over some time, see Chandon, Morwitz and Reinartz (2004).

The literature on survey effects shows that this aspect of consumer behavior should not be neglected in empirical analysis. One way to reduce these effects is to try to phrase questions in such a way that their effect on subsequent behavior is small. Another approach, which is perhaps more easy, is to create econometric models and estimation techniques that explicitly incorporate the possible presence of survey effects. In this paper we follow this second line, and we propose a simple methodology to deal with survey effects that may be present in survey data.

The idea to formulate explicit models for survey effects was applied by Chandon, Morwitz and Reinartz (2005), where these authors use regression methods. Their approach is characterized by the following two key features. First, their model incorporates terms that account for the possible existence of self-generated validity effects. Second, their method corrects for the fact that latent purchase intentions can only be imperfectly measured in surveys. Stated in technical terms, the measurement errors lead to endogeneity, which means that a direct regression of the observed purchase behavior on the measured intention gives inconsistent estimators and, therefore, biased results.

In this paper, we extend and modify their method to the situation where the measured purchase decision has two possible outcomes, 'yes' or 'no', so that the purchase variable is binary. The proposed modifications are needed for reliable inference, that is, to get consistent estimators of purchase behavior and of the involved uncertainties as measured by standard errors. We will use an econometric method to correct for endogeneity in binary models, as proposed by Rivers and Vuong (1988), see also Wooldridge (2002, Section 16.6.2). We refer to Cameron and Trivedi (2005, Chapter 14) for a general introduction to econometric methods for binary outcome models.

In addition to good models and methods that include the involved behavioral aspects and statistical issues, it is of course of crucial importance to collect appropriate data. We describe a field study, where customers of supermarkets were studied in their purchases of easy-to-prepare food products. This study is explicitly designed in such a way that it becomes relatively easy to measure the magnitude of the survey effects and to correct for these effects with the purpose to extrapolate consumer behavior beyond the experimental group. As relatively few customers purchase the products of interest, we inflated the data set by adding synthetic data, replicating the data of actual buyers several times in the inflated data set. This is done to show the potential merits of our methodology in situations were a sufficiently large number of survey data is available.

This paper has the following structure. First, we discuss the general methodology. Here we describe a prototypical situation that regularly occurs in practice, where purchase intentions and actual behavior are measured. We translate this situation into an econometric model, which explicitly takes into account that asking for purchase intentions can cause people to behave differently. We also describe the typical data that one needs to collect in order to be able to correct for such survey effects. In the second part of this methodological section, we discuss the proper estimation method for this model, and we discuss variants of this method in case the purchase variable of interest takes only two values, 'yes' or 'no'. Next, we illustrate the empirical relevance of our methodology. Here we analyze data that were recently collected, concerning health issues and their impact on sales of easy-to-prepare food products. The model with survey effects requires a slightly modified version of the prototypical model and of the estimation method, and technical details of this modification are relegated to an appendix. The empirical results with the inflated data set indicate the potential magnitude of the involved survey effects. Further, we show that simpler methods may give biases in estimated behavior and hence in purchase forecasts. Finally, we summarize the results and conclude that it matters to take the right decisions in collecting data, constructing models, and estimating parameters. We also discuss the limitations of our results, which are mainly confined to the available data. We argue that proper correction of survey effects may require a substantial amount of informative data.

METHODOLOGY

Modelling of Survey Effects

In our analysis of potential survey effects, we take the model of Chandon, Morwitz and Reinartz (2005) as our starting point. We briefly summarize their approach, and for ease of reference we use the same notation as in their paper.

Data are available for two groups of customers. In the survey group, the data consists of (i) stated subjective purchase intentions (expressed in the survey before the actual decision, and denoted by MI, the 'Measured Intention'), (ii) actual purchase decisions ('yes' or 'no', denoted by the binary variable B with value 1 for 'yes' and value 0 for 'no'), and (iii) a set of objective customer characteristics (such as gender, age, and other available indicators of latent purchase intention, which we jointly denote by Z). In the control group, the data consists of the purchase decision B and the set of characteristics Z, but no data are available on MI. Customers in the control group are questioned after they have made their purchases, and so their purchase decisions can not affected by the survey questions.

As the expressed intentions MI in the survey group may differ from the actual latent intentions (denoted by LI, which is unobserved), the purchase model with potential survey effects is formulated by the following two equations. First, the behavioral equation, linking purchase decisions to latent intentions and survey effects, is described by

$$B = \alpha + \beta_1 LI + \beta_2 S + \beta_3 (LI \times S) + \varepsilon. \tag{1}$$

Here S indicates the experimental group, with value S=1 in the survey group and S=0 in the control group, and ε stands for all unmeasured factors that affect the purchase decision. The level effect β_2 of surveying is called 'intention modification', and the slope effect β_3 is known as 'self-generated validity'. All parameters in this equation are of interest, as (β_2, β_3) measure the survey effects and (α, β_1) measure customer behavior that is not affected by the survey (with S=0). Second, the measurement equation linking latent intentions to the measured intentions in the survey group is given by

$$MI = LI + \delta_{MI}, \tag{2}$$

where δ_{MI} denotes the measurement error of the true, latent intention. If we substitute this in the behavioral Equation 1, we get

$$B = \alpha + \beta_1 MI + \beta_2 S + \beta_3 (MI \times S) + \mu, \tag{3}$$

where $\mu = \varepsilon - \beta_1 \delta_{MI} - \beta_3 S \delta_{MI}$.

The parameters of interest $(\alpha, \beta_1, \beta_2, \beta_3)$ can not be estimated by applying ordinary least squares (OLS) to Equation 3, because of the following three complications. First, the measured intention is available only in the survey group (with S=1), so that only $(\alpha + \beta_2)$ and $(\beta_1 + \beta_3)$ can be estimated in this way. Second, the measured intention MI and the composite error term μ both depend on the measurement error δ_{MI}, so that MI is endogenous and OLS will be inconsistent. This means that, even in very large samples, the OLS estimates of the parameters differ from the true underlying behavior. Third, the binary character of the purchase variable B is neglected, and OLS is not the right method anyway to estimate models with binary dependent variables.

In Chandon, Morwitz and Reinartz (2005), the first two complications are solved in an elegant way, as follows. First of all, the measured intention MI in the survey group is related to the r customer characteristics $Z = (z_1, \ldots, z_r)$, by means of

$$MI = \gamma_1 z_1 + \ldots + \gamma_r z_r + \eta. \tag{4}$$

The parameters in this equation are estimated by OLS, and the result is written as

$$MI = \hat{\gamma}_1 z_1 + \ldots + \hat{\gamma}_r z_r + e = FMI + e, \tag{5}$$

where FMI $= \hat{\gamma}_1 z_1 + \ldots + \hat{\gamma}_r z_r$ is the fitted intention and e is the residual term. This equation can be used to construct the fitted intentions FMI also for the control group, as Z is measured in this group. This solves the first complication, as MI, which is available only in the survey group, is replaced by FMI that is available also in the control group. The second complication, endogeneity of MI, is solved by replacing MI in Equation 3 by FMI. Under the assumption that the measurement errors δ_{MI} are independent of the customer characteristics Z, which seems a logically sound assumption, the endogeneity bias disappears. The parameters of interest are estimated by the regression

$$B = \alpha + \beta_1 FMI + \beta_2 S + \beta_3 (FMI \times S) + \omega. \tag{6}$$

This two-step method is known in econometrics as 'two-stage least squares' (2SLS). The first stage consists of the regression in Equation 4 for the survey group alone, and the second stage of the regression in Equation 6 for the survey and control groups jointly. However, the method differs from conventional 2SLS where the estimation sample should be the same at both stages.

Consistent Estimation of Survey Effects

The above two-step method gives consistent estimators of the (behavioral and survey effect) parameters of interest if the purchase variable B is an unrestricted scale variable, that is, a variable which can take any value. However, the standard errors of the coefficients obtained in the second stage are not valid and they may well underestimate the actual uncertainty in the estimates. Instead, 2SLS standard errors can be used, and these are available in many statistical packages. Another shortcoming is that the dependent variable B does not satisfy the conditions required for consistency of 2SLS. In our application, B is a binary variable. It is well-known that 2SLS is not consistent in this case, although it may still give useful estimates of average effects, see Wooldridge (2002, p. 472). Even if B is a continuous variable, for instance the amount of money spent on the products of interest, B is still limited to be non-negative. This situation is known as censored regression, and 2SLS is also inconsistent for this type of data, see Wooldridge (2002, Section 16.6.2, pp. 530-3).

To get parameters estimates and standard errors that are valid in large enough samples, we follow a modified two-step method for binary data as suggested by Rivers and Vuong (1988). Here we summarize the required steps, and we refer readers interested in further details of this method to Wooldridge (2002, Section 15.7.2, pp. 472-7). For simplicity, we assume for the moment that the samples are the same at both stages, as in standard 2SLS. In particular, it is assumed that the measured intention is known also in the control group. The case of unequal samples, which is relevant because of the missing intention data in the control group, is treated in the Appendix.

As a preliminary step, we first briefly discuss the standard model for binary dependent data, see Cameron and Trivedi (2005, Chapter 14) for a more extensive discussion. As an alternative to the behavioral Equation 3, the purchase decision B of a customer ('yes', with B=1, or 'no', with B=0) is described as follows. The decision of this customer is modelled in terms of the latent stimulus B* $= \alpha + \beta_1 MI + \beta_2 S + \beta_3 (MI \times S) - \nu$, where ν is an unobserved customer-specific component, such that

$$B = 1 \text{ if and only if } B^* > 0.$$

Let Φ be the cumulative distribution function of v, then the probability of a purchase, that is, B=1, is equal to the probability that $\nu < \alpha + \beta_1 MI + \beta_2 S + \beta_3 (MI \times S)$, so that

$$\text{Prob}(B = 1) = \Phi\Big(\alpha + \beta_1 MI + \beta_2 S + \beta_3 (MI \times S)\Big).$$

If ν is assumed to follow a standard normal distribution, then this is called the probit model. The parameters can be estimated by maximum likelihood, and most software packages contain probit procedures that are easy to apply. The parameters are estimated consistently in this way, provided that the model assumptions are correct. In particular, if one or more of the explanatory factors are endogenous, as is the case in our application, then consistency requires a simple correction of the standard probit method. This correction consists of adding the residuals of the first stage as an additional explanatory factor.

Details of Estimation Method

The following two-stage method gives consistent estimators and standard errors for binary dependent variable models with endogenous regressors, provided that these endogenous regressors are measured on a continuous scale. The first stage consists of OLS in Equation 4, as before. Let e be the residuals of this first stage regression, and let s^2 be the corresponding residual variance. The second stage is probit in the model

$$\text{Prob}(B=1) = \Phi\Big(\alpha^* + \beta_1^* MI + \beta_2^* S + \beta_3^*(MI \times S) + \theta e\Big), \tag{7}$$

where Φ is the standard normal cumulative distribution function. This gives consistent estimators and standard errors of scaled versions of the parameters of interest. Consistent estimators of the original parameters are obtained by appropriate re-scaling. Let $(\hat{\alpha}^*,\ \hat{\beta}_1^*,\ \hat{\beta}_2^*,\ \hat{\beta}_3^*, \hat{\theta})$ be the above second stage probit estimates, then consistent estimators of the behavioral and survey parameters $(\alpha, \beta_1, \beta_2, \beta_3)$ in Equation 1 are obtained as follows:

$$(\hat{\alpha},\ \hat{\beta}_1,\ \hat{\beta}_2,\ \hat{\beta}_3) = \Big(1/\sqrt{1+\hat{\theta}^2 s^2}\Big) \times (\hat{\alpha}^*,\ \hat{\beta}_1^*,\ \hat{\beta}_2^*,\ \hat{\beta}_3^*). \tag{8}$$

The probit t-statistic of $\hat{\theta}$, the estimated coefficient of e in Equation 7, is a valid test of the null hypothesis that MI is exogenous, that is, that the latent intention is measured without error. We summarize the required steps.

1. Perform OLS in Equation 4, with residuals e.

2. Estimate Equation 7 by probit.

3. Re-scale the coefficients of Stage 2 by means of Equation 8.

Estimation of Marginal Effects

The interpretation of probit estimates differs somewhat from that of regression coefficients. This is because in linear regression models, the coefficients measure constant marginal effects, whereas in the (nonlinear) probit model, the marginal effects are not constant across the sample. The marginal effect of the measured intention on the probability of a purchase is estimated consistently by

$$\frac{\partial \mathrm{Prob}(B=1)}{\partial MI} = (\hat{\beta}_1 + \hat{\beta}_3 S) \times \phi\Big(\hat{\alpha} + \hat{\beta}_1 MI + \hat{\beta}_2 S + \hat{\beta}_3 (MI \times S)\Big),$$

where $\phi(z) = \exp\{-z^2/2\}/\sqrt{2\pi}$ is the standard normal density, see Wooldridge (2002, p. 475). This marginal effect depends on the intention MI and it differs between customers in the survey group (with S=1) and other customers (S=0).

As the survey variable S is binary, the survey effect on the purchase probability can be measured by $\mathrm{Prob}(B=1 \mid S=1) - \mathrm{Prob}(B=1 \mid S=0)$. These probabilities depend on the intention MI and the effect is estimated consistently by

$$\Phi\Big((\hat{\alpha} + \hat{\beta}_2) + (\hat{\beta}_1 + \hat{\beta}_3) MI\Big) - \Phi\Big(\hat{\alpha} + \hat{\beta}_1 MI\Big).$$

ILLUSTRATION

Experiment Design

To investigate the potential effects of surveys on consumers' decisions, we performed a field study on easy-to-prepare food products in two supermarkets . The experimental group is split in two parts, a survey group and a control group. Customers in the survey group are interviewed on entering the shop. Here they answer a list of questions, including their past purchase frequency of the products (which measures their purchase intentions MI), their opinion on health aspects of the products, and several personal aspects like gender, age, smoking habits, and working hours. As our methodology requires that the measured purchase intention in the survey group is continuous, this question was phrased as follows: 'How often do you use this product in your dinners'. The possible answers are 'never' (MI=0), 'about once a month' (MI=1), 'about once a week' (MI=4), or 'more than once a week' (MI=8). With this question and coding, we tried to reconcile the need for quantitative information on past purchase decisions with the possibly limited recollection of consumers of their eating habits.

A distinction is made between two categories of easy-to-prepare food products, that is, (i) ready-made-food consisting of a complete meal and (ii) cooked vegetables in can or glass. Each product is immediately ready for consumption after heating up. A further distinction is made between positive surveys, where the interviewer mentions some positive health aspects of the product, and negative surveys where some negative health aspects are mentioned. This gives in total four survey groups. The actual purchase decision (B) is measured when the customer leaves the shop, with outcome 'yes' if the customer bought a product from the category under consideration and 'no' otherwise. Customers in the control group are interviewed after their shopping, so that they are not affected by the survey in their purchase decisions.

In performing the study, much attention is paid to the requirement that the control group should not know that an experiment is going on. Therefore, during each time segment of the experiment, the control group is sampled first and the survey groups afterwards. Further, as the purchase intentions in the control group are predicted from Equation 5, which is derived from the survey group, it is necessary that the customer characteristics in the control group are similar to those in the survey group. Therefore, the survey sample is taken immediately after the control sample, so that the time of the day and other circumstances are comparable for both groups.

One of the considerations in the design is the magnitude of the experimental groups. For each of the two products, let n_{s+} be the size of the positive survey group and n_{s-} that of the negative survey group, let $n_s = n_{s+} + n_{s-}$ be the total size of the survey group, and let n_c be the size of the control group. A rough statistical analysis of standard errors suggests to take $n_s = n_c$, that is, equal sizes in survey and control groups.[1] Further, we take $n_{s+} = n_{s-} = 50$ for each product category. This means that the total sample size is 400, with 200 observations for both product categories: 100 in the control group, 50 in the positive survey group, and also 50 in the negative survey group.

Collected Data

The previously discussed design was applied to collect data during several days in the Spring of 2006, in periods free from special actions or unusual attention for the products of interest and their substitutes. This delivered a data set on 360 customers, as around 40 of the interviews were dropped because of missing observations, and the number of observations available in estimation is 354 for similar reasons.

Table 1 shows some summary statistics of the collected data set. As the total number of purchases is only 39, we decided to join the two product categories in our analysis of survey effects. More precisely, we allow for differences in purchase intentions for the two product categories by including a product indicator (denoted by 'Product') in the first stage regression in Equation 4, with value 1 for cooked vegetables in can or glass and value 2 for ready-made-food. However, no distinction between the two products is made in the second stage probit in Equation 7.

Table 1 to be inserted about here.

As the purchase frequency is low, around 10%, it is not surprising that this greatly limits the strength of the conclusions that can be drawn from these data. To illustrate the potential merits

[1] This analysis is based on linear regression in Equation 6. The standard errors of the intention parameters α and β_1 are proportional to $1/\sqrt{n_c}$ and those of the survey effect parameters to $\sqrt{(n_s + n_c)/n_s n_c}$. The choice $n_s = n_c$ minimizes the standard errors of the survey effect parameters. Here we do not provide further details, which are available on request.

of our method, we decided to inflate part of the data. We added synthetic data by replicating the data of the 39 purchasing customers several times, to increase the purchase frequency to around 50%. The original data set contains 321 non-buyers and 39 buyers, and we add seven copies of the buyer data. So, the inflated data set has $360 + 7 \times 39 = 633$ observations, with 321 real non-buyers and eight copies of 39 real buyers, giving a total of 312 'buyers'. The number of observations available in estimation is 620, due to missing values. Table 1 summarizes the sample sizes in the inflated data set.

We screened the collected data on about thirty variables for their explanatory power. Table 2 summarizes the correlations between the variables that we will use in our analysis.

Table 2 to be inserted about here.

Some of the variables were explained before, and the other variables have the following meaning: 'Male' indicates the gender (1 if male, 0 if female), 'Age' is the reported age in years, 'Smokes' indicates current smoking behavior (1 if 'yes', 0 if 'no'), 'Shift Work' indicates the type of work in case the respondent has a job (1 if work in shifts, with irregular times, and 0 otherwise), and 'Surveypos' and 'Surveyneg' are survey indicators (1 if the indicated type of survey applies, 0 otherwise). In Stage 1, the explained variable is the measured intention (MI) in the survey group. This variable is denoted by 'Past Use', with values 0 (never), 1 (once a month), 4 (once a week), and 8 (more than once a week). In Stage 2, the explained variable is the purchase decision (B), denoted by 'Purchase' with values 1 ('yes') and 0 ('no').

Most correlations are rather small. Some correlations are larger in the inflated data set than in the collected data set, most notably so for the (positive and negative) survey effects. The correlation between the purchase indicator and the survey treatment is, as expected, positive for a positive treatment and negative for a negative treatment. In the inflated data set, the survey treatment is correlated with the intention variable ('Past Use'), notwithstanding our efforts in the collected data set to assign customers randomly to both treatment groups. As we will see in the sequel, survey effects are more easily detected within the negative treatment group than in the positive treatment group. This may be partly due to smaller purchase intentions in the positive survey group. Another possible explanation is that the interviewed customers may be more sensitive to negative information on the product than to positive information.

Estimation method

In the foregoing, we described our two-stage estimation methodology under two simplifying assumptions, that is, the observation samples are the same for both stages and there exists a single type of survey treatment. In the Appendix, we describe the modifications that are required for the application at hand, where the first stage regression has to be applied on a sub-sample (excluding the control group, as the purchase intention is not measured in this group) and where the survey is of two types, positive and negative. The outcome is that the endogeneity correction term should be applied only in the survey group, and not in the control group. This is also intuitively evident, as the endogenous variable (MI) is measured only in the survey group and not in the control group.

This leads to the following method for consistent estimation of all the parameters of interest. Here we use the following notation. The subindex s (c) denotes the survey (control) group, Z denotes the set of instrument variables $(z_1, \ldots, z_r)$ in Equation 4, and S_+ and S_- denote the survey treatment, with $S_+ = 1$ for a positive survey and $S_+ = 0$ otherwise, and $S_- = 1$ for a negative survey and $S_- = 0$ otherwise.

1. Perform OLS in Equation 4 for the survey group. Write the estimated Equation 5 as $MI_s = Z_s\hat{\gamma} + e_s = FMI_s + e_s$, with fitted values $FMI_s = Z_s\hat{\gamma}$ and residuals e_s.

2a. Estimate the following equation by probit for the survey group:

$$\text{Prob}(B = 1) = \Phi\Big(\alpha_+ S_+ + \alpha_- S_- + \beta_+(S_+ \times FMI_s) + \beta_-(S_- \times FMI_s) + \theta e_s\Big).$$

2b. Estimate the following equation by probit for the control group, where $FMI_c = Z_c\hat{\gamma}$ with $\hat{\gamma}$ obtained from Stage 1:

$$\text{Prob}(B = 1) = \Phi\Big(\alpha_c + \beta_c FMI_c\Big).$$

3. Re-scale the coefficients of Stage 2a, similar to Equation 8, that is, by dividing the coefficients by $\sqrt{1+\hat{\theta}^2 s^2}$ where s^2 is the residual variance of e_s in Stage 1.

The structural parameters of the behavioral Equation 1 are related to the above model parameters, as follows: $\alpha = \alpha_c$, $\beta_1 = \beta_c$, for the positive survey effect $\beta_{2+} = \alpha_+ - \alpha_c$ and $\beta_{3+} = \beta_+ - \beta_c$, and for the negative survey effect $\beta_{2-} = \alpha_- - \alpha_c$ and $\beta_{3-} = \beta_- - \beta_c$. Consistent estimators of these structural parameters are obtained by substituting the corresponding coefficients obtained by the above estimation method.

Results

We applied the above estimation methodology both to the collected data set and to the inflated data set. The estimation results are in Table 3. Here 'Intention' indicates the estimated purchase intention (FMI) obtained from Stage 1, and 'OLS Residual' is the series of residuals (e_s) of the regression in Stage 1. Further, 'Surveypos' ('Surveyneg') is a dummy variable with value 1 if the consumer is in the positive (negative) survey group and with value 0 otherwise, and 'Surveypos × Intention' ('Surveyneg × Intention') is an interaction term consisting of the product of the two indicated variables. The Stage 1 regression is for the survey group only (with sample size 172 or 277), and the Stage 2 probit is for the survey and control groups jointly (with sample size 354 or 620).

Table 3 to be inserted about here.

The Stage 1 regression for the collected purchase intention data is reasonably satisfactory. The 'Past Use' is lower for ready-made-food ('Product' = 2) than for cooked vegetables in can or glass ('Product' = 1). Males and smokers tend to use the products more than females and non-smokers, and the use is estimated to be minimal around an age of $(100 \times 0.146)/(2 \times 0.132) = 55$ years. Customers with a job involving work in shifts tend to use the products less than average.

This may be contrary to expectations, as irregular working times may increase the attractiveness of easy-to-prepare food. Note, however, that some competitive fast food products, for instance, pizza and French fries, were not included in the analysis. All coefficients are significant, and the R-squared is 0.24. The first stage results for the inflated data set are roughly similar, with increased significance due to the larger number of data and with an R-squared of 0.19.

The Stage 2 probit estimates are not significant for the collected data set, except for the negative intention modification effect of a negative survey. The significance is increased for the inflated data set, with a positive intention modification effect for positive surveys and with both an intention modification and a self-generated validity effect for negative surveys. The effects of negative surveys are much more pronounced than those of positive surveys. The coefficient of the fitted purchase intention is positive, as expected. Further, the significance of the OLS residual term implies that the endogeneity is significant. That is, in terms of Equation 2, the measurement errors $\delta_{MI,s}$ in measuring the latent purchase intentions LI_s in the survey group should not be neglected. Finally, the standard errors of the scaled coefficients in the last column in Table 3 can not easily be computed, but their significance will in general be close to that of the unscaled coefficients in the preceding column.

Table 4 compares the outcomes of our method in Table 3 with those of alternative methods. For ease of comparison, columns (5), (10) and (11) replicate the second stage results of Table 3. Further, 'Spos $\times$ Int' denotes the product of the variables 'Surveypos' and 'Intention', and 'Sneg $\times$ Int' is the product of 'Surveyneg' and 'Intention'.

The estimates in columns (1), (3), (6) and (8) are based on the survey group with negative treatment only. These results would be obtained if survey effects were neglected and the sample consisted of the negative survey group alone. The estimated intention effect (in the row 'Intention') clearly differs from the results of the other methods, which illustrates the bias caused by neglecting survey effects.

The results in columns (2) and (7) correspond to the method of Chandon, Morwitz and Reinartz (2005), where the binary character of the purchase variable is neglected. The probit results in columns (4) and (9) neglect the endogeneity correction term that is included in columns (5) and (10). Because of the special structure of the model, with value 0 in the control group for the four survey-related explanatory variables, the coefficients of the constant term and of the variable 'Intention' in the control group are the same in columns (4) and (5) and also in columns (10) and (11). That is, if the endogeneity would be neglected then the behavioral parameters in the control group will still be estimated consistently, but this does not hold true for the survey effects in the survey group. A further explanation of this fact is given in the Appendix, where the estimation of the parameters is analyzed in more detail.

Table 4 to be inserted about here.

Evaluation

The results in Tables 3 and 4 can be used to evaluate the effects on the purchase probability of the survey and of changes in the latent purchase intention.

First we consider the marginal effect of the purchase intention on the purchase probability. The effect in the control group is of most interest, as these customers are free from survey effects and their behavior can therefore be considered to be representative for all consumers. The marginal effect is constant in linear regression models, but in probit models it depends on the intention level. We consider five intention levels, which are derived from the estimates of Stage 1 using the inflated data set. This gives a set of 620 estimated intentions, with minimum, quartiles, and maximum shown in the column 'Intention' in Table 5. These intentions are substituted in four purchase models, with codes corresponding to those used in Table 4: methods (6) and (8) neglect the survey effect, (7) corrects for this effect but neglects the binary character of the purchase data, and (11) takes both the survey effect and the binary aspect into account. Consumers of interest are the non-surveyed ones, with 'Surveypos = 0' and 'Surveyneg = 0' in models (7) and (11).

The differences between methods (6) and (7) and also those between (8) and (11) show that the required endogeneity correction has a very substantial effect on the estimated marginal effects. The distinction between the (wrongly specified) methods (6) and (8) are more substantial than those between (7) and (11). In the linear model of Chandon, Morwitz and Reinartz (2005), the marginal effect is 2.6%, whereas in our probit model it ranges from 2.4% to 2.6% for medium purchase intention. In probit models, the marginal effects can in principle differ widely for different intention levels. In our application, however, the coefficient of 'Intention' in the control group is relatively small, that is, 0.065, see Table 4, method (11). As the fitted purchase intentions (FMI) vary between -3.85 and 5.94, the marginal effect is nearly flat over the considered range.

Table 5 to be inserted about here.

Next we consider the size of survey effects on the purchase probability. Because of the results in Table 4, we restrict the attention to negative surveys, as the effects of a positive survey were found to be relatively small. According to the results for our method in column (11) of Table 4, for a given level FMI of the purchase intention, the purchase probability in the control group is $\Phi(-0.039+0.065\text{FMI})$ and in the negative survey group it is $\Phi(-1.534+0.349\text{FMI})$. To evaluate the merits of our method, we also estimate these probabilities by means of method (7), where the binary character of the purchase variable is neglected, and (9), where the endogeneity is neglected. The results are in Table 6. The results of methods (7) and (9) are quite similar in most cases, indicating that neglecting either one of the data aspects, their binary character or endogeneity, leads to comparable biases. The survey effects (in the rows 'Effect') estimated by the consistent method (11) are systematically larger than the estimates obtained by the inconsistent methods (7) and (9). As compared to method (7), our method provides an upward correction of the survey effects by a factor of about 25% for median and above-average buyers and of 10% for below-average buyers.

Table 6 to be inserted about here.

CONCLUSIONS

Surveys can help to understand consumer behavior, provided that proper attention is paid to the nature of the data and of the data collection process. In this paper, we considered binary purchase decisions ('yes' / 'no') in situations where a survey interview may affect customers in their purchase behavior. Often, the real purpose of survey studies is to get reliable predictions of consumer decisions outside the survey sample. To achieve this goal, it is essential to collect the best possible data and to construct an econometric model that takes the properties of the data and of the survey process into account. In statistical terms, focusing on behavioral parameters and their standard errors, it is of importance to use consistent methods. We proposed a consistent method for modelling binary decisions in case of imprecisely measured purchase intentions. Our methodology extends the approach of Chandon, Morwitz and Reinartz (2005) to the case of binary data. It consists of an adaptation of the method of Rivers and Vuong (1988) to the situation where the purchase intention is measured only in the survey group and not in the control group.

In our application, methods that neglect the binary character of the data or the endogeneity of the measured intentions tend to under-estimate the survey effects. For surveys with negative information on the products of interest, the required correction factors obtained from our consistent method are about 10% for consumers with below-average purchase intentions, and about 25% for consumers with median and above-average purchase intentions.

The proposed methodology is simple to use and easy to understand. However, collecting the right data is less easy. Our field study, with 360 customers divided in approximately equal survey and control groups, turned out to be too small to get sufficiently significant results. Various observational factors affect the data, among which are measurement errors of latent intentions and survey effects. It requires a substantial amount of data to disentangle these observation effects from the underlying structural behavior of consumers which is needed to extrapolate the survey data to future decisions. The main message of our paper is that these observational factors should not be neglected, as otherwise one may get the wrong picture of actual behavior. To limit the required sample sizes, it helps to design the survey experiment carefully so that the maximum amount of relevant information is gathered. This includes the set-up of the experiment, with survey and control groups, and asking the right questions, including those that help in estimating the involved measurement errors and survey effects.

APPENDIX

We describe a consistent method to estimate the parameters of Equation 1. We write B_s^* (B_c^*) for the latent variable B^* in the survey (control) group. A purchase of the product ($B = 1$) occurs if and only if $B^* > 0$. Let $\alpha_c = \alpha$, $\beta_c = \beta_1$, $\alpha_s = \alpha + \beta_2$ and $\beta_s = \beta_1 + \beta_3$, then Equations 1, 2 and 4 become

$$\begin{aligned} B_s^* &= \alpha_s + \beta_s LI_s + \varepsilon_s \quad , \quad B_c^* = \alpha_c + \beta_c LI_c + \varepsilon_c \\ MI_s &= LI_s + \delta_{MI,s} = Z_s\gamma + \eta_s. \end{aligned}$$

Here Z_s is the matrix with instrument scores (for the variables gender, age and so on) for the survey group. Note that MI is not measured in the control group, but that instrument scores Z_c are available in the control group. As it assumed that the control group and the survey group are selected randomly, it makes sense to assume that the unmeasured MI_c satisfies $MI_c = LI_c + \delta_{MI,c} = Z_c\gamma + \eta_c$, where $(\varepsilon_c, \delta_{MI,c}, \eta_c)$ is independent of $(\varepsilon_s, \delta_{MI,s}, \eta_s)$. We now rewrite the model in terms of the latent variables (B_s^*, B_c^*) and the measured variables (MI_s, Z_s, Z_c), so that

$$B_s^* = \alpha_s + \beta_s MI_s + \omega_s \tag{9}$$

$$MI_s = Z_s\gamma + \eta_s \tag{10}$$

$$B_c^* = \alpha_c + Z_c\beta_c\gamma + \omega_c \tag{11}$$

$$\omega_s = \varepsilon_s - \beta_s\delta_{MI,s} \quad , \quad \omega_c = \varepsilon_c - \beta_c\delta_{MI,c} + \beta_c\eta_c.$$

The central complication of this model, endogeneity of MI_s, is due to the fact that $\mathrm{cov}(\omega_s, \eta_s) = \mathrm{cov}(\omega_s, MI_s) = -\beta_s \mathrm{var}(\delta_{MI,s}) \neq 0$ (the first equality follows from the exogeneity of Z, and the second equality follows from the assumptions that the variables LI_s, ε_s and $\delta_{MI,s}$ are mutually independent). However, under the intuitively reasonable assumption of complete independence of the error terms of the control group from those of the survey group, it follows that $\mathrm{cov}(\omega_c, \omega_s) = \mathrm{cov}(\omega_c, \eta_s) = 0$.

The above three-equation model (9)-(11) can be estimated in several ways. One method is full Maximum Likelihood, where care should be taken to account for the difference in variance of the error terms ω_s and ω_c, so that the probit part is heteroskedastic. Also note the parameter restriction that the coefficient vectors γ of Z_s in Equation 10 and $\beta_c\gamma$ of Z_c in Equation (11) should be proportional to each other.

We will use a much simpler, consistent, though (asymptotically) less efficient procedure, which is inspired by the method of Rivers and Vuong discussed in the main text. In Stage 1, estimate γ consistently by regressing MI_s on Z_s, with fitted regression $MI_s = Z_s\hat{\gamma} + e_s = FMI_s + e_s$. For the probit estimates in Stage 2, we get in the survey group $B_s^* = \alpha_s + \beta_s FMI_s + \beta_s e_s + \omega_s$, and, with $FMI_c = Z_c\hat{\gamma}$, in the control group $B_c^* = \alpha_c + \beta_c FMI_c + \omega_c^*$ where $\omega_c^* = \omega_c + Z_c(\gamma - \hat{\gamma})$. That is, the endogeneity correction term e_s is present only in the survey group and not in the control group. The equations can be combined by defining $e_c = 0$ in the control group, so that we can write the equation in combined form for both groups as

$$B^* = \alpha_c + \beta_c FMI + (\alpha_s - \alpha_c)S + (\beta_s - \beta_c)FMI \times S + \theta e + \omega^*, \tag{12}$$

where $S = 1$ in the survey group and $S = 0$ in the control group. Here we omit the parameter restriction that $\theta = \beta_s$, which reduces efficiency but does not affect the consistency of the estimators. This probit model is heteroskedastic, as in the survey group $\omega_s^* = \omega_s = \varepsilon_s - \beta_s\delta_{MI,s}$ whereas in the control group $\omega_c^* = \omega_c + Z_c(\gamma - \hat{\gamma}) = \varepsilon_c - \beta_c\delta_{MI,c} + \beta_c\eta_c + Z_c(\gamma - \hat{\gamma})$. Because of the special structure of the model, with separate parameters for the two groups and with $e_c = 0$ in the control group, this heteroskedasticity can be neglected. Indeed, because the samples in the survey and control groups are mutually independent, the joint log-likelihood function (logL) for Stage 2 is simply the sum of the log-likelihoods for the two groups, and with $\sigma_s^2 = \mathrm{var}(\omega_s)$ and $\sigma_c^2 = \mathrm{var}(\omega_c^*)$ we get

$$\mathrm{logL} = \sum_{\text{survey}} \Phi\Big(\frac{\alpha_s + \beta_s FMI_s + \theta e_s}{\sigma_s}\Big) + \sum_{\text{control}} \Phi\Big(\frac{\alpha_c + \beta_c FMI_c}{\sigma_c}\Big).$$

Clearly, the maximum likelihood estimates can be obtained by two separate probit estimates, one for the survey group (including the endogeneity correction term e_s as an additional factor) and another (standard) one for the control group.

A final adjustment is needed, as the survey is of two different types, positive and negative. In this case, the term $(\alpha_s + \beta_s FMI_s + \theta e_s)$ in the above expression for the log-likelihood in the survey group should be replaced by

$$\alpha_+ S_+ + \alpha_- S_- + \beta_+ (S_+ \times FMI_s) + \beta_- (S_- \times FMI_s) + \theta e_s.$$

This leads to the estimation method described in the main text.

References

[1] Cameron, A. Colin and Pravin K. Trivedi (2005), *Microeconometrics - Methods and Applications*. New York: Cambridge University Press.

[2] Chandon, Pierre, Vicki G. Morwitz and Werner J. Reinartz (2004), "The short- and long-term effects of measuring intent to repurchase," *Journal of Consumer Research*, 31 (3), 566-572.

[3] Chandon, Pierre, Vicki G. Morwitz and Werner J. Reinartz (2005), "Do intentions really predict behavior? Self-generated validity effects in survey research," *Journal of Marketing*, 69 (2), 1-14.

[4] Dholakia, Utpal M. and Vicki G. Morwitz (2002), "The scope and persistence of mere-measurement effects: Evidence from a field study of customer satisfaction measurement," *Journal of Consumer Research*, 29 (2), 159-167.

[5] Morwitz, Vicki G. (2005), "The effect of survey measurement on respondent behaviour," *Applied Stochastic Models in Business and Industry*, 21 (5), 451-455.

[6] Rivers, Douglas and Quang H. Vuong (1988), "Limited information estimators and exogeneity tests for simultaneous probit models," *Journal of Econometrics*, 39 (3), 347-366.

[7] Wooldridge, Jeffrey M. (2002), *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.

Table 1: EXPERIMENTAL SAMPLE SIZES

Data	Group	Survey			Control	Total	Estimation
		Positive	Negative	Both			
Collected	All	91	84	175	185	360	354
	Buyers	11	5	16	23	39	38
Inflated	All	168	119	287	346	633	620
	Buyers	88	40	128	184	312	304

Notes: Size of consumer groups (two survey groups, with positive and negative treatment, and control group) and number of buyers of the product, for the collected data and also for an inflated data set with synthetically added buyers. The estimation sample sizes are reduced due to missing observations for some of the variables.

Table 2: DATA CORRELATIONS

Data Set	Explained	Past Use	Product	Male	Age	Smokes	Shift Work	Surveypos	Surveyneg
Collected	Past Use	1.00	-0.39	0.13	-0.13	0.17	-0.11	-0.02	0.02
	Purchase	0.18	-0.16	-0.08	0.10	0.04	0.02	0.02	-0.09
Inflated	Past Use	1.00	-0.33	0.17	-0.02	0.19	-0.10	-0.14	0.14
	Purchase	0.26	-0.26	-0.14	0.15	0.06	0.03	0.04	-0.15

Notes: Pairwise correlations of explained variables ('Past Use' and 'Purchase') with explanatory factors, both for the collected data and for an inflated data set with synthetically added buyers. The meaning of the variables is explained in the text.

Table 3: ESTIMATION RESULTS

	Collected Data		Inflated Data Set		
	OLS (Stage 1)	Probit (Stage 2)	OLS (Stage 1)	Probit (Stage 2)	Scaled
Explained	Past Use	Purchase	Past Use	Purchase	Purchase
Constant	7.944***	-1.301***	7.863***	-0.039	-0.039
Product	-1.898***		-1.877***		
Male	0.637*		1.495***		
Age	-0.146**		-0.155**		
$Age^2/100$	0.132**		0.155**		
Smokes	0.805**		0.927***		
Shift Work	-3.226***		-4.108***		
Intention		0.097		0.065*	0.065
Surveypos		0.008		0.046	0.054
Surveyneg		-1.072*		-1.277***	-1.495
Surveypos × Intention		-0.031		-0.030	-0.035
Surveyneg × Intention		0.238		0.243**	0.284
OLS Residual		0.077		0.105***	
Sample Size	172	354	277	620	620
Significance	0.000	0.134	0.000	0.000	

Notes: Estimated coefficients (Stage 1 and Stage 2) for the collected data and for an inflated data set with synthetically added buyers. A *** denotes significance at 1% significance level, ** at 5%, and * at 10%. The 'Significance' shows the P-value for the test on the joint significance of all coefficients in the equation.

Table 4: COMPARISON OF METHODS

	Collected Data					Inflated Data Set					
	(1) OLS	(2) OLS	(3) Probit	(4) Probit	(5) Corrected	(6) OLS	(7) OLS	(8) Probit	(9) Probit	(10) Corrected	(11) Scaled
Constant	-0.008	0.094	-2.380	-1.301	-1.301	0.055	0.484	-1.362	-0.039	-0.039	-0.039
Intention	0.032	0.021	0.349	0.097	0.097	0.103	0.026	0.334	0.065	0.065	0.065
Surveypos		0.004		-0.005	0.008		0.023		0.057	0.046	0.054
Surveyneg		-0.102		-1.078	-1.072		-0.429		-1.323	-1.277	-1.495
Spos $\times$ Int		-0.007		-0.016	-0.031		-0.016		-0.040	-0.030	-0.035
Sneg $\times$ Int		0.011		0.251	0.238		0.077		0.268	0.243	0.284
OLS Residual					0.077					0.105	
Sample Size	82	354	82	354	354	110	620	110	620	620	620
Significance	0.153	0.247	0.142	0.146	0.134	0.004	0.000	0.003	0.000	0.000	

Notes: Estimated coefficients (Stage 2, with explained variable ‘Purchase’) for the collected data and for an inflated data set with synthetically added buyers. The corrected estimates in columns (5,10,11) are identical to the results of Stage 2 in Table 3.

Table 5: EFFECT OF CHANGE IN PURCHASE INTENTION

	Intention	(6) OLS	(7) OLS	(8) Probit	(11) Corrected
Minimum	-3.85	10.3	2.6	0.4	2.5
1-st Quartile	0.88	10.3	2.6	7.5	2.6
Median	2.20	10.3	2.6	10.9	2.6
3-rd Quartile	3.22	10.3	2.6	12.8	2.6
Maximum	5.94	10.3	2.6	11.0	2.4

Notes: Estimated marginal effect of the purchase intention on the probability (expressed as percentage) of a purchase for non-surveyed consumers. See Table 4 for the coefficients of methods (6, 7, 8, 11).

Table 6: ESTIMATES OF NEGATIVE SURVEY EFFECTS

	Intention	Group	(7) OLS	(9) Probit	(11) Corrected	(11)-(7) Difference	$100 \times \frac{(11)-(7)}{(7)}$ % Difference
1-st Quartile	0.88	Control	50.69	50.72	50.72	0.03	
		Neg. Survey	14.56	14.25	10.99	-3.57	
		Effect	36.13	36.47	39.73	3.60	10
Median	2.20	Control	54.12	54.14	54.14	0.02	
		Neg. Survey	28.16	26.45	22.18	-5.98	
		Effect	25.96	27.69	31.96	6.00	23
3-rd Quartile	3.22	Control	56.77	56.76	56.76	-0.01	
		Neg. Survey	38.67	38.60	34.08	-4.59	
		Effect	18.10	18.16	22.68	4.58	25

Notes: Estimated effect of a negative survey on the probability of a purchase (expressed as percentage). The 'Effect' rows show the difference between the purchase probabilities in the control group and in the negative survey group. See Table 4 for the coefficients of the methods (7,9,11).