

RESEARCH SUPPORT
UNIVERSITY OF MICHIGAN BUSINESS SCHOOL

OCTOBER 1996

**THE STOCHASTIC MODELING OF PURCHASE
INTENTIONS AND BEHAVIOR**
Working Paper #9612-25

Martin Young
University of Michigan Business School

Wayne S. DeSarbo
Smeal College of Business
Pennsylvania State University

Vicki G. Morwitz
Leonard N. Stern School of Business
New York University

Michigan Business School

Research Papers

**THE STOCHASTIC MODELING OF PURCHASE
INTENTIONS AND BEHAVIOR**
Working Paper #9612-25

Martin Young
University of Michigan Business School

Wayne S. DeSarbo
Smeal College of Business
Pennsylvania State University

Vicki G. Morwitz
Leonard N. Stern School of Business
New York University

Research Support

- School of Business Administration
 - University of Michigan
 - Ann Arbor, Michigan 48109-1234
-

The Stochastic Modeling of Purchase Intentions and Behavior

To appear in *Management Science*

Martin R. Young

Department of Statistics & Management Science
University of Michigan School of Business Administration

Wayne S. DeSarbo

Department of Marketing
Smeal College of Business
Pennsylvania State University

Vicki G. Morwitz

Department of Marketing
Leonard N. Stern School of Business, New York University

October, 1996

Abstract

A common objective of social science and business research is the modeling of the relationship between demographic/psychographic characteristics of individuals and the likelihood of certain behaviors for these same individuals. Frequently, data on actual behavior is unavailable; rather, one has available only the self-reported intentions of the individual. If the reported intentions imperfectly predict actual behavior, then any model of behavior based on the intention data should account for the associated measurement error, or else the resulting predictions will be biased. In this paper, we provide a method for analyzing intentions data that explicitly models the discrepancy between reported intention and behavior, thus facilitating a less biased assessment of the impact of designated covariates on actual behavior. The application examined here relates to modeling relationships between demographic characteristics and actual purchase behavior among consumers. A new Bayesian approach employing the Gibbs sampler is developed and compared to alternative models. We show, through simulated and real data, that, relative to methods which implicitly equate intentions and behavior, the proposed method can increase the accuracy with which purchase response models are estimated.

Key Words: *Bayesian Methods, Hierarchical Bayes, Markov Chain Monte Carlo, Measurement Error, Probit Regression, Purchase Intentions, Stochastic Models*

The authors wish to thank the Departmental Editor, Linda Green, and two anonymous reviewers for their insightful comments which improved the content of this manuscript.

1 Introduction

A common objective of applied social science and business research is the modeling of the relationship between the demographic/psychographic characteristics of individuals and the behavior of these individuals. For example, in many marketing research applications, the goal of many concept and product tests is to understand the purchase behavior of a population of consumers, to determine whether or not to introduce a product to the market, and, if so, to which market segments. Typically in such concept and product tests, one cannot observe the actual purchases of individuals, so self-reported purchase intentions for the new product or concept are measured and used as a proxy variable. Intentions are also often used to predict sales over time for existing products among different segments of customers. In order to determine the purchase propensity of members of different market segments, models relating demographic, psychographic, or other characteristics of consumers to purchase intentions are often developed. When there is an imperfect correspondence between self-reported purchase intentions and actual behavior, estimates concerning consumer characteristics obtained from these models will be biased.

A key question, then, is what is the relationship between self-reported purchase intentions and actual behavior? Research in social psychology suggests that intentions should be the best predictor of an individual's behavior because they allow each individual to independently incorporate all relevant factors that may influence his or her actual behavior (Fishbein and Ajzen 1975). However, most empirical evidence suggests that purchase intentions cannot be taken literally.

Over the last fifty years, several studies have examined the relationship between stated purchase intentions and actual purchase behavior. The U.S. government conducted studies and experiments concerning purchase intentions between the 1940s and the 1970s. Their goal was to predict short term movements in future expenditures based on stated purchase intentions and the financial position and demographic characteristics of American households. In many of these studies, significant relationships between intentions to buy

durable goods and subsequent purchase were found using various econometric models on panel data (Juster 1966; Tobin 1959). Although the relationships are significant, Juster describes the performance of binary intentions measures as unimpressive. He notes that although intender purchase rates are higher than those of non-intenders, intender purchase rates are significantly lower than one, and non-intender purchase rates are significantly greater than zero. In fact, since for most durable goods, the majority of households report that they are non-intenders, most of the actual purchases are made by non-intender households. Juster therefore advocated the use of purchase probability measures over direct binary intent measures. However, he noted that self-reported purchase probabilities also provide biased estimates of actual purchasing, typically underestimating the actual purchase rate.

Several other studies have examined the relationship between purchase intentions and purchase behavior for durable goods (Adams 1974; Clawson 1971; Ferber and Piskie 1965; Granbois and Summers 1975; McNeil 1974; Pickering and Isherwood 1974) and for nondurable goods (Gormley 1974; Tauber 1975; Taylor, Houlahan, and Gabriel 1975; Warshaw 1980). The observed relationship between intentions and purchase is generally positive and significant, however, the strength of the relationship varies from study to study. For example, in Juster (1966), of those respondents who claimed they would “definitely or probably” purchase a car in the next six months, only 50 percent actually purchased. Jamieson and Bass (1989) and Morwitz and Schmittlein (1992) found similar results in different marketing contexts. Table 1 summarizes the results from these studies.

[Insert Table 1 Here]

Overall, based on the empirical evidence, intentions appear to almost always provide biased measures of purchase propensity, sometimes underestimating actual purchasing and other times overestimating actual purchasing. Manski (1990) provides further support that intentions data should not be taken literally. Manski develops a model of the relationship between binary intentions measures and subsequent behavior under the “best-case” situation where respondents have rational expectations. In this situation,

respondents' reported intentions are their best predictions of their future behavior. However, respondents typically do not have perfect information about changes that may occur in the future that may affect their probability of purchase. Based on his model, Manski demonstrates that intentions data do not identify a specific probability of behavior, but rather bound the probability of behavior, and that the bounds are nonparametrically estimable. Manski notes that, in contradiction to prior assertions, there is no reason to expect individual-level differences between intentions and behavior to average out in the aggregate. In other words, we should expect to observe that not all intenders purchase and that some non-intenders do purchase even with perfectly rational respondents.

More recent studies of purchase intentions have developed models that incorporate the discrepancies between stated intentions and actual behavior. The psychometric beta binomial model of Morrison (1979) is a descriptive model of the relationship between stated purchase intent and subsequent purchase. Morrison incorporated the observed discrepancies between intent and purchase by taking into account measurement error of intentions (the difference between true intent and stated intent), the impact of exogenous events (that true intent may change over time), and systematic bias (the systematic tendency for intentions to overestimate or underestimate purchase). Kalwani and Silk (1982) report further analyses and applications of Morrison's model. Bemmaor (1995) more recently extended the Morrison (1979) model to the case of heterogeneous switching probabilities to attempt to explain the sign and magnitude of the discrepancy between overall mean purchase intent and subsequent proportion of buyers. In a different modeling effort, Infosino (1986) assumed that one can interpret a purchase intention response as a monotonic transformation of value (willingness-to-pay minus price), truncated to an integer on the intention scale. Infosino further assumed that, at the individual level, willingness-to-pay is stochastic. He also assumed that exogenous events, such as changes in the marketing mix, shift, but do not change the shape of the distribution of willingness-to-pay. According to this model, the proportion of customers who purchase is equal to the proportion of the transformed distribution of value above some threshold value.

The Morrison (1979) and the Infosino (1986) models both capture systematic biases in intentions measurement. The goal of several other intentions studies has been to identify causes of systematic biases in intent measurement. One determinant of the magnitude and direction of bias in intent measurement is the type of product under consideration. Research by Kalwani and Silk (1982) and Jamieson and Bass (1989) demonstrates that the relationship between purchase intentions and purchase behavior is different for durable goods than for non-durable goods. For example, Kalwani and Silk find that for durable goods, a linear model provides a good fit between intentions and purchase behavior, but for nondurable goods, a piecewise linear model provides a better fit. Morwitz and Schmittlein (1992) find that the relationship between purchase intentions and subsequent purchase also varies across demographic and product usage based segments. They find that some segments of intenders are more likely to fulfill their intentions than others, and that some segments of non-intenders are more likely to purchase despite their intentions than others. One additional cause of systematic bias in intent measurement is the effect of measuring intentions on actual purchasing. Morwitz, Johnson, and Schmittlein (1993) demonstrate that merely asking respondents whether or not they intend to purchase a durable good actually increases subsequent purchasing of the product.

In summary, based on past empirical research on the relationship between intentions and behavior, we know that although intentions are often used as a proxy variable for actual purchase in applied marketing research, not everyone who says they intend to buy will actually purchase, and conversely not everyone who says they do not intend to buy will not make a purchase. The reasons that intentions are not perfect measures of behavior include measurement error in measuring intentions, changes that occur between the time intent is measured and the purchase occasion, systematic biases that might arise from the effect of product characteristics, the effect of respondent characteristics, and the effect of measuring intentions on behavior.

The imperfect correspondence between purchase intentions and behavior suggests that standard (regression) models relating demographic/psychographic covariates to intentions can provide inaccurate estimates of the relationship between these covariates and actual

purchase behavior. In this paper, we develop a model and estimation procedure for analyzing purchase intention data that explicitly take into account the discrepancy between purchase intentions and purchase behavior. While the models previously described are concerned with identifying the specific mechanisms by which purchase intention deviates from purchase behavior, our methodology takes this discrepancy between intention and behavior as a given, and then uses this known discrepancy to help correctly identify correlates of *behavior* based on data from *intentions* only. An extension of the model is developed for the case in which intentions data is available for multiple brands within a product category; this extended model uses hierarchical Bayes methods (e.g., Berger 1985) to achieve optimal pooling of information across brands.

The estimation procedure introduced in this paper makes use of Markov chain Monte Carlo techniques (Gelfand and Smith 1990; Roberts and Smith 1993) to identify model parameters. We show via a simulation study that, when the model underlying the proposed methodology is valid, the proposed method can accurately estimate coefficients driving purchase activity, while methods that ignore the inequivalence between intentions and behavior lead to substantially biased parameter estimates. We also show that, given data on multiple brands, the optimally pooled estimates derived in this paper can provide significant gains in efficiency relative to the natural unpooled estimates. Finally, we consider a marketing dataset collected in Morwitz and Schmittlein (1992) in which personal computer consumers were surveyed regarding their purchase intentions, and then surveyed again, a year later, to discover their actual purchase histories. We demonstrate that our proposed method for analyzing intentions data provides a more accurate estimate of the relationship between covariates and purchase behavior, suggesting that the proposed method may be valuable to researchers working in the usual setting in which only purchase intentions data are available, but in which inferences about purchase behavior are desired.

2 The Model

The model presented here is used for analyzing the simplest form of intentions data, namely binary response questions, but can be generalized to categorical, ordinal, or continuous response. The model states that purchase behavior is related to covariates according to a binary regression model, and that the purchase intention data represents a randomly distorted version of the purchase behavior data.

Let variables y_i and w_i denote purchase intention and purchase behavior, respectively:

$$y_i = \begin{cases} 1 & \text{if consumer } i \text{ indicates intention to purchase the product} \\ & \text{within the designated time horizon,} \\ 0 & \text{otherwise} \end{cases}$$
$$w_i = \begin{cases} 1 & \text{if consumer } i \text{ actually purchases the product} \\ & \text{within a designated time horizon,} \\ 0 & \text{otherwise.} \end{cases}$$

The declaration of intention, or lack of intention, to purchase the product necessarily occurs prior to the actual purchase. The variable w_i is postulated to be directly dependent on covariates \mathbf{x}_i , according to the model:

$$P(w_i = 1) = \Phi(\beta' \mathbf{x}_i), \tag{1}$$

where $\Phi(\cdot)$ is a cumulative probability function; if $\Phi(\cdot)$ is the cumulative normal or cumulative logistic, then (1) denotes a probit or logistic regression equation. The covariates \mathbf{x}_i in (1) typically represent demographic or psychographic characteristics of respondent i ; e.g., age, gender, present job category, etc. The objective in many marketing research studies is to identify the relationship between the covariates \mathbf{x}_i and purchase behavior w_i for market segmentation purposes; i.e., to estimate the parameters β .

Since the w_i 's are *not* observed in a typical marketing research survey, one cannot

perform a probit or logistic regression of w versus \mathbf{x} to infer the coefficients β . If the observed purchase intentions y_i exactly correspond to the w_i , then one can clearly perform binary regression of y versus \mathbf{x} to obtain consistent estimates of the coefficients in (1). Suppose, though, that the observed intentions y_i are only imperfect proxies for the actual purchase actions w_i , related to one another by the probabilities:

$$P(y_i = 1 | w_i = 1) = p_{11}, \quad (2a)$$

$$P(y_i = 0 | w_i = 0) = p_{00}. \quad (2b)$$

If either p_{11} or p_{00} is different from 1, then the observed intentions will not perfectly match the purchase behavior, and so regression of y versus \mathbf{x} will give biased estimates of the relationship between the covariates and actual purchase. We will assume, for the present, that the probabilities p_{11} and p_{00} are known, at least approximately. As seen in Table 1, such probabilities are well established in the marketing literature for many product categories.¹

The model proposed, then, is a corrupted binary regression model; Copas (1988) presents a somewhat similar model in which the corruption probabilities p_{11} and p_{00} are necessarily equal. As in Albert and Chib (1993), the analysis of the model is facilitated by *data augmentation*; i.e., by the introduction of particular latent variables. Let z_i denote a latent utility for each consumer that determines product purchase; in particular, consumer i will purchase whenever the utility z_i is positive. This utility is assumed to be determined by the covariates \mathbf{x}_i according to the linear model:

$$z_i = \beta' \mathbf{x}_i + \epsilon_i, \quad \epsilon_i \sim N(0, 1). \quad (3)$$

¹The probabilities in Table 1 refer to $P(w | y)$, the probability of purchase given the stated intention, while the probabilities p_{11} and p_{00} refer to $P(y | w)$. However, as is discussed in section 2.2, the necessary quantities $P(y | w)$ can be inferred from $P(w | y)$ using Bayes theorem.

The purchase behavior, w_i , then is determined by:

$$w_i = \begin{cases} 1 & \text{if } z_i \geq 0, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

Equations (3) and (4) are in fact equivalent to equation (1). The stated purchase intention depends on w_i as before:

$$P(y_i = j | w_i = k) = p_{jk}, \quad j = 0, 1, \quad k = 0, 1. \quad (5)$$

The data for the problem consist of independent observations (y_i, \mathbf{x}_i) , $i = 1, \dots, n$. The objective of the study is to infer the parameters β relating the \mathbf{x}_i to actual, but unobservable, purchases w_i (which may occur in some future time period). It is this modelling objective, the identification of the demographic factors predictive of purchase behavior, which distinguishes the model presented here from other marketing models for purchase intention and behavior such as those of Morrison (1979) and Bemmaor (1995). As is shown in the following sections, the β coefficients in the model, which measure the effect of covariates on purchase behavior, can be estimated via Bayes estimation implemented with Gibbs sampling (Gelfand and Smith 1990).

2.1 Bayes Estimation with Gibbs Sampling

There are advantages to adopting a Bayesian estimation approach for the problem of analyzing purchase intentions data. One potential benefit of applying a Bayesian analysis to purchase intention data is that the inferences are exact for finite samples (conditional on the assumption that the model (3)–(5) is correctly specified), while confidence intervals obtained through maximum likelihood estimation are valid only asymptotically. Another advantage of the Bayesian approach is that, as is to be discussed in section 2.2.1, it is possible to incorporate uncertainty about the probabilities p_{00} and p_{11} into the estimation procedure. A final advantage is that, if parameter vectors are to be estimated for a number of related product lines, a hierarchical Bayes approach can be used to achieve

an optimal pooling of data across products; this technique can significantly improve estimation accuracy, as is shown in section 2.3.

The posterior for β will be proportional to the likelihood function times the prior. The likelihood function for the parameters β in terms of the observable data (y_i, \mathbf{x}_i) , $i = 1, \dots, n$ is given by:

$$\mathcal{L}(\mathbf{y}|\beta, \mathbf{X}, p_{00}, p_{11}) = \prod_{i=1}^n [P(y_i = 1 | \beta, \mathbf{x}_i, p_{00}, p_{11})^{y_i} P(y_i = 0 | \beta, \mathbf{x}_i, p_{00}, p_{11})^{1-y_i}]. \quad (6)$$

Here,

$$\begin{aligned} P(y_i = 1 | \beta, \mathbf{x}_i, p_{00}, p_{11}) &= \\ P(y_i = 1 | \beta, \mathbf{x}_i, p_{00}, p_{11}, w_i = 1)P(w_i = 1 | \beta, \mathbf{x}_i, p_{00}, p_{11}) &+ \\ P(y_i = 1 | \beta, \mathbf{x}_i, p_{00}, p_{11}, w_i = 0)P(w_i = 0 | \beta, \mathbf{x}_i, p_{00}, p_{11}) & \\ &= p_{11}\Phi(\beta' \mathbf{x}_i) + (1 - p_{00})(1 - \Phi(\beta' \mathbf{x}_i)) \\ &= p_{10} + (p_{11} - p_{10})\Phi(\beta' \mathbf{x}_i), \end{aligned} \quad (7)$$

and

$$\begin{aligned} P(y_i = 0 | \beta, \mathbf{x}_i, p_{00}, p_{11}) &= 1 - P(y_i = 1 | \beta, \mathbf{x}_i, p_{00}, p_{11}) \\ &= p_{00} - (p_{00} - p_{01})\Phi(\beta' \mathbf{x}_i), \end{aligned} \quad (8)$$

where $p_{10} = 1 - p_{00}$, and $p_{01} = 1 - p_{11}$.

Thus, the likelihood function is given by:

$$\mathcal{L}(\mathbf{y}|\beta, \mathbf{X}, p_{00}, p_{11}) = \prod_{i=1}^n \left[(p_{10} + (p_{11} - p_{10})\Phi(\beta' \mathbf{x}_i))^{y_i} \cdot (p_{00} - (p_{00} - p_{01})\Phi(\beta' \mathbf{x}_i))^{1-y_i} \right] \quad (9)$$

Regardless of the form of the prior, the posterior distribution for β will not be from

a standard family, so quantities like the posterior mean and standard deviation cannot be computed analytically, but can only be obtained through some numerical integration approach. Here, we use a Markov chain Monte Carlo method (Roberts and Smith 1993) to perform this integration, by generating a Markov chain whose steady-state distribution is equal to the posterior distribution of β . One particular Monte Carlo technique which is convenient for this problem is Gibbs sampling (Gelfand and Smith 1990; Roberts and Smith 1993; Albert and Chib 1993). Gibbs sampling is a multivariate random number generation technique that may be used in the frequent case in which a joint density to be sampled from is complex, but in which the conditional distributions are simple. As is described in Tanner (1993), the Gibbs sampler works by sampling alternatively from the full conditional posterior distribution of each parameter in turn. Under appropriate conditions (Tierney 1994), the stationary distribution of the chain produced by this alternating technique is in fact the desired, complex, multivariate posterior density of the unknown model parameters. Since the chain is ergodic (Geman and Geman 1984), the posterior mean and standard deviation of the parameter of interest, β , can be obtained by simply computing the sample mean and standard deviation over the generated samples (in practice, the first several hundred samples generated are typically discarded, to avoid transient effects). Allenby and Lenk (1994) and McCulloch and Rossi (1994) present further applications of Gibbs sampling to marketing models.

In the context of the purchase intentions data, the entire set of unobservable parameters consist of β , as well as the z_i and the w_i . The full conditional distributions for these unknown parameters are all easily obtained. If the z_i 's are known, and if the prior on β is the non-informative uniform distribution, then the conditional distribution for the coefficient β is obtained by standard results on Bayesian linear models (e.g., Zellner 1971):

$$\pi(\beta|z, y, \mathbf{X}) \sim N((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}, (\mathbf{X}'\mathbf{X})^{-1}), \quad (10)$$

where:

$$\mathbf{z} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} 1 & \mathbf{x}_1 \\ \vdots & \\ 1 & \mathbf{x}_n \end{pmatrix}.$$

The mean of the full posterior conditional distribution for β in (10) is just the usual ordinary least-squares estimate obtained by regressing z versus \mathbf{x} , and the variance of the distribution is the usual sampling variance for the least squares estimate of β .

If one has available an informative prior on β which is multivariate normal, with mean μ and covariance \mathbf{T}^{-1} , then the conditional posterior for beta is also of simple form:

$$\pi(\beta|z, y, \mathbf{x}, \mu, \mathbf{T}) \sim N\left((\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}(\mathbf{X}'\mathbf{z} + \mathbf{T}\mu), (\mathbf{X}'\mathbf{X} + \mathbf{T})^{-1}\right). \quad (11)$$

Also, as is discussed in section 2.3, it is possible to estimate the informative prior empirically, given purchase intention data on multiple related brands.

If β and the w_i are known, then the full conditional of z_i is obtained as:

$$\pi(z_i|\beta, w, y, \mathbf{x}) \sim \begin{cases} N(\beta'\mathbf{x}_i, 1)I(z_i \geq 0) & \text{if } w_i = 1 \\ N(\beta'\mathbf{x}_i, 1)I(z_i < 0) & \text{if } w_i = 0. \end{cases} \quad (12)$$

That is, if $w_i = 1$, then z_i is restricted to be positive, while if $w_i = 0$, then z_i is restricted to be negative (Albert and Chib 1993).

Finally, the unobservable w_i have to be generated as well. Their posterior distributions can be obtained via Bayes theorem:

$$\pi(w_i = k|\beta, y_i = j, \mathbf{x}_i) = \frac{\pi(y_i = j|w_i = k, \beta, \mathbf{x}_i)\pi(w_i = k|\beta, \mathbf{x}_i)}{\sum_{l=0}^1 \pi(y_i = j|w_i = l, \beta, \mathbf{x}_i)\pi(w_i = l|\beta, \mathbf{x}_i)}. \quad (13)$$

The prior, $\pi(w_i = 1|\beta, \mathbf{x}_i)$, is given by $\Phi(\beta'\mathbf{x}_i)$, by equations (3) and (4). The likelihood, $\pi(y_i = 1|w_i = 1, \beta, \mathbf{x}_i)$, is given by p_{11} , from equation (5). The necessary full conditional

probabilities are given by:

$$\pi(w_i = 1 | \beta, y_i = 1, \mathbf{x}) = \frac{p_{11}\Phi(\beta'\mathbf{x}_i)}{p_{11}\Phi(\beta'\mathbf{x}_i) + p_{10}(1 - \Phi(\beta'\mathbf{x}_i))} \quad (14a)$$

$$\pi(w_i = 1 | \beta, y_i = 0, \mathbf{x}) = \frac{p_{10}\Phi(\beta'\mathbf{x}_i)}{p_{10}\Phi(\beta'\mathbf{x}_i) + p_{11}(1 - \Phi(\beta'\mathbf{x}_i))}. \quad (14b)$$

The Gibbs sampler then alternates between the three steps of generating the coefficients β from (10), generating the z_i from (12), and generating the w_i from (14a–14b). This sequence forms a Markov chain; at convergence, the values of β come from the desired posterior distribution. These values reflect the imperfect match between w_i and y_i ; the correction is accomplished through appropriate random generation of the w_i , by equations (14a) and (14b). The chain can be started at some rational starting point, for example with $w_i = y_i$, and with $z_i = 1$ or -1 according to the value of y_i . If the values of β generated by the Markov chain scheme are denoted $\beta^{(1)}, \dots, \beta^{(G)}$, then the usual Bayes estimate for β is the sample mean of the $\beta^{(g)}$, ignoring the first B samples; i.e., $\hat{\beta} = (G - B) \sum_{g=B+1}^G \beta^{(g)}$. The posterior variance is similarly estimated as the sample variance for the last $G - B$ iterates. The first B samples are discarded in order to guarantee that the Markov chain has reached the correct steady state distribution.

Considerable research has been devoted to diagnosing convergence of the Gibbs sampler; that is, to identifying acceptable values of G and B above. Tanner (1993) provides guidance for implementation of Gibbs sampling techniques, and Roberts (1992), Robert (1995), and Zellner and Min (1995) describe diagnostic measures for assessing convergence of the Markov chain.

2.2 Determination of p_{11} and p_{00}

The quantities p_{11} and p_{00} defined in (2) are the probabilities for intentions conditional on actual purchase behavior, while the probabilities described in Table 1 describe the probabilities of purchase behavior conditional on purchase intention. Since it is these latter probabilities for which published historical values (or product class level estimates vis a vis

industry standards) exist, it would be desirable to be able to express the required inputs p_{11} and p_{00} in terms of these more familiar probability figures. Let $q_{jk} = P(w = j | y = k)$, so that, for example, q_{11} denotes the proportion of survey respondents who indicate an intention to buy that actually do complete a purchase. The p_{jk} can be written in terms of the q_{jk} via Bayes theorem:

$$\begin{aligned}
 p_{11} &= P(y = 1 | w = 1) \\
 &= \frac{P(w = 1 | y = 1) P(y = 1)}{P(w = 1 | y = 1) P(y = 1) + P(w = 1 | y = 0) P(y = 0)} \\
 &= \frac{q_{11} P(y = 1)}{q_{11} P(y = 1) + q_{10} P(y = 0)}, \tag{15}
 \end{aligned}$$

$$\begin{aligned}
 p_{00} &= P(y = 0 | w = 0) \\
 &= \frac{P(w = 0 | y = 0) P(y = 0)}{P(w = 0 | y = 0) P(y = 0) + P(w = 0 | y = 1) P(y = 1)} \\
 &= \frac{q_{00} P(y = 0)}{q_{00} P(y = 0) + q_{01} P(y = 1)}. \tag{16}
 \end{aligned}$$

The quantities $P(y = 1)$ and $P(y = 0)$ can be calculated by the marketing researcher based on the survey of intentions at hand, since $P(y = 1)$ simply denotes the proportion of respondents indicating an intention to buy. Thus, using these values, and the industry standards for q_{00} and q_{11} as listed in Table 1, one can obtain approximate estimates of the required probabilities p_{jk} via equations (15) – (16).

2.2.1 Uncertainty about p_{11} and p_{00}

Since the appropriate probabilities p_{11} and p_{00} may not be known exactly, one may wish to incorporate uncertainty about these parameters into the estimation procedure. Instead of regarding these probabilities as fixed and known, one can instead assign prior distributions to these quantities, and can then draw random values of p_{00} and p_{11} from their full conditional posterior distributions at each iteration of the Markov chain. The

beta probability distribution provides a flexible family of distributions on $[0, 1]$ that can be used to express knowledge about p_{11} and p_{00} . Suppose that the prior for p_{11} is $\text{beta}(\alpha_1, \alpha_0)$: $\pi(p_{11}) \propto p_{11}^{\alpha_1-1}(1-p_{11})^{\alpha_0-1}$. Let n_{jk} denote the number of observations in the dataset for which $y_i = j$ and $w_i = k$. Then the full conditional posterior density for p_{11} is proportional to the prior times the likelihood, and the likelihood is proportional to $p_{11}^{n_{11}}(1-p_{11})^{n_{01}}$; thus, the full conditional posterior is proportional to $p_{11}^{\alpha_1+n_{11}-1}(1-p_{11})^{\alpha_0+n_{01}-1}$. I.e., the density is $\text{beta}(\alpha_1 + n_{11}, \alpha_0 + n_{01})$. Similarly, if the prior for p_{00} is $\text{beta}(\gamma_0, \gamma_1)$, then the full conditional posterior for p_{00} is $\text{beta}(\gamma_0 + n_{00}, \gamma_1 + n_{10})$. Note that the data is informative about p_{11} and p_{00} , since these parameters appear in the likelihood function for the observable data (y_i, \mathbf{x}_i) in equation (9).

2.3 Hierarchical Bayes Estimation

Frequently, marketing researchers will obtain purchase intention data on multiple brands and/or product lines within a common product class; a typical survey may ask the survey respondents: “Do you intend to purchase Brand 1 in the next year? Do you intend to purchase Brand 2 in the next year? ... Do you intend to purchase Brand M in the next year?” Given this data on multiple (M) brands, one may wish to identify the coefficient vector β_k for each brand k , $k = 1, \dots, M$.

One approach to estimating the β_k parameter vectors is to simply estimate model (3)–(5) separately for each brand, in effect, assuming that the β_k are completely unrelated. Alternatively, one could make the strong assumption that the β_k vectors for the different brands are in fact all equal; under this assumption, one could then pool the data from the M brands, thus considerably multiplying the amount of data available. The advantage of the former approach, of estimating the M models separately, is that it makes no assumptions about the data, and thus cannot introduce bias. The advantage of the latter approach, pooling the data, is that it increases the effective sample size, and hence reduces the sampling variance of the estimates. The former method will be superior in cases when the vectors β_k are very dissimilar for the different brands, while the latter will be superior when the β_k are homogeneous. To appreciate the potential gain from pooling, if the β_k

are truly all equal and data from M brands are pooled, the sampling standard error of the parameter estimates will decrease by a factor of \sqrt{M} ; thus, with 9 brands, the unpooled estimates will have 200% larger sampling error than the pooled estimates.

Hierarchical Bayes and empirical Bayes methods (e.g., Berger 1985) are designed to obtain the optimal tradeoff between a completely unpooled and a completely pooled estimator. The methods are based on a supposition that the β_k vectors are not identical, but that they arise from a common distribution. Let $\pi(\beta)$ denote this common prior distribution, and assume that the distribution is multivariate normal, with mean μ and covariance $\Sigma = \mathbf{T}^{-1}$. In the Gibbs sampling framework, each vector β_k will be generated using equation (11), based on the informative prior $\pi(\beta) \sim N(\mu, \Sigma)$. From (11), if \mathbf{T} is zero, indicating the variance of the β 's, Σ , is very large and the brands are heterogeneous, then the posterior mean for β_k will be $(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{z}_k$ which is equivalent to the mean given in equation (10), the formula for the simple, unpooled estimate for brand k . Alternatively, if \mathbf{T} is infinite, indicating that the brands are completely homogeneous, then the conditional posterior mean for β_k is equal to μ for all k ; i.e., the estimate behaves like a pooled estimate.

Thus, if appropriate prior parameters μ and Σ are available, the Bayes estimator behaves as desired, acting as a disaggregate estimator when the brands are heterogeneous, and as a pooled estimator when the brands are homogeneous. However, because response data is available on multiple brands, the prior parameters can in fact be estimated from the data themselves. Suppose at a given iteration in the Gibbs sampling scheme, one has samples $(\beta_1, \dots, \beta_M)$ for the M brands. Then sensible point estimates of the prior parameters are:

$$\hat{\mu} = M^{-1} \sum_{k=1}^M \beta_k \quad (17)$$

$$\hat{\Sigma} = M^{-1} \sum_{k=1}^M (\beta_k - \hat{\mu})(\beta_k - \hat{\mu})'; \quad (18)$$

i.e., the prior mean and covariance for the β 's is estimated by the sample mean and

covariance of the β 's. For the true hierarchical Bayes method, one would not use a point estimate of (μ, Σ) but would instead sample values from their respective conditional posteriors. To obtain these conditional posteriors, one must specify a prior on the hyper-parameters. A hyper-prior on (μ, Σ) which is proportional to $\Sigma^{-\frac{p}{2}}$, with p denoting the dimension of β , can be said to represent "maximal uncertainty" about these parameters (Wakefield et al. 1994). Under this non-informative prior, the conditional posteriors for the hyper-parameters are:

$$\pi(\mu|\beta_1, \dots, \beta_k, \Sigma) \sim N\left(M^{-1} \sum_{k=1}^M \beta_k, M^{-1} \Sigma\right) \quad (19)$$

$$\pi(\Sigma^{-1}|\beta_1, \dots, \beta_k, \mu) \sim W\left(M + p, \sum_{k=1}^M (\beta_k - \mu)(\beta_k - \mu)'\right), \quad (20)$$

where $W(d, A)$ denotes the Wishart distribution with d degrees of freedom and expected value dA^{-1} (Wakefield et al. 1994). Odell and Feiveson (1966) present a procedure for generating deviates from the Wishart matrix distribution.

To summarize, the Gibbs sampling estimation algorithm for the hierarchical Bayes model involves iterative repetition of the following steps:

1. For each brand k , generate the vector β_k using the prior $\pi(\beta) \sim N(\mu, \Sigma)$, via equation (11).
2. For each brand k and subject i , generate the values w_{ik} using equation (14), conditional on the current value for β_k .
3. For each brand k and subject i , generate the values z_{ik} using equation (12), conditional on the current value for β_k .
4. Given the values β_1, \dots, β_M , and Σ , generate the prior mean μ using equation (19).
5. Given the values β_1, \dots, β_M , and μ , generate the prior covariance Σ using equation (20).

The procedure can be initialized by, for example, setting $\mu = \mathbf{0}$, $\Sigma = \mathbf{I}$. See Lenk and Rao (1990), Allenby and Lenk (1994) and Allenby and Lenk (1995) for further applications of

hierarchical Bayes models in marketing science.

2.4 Potential Degeneracies

There is a potential problem that can arise with model (3) – (5), regardless of estimation technique employed. The problem is that under certain conditions, the likelihood is maximized for infinite values of the β coefficients. To illustrate, consider a simple case in which there is just one covariate, x , which is binary, taking values 0 or 1. Then, from equation (7), $P(y_i = 1 | \beta, x = 0, p_{00}, p_{11}) = p_{10} + (p_{11} - p_{10})\Phi(\beta_0)$; similarly, $P(y_i = 1 | \beta, x = 1, p_{00}, p_{11}) = p_{10} + (p_{11} - p_{10})\Phi(\beta_0 + \beta_1)$. If π_0 denotes the empirical proportion of observations for which $y_i = 1$ given that $x_i = 0$, and π_1 denotes the proportion of observations for which $y_i = 1$ given that $x_i = 1$, then the MLE's of (β_0, β_1) can be found simply by setting the probabilities implied by the model to be equal to the observed proportions:

$$\pi_0 = p_{10} + (p_{11} - p_{10})\Phi(\beta_0) \quad (21)$$

$$\pi_1 = p_{10} + (p_{11} - p_{10})\Phi(\beta_0 + \beta_1), \quad (22)$$

and the solution is:

$$\hat{\beta}_0 = \Phi^{-1}\left(\frac{\pi_0 - p_{10}}{p_{11} - p_{10}}\right) \quad (23)$$

$$\hat{\beta}_1 = \Phi^{-1}\left(\frac{\pi_1 - p_{10}}{p_{11} - p_{10}}\right) - \hat{\beta}_0. \quad (24)$$

The degeneracy arises if the argument of $\Phi^{-1}(\cdot)$ in (23) or (24) falls outside the interval $(0, 1)$; this event occurs if either π_0 or π_1 falls outside the interval (p_{10}, p_{11}) . In such a degenerate situation, the likelihood function is maximized for coefficient values that are infinite. The model given by equations (3)–(5) implies that for any value of x , the conditional probability that $y = 1$ must be between p_{10} and p_{11} ; the degeneracy arises when the data conflicts with this assumption.

Degenerate cases are diagnosed by observing the Gibbs samples of the β coefficients

wandering toward infinity. Degeneracy may, but does not necessarily, indicate that the values assumed for p_{00} and p_{11} are inappropriate for the given dataset. The degeneracy can be prevented by putting constraints or bounds on the parameters. In the Gibbs sampling framework, this implies rejecting any generated set of coefficients β that fall outside of some “reasonable” neighborhood. Alternatively, the use of a proper prior distribution on β , such that the prior probability of infinite β is sufficiently small, will prevent the incidence of the degeneracy.

3 A Simulation Study

A modest simulation study was performed to evaluate the potential ability of the Bayes estimators to recover the true parameters relating covariates to product purchase, based just on data from purchase intentions. Data were generated from the model described in (3) – (5), with $\beta = (\beta_0, \beta_1)$, $\mathbf{x}_i = (1, x_i)$, and with the x_i generated uniformly over $[0,1]$. Four different simulation blocks, with different sets of generating parameters, were evaluated. Table 2 lists the parameter values $(n, \beta_0, \beta_1, p_{00}, p_{11})$ used in the study. The values were chosen to examine the performance of the estimation method over a range of possible parameter settings. Table 2 also displays the results obtained by analyzing the data using both the Gibbs sampling technique described in this paper, and the results obtained using the naive approach of simply performing maximum likelihood probit regression of the intention data y_i versus the \mathbf{x}_i . This latter approach is essentially equivalent to applying the method of this paper, but erroneously assuming that $p_{00} = p_{11} = 1$. The Gibbs sampling procedure was run for 2000 iterations, and the final Bayes estimate was chosen as the mean of the last 1500 samples generated.

[Insert Table 2 Here]

In the first simulation block, a modest amount of discrepancy between intentions and purchase was assumed: $p_{00} = p_{11} = 0.8$. Note that even for this apparently small amount of error, the naive parameter estimates are grossly biased; the mean naive estimate for

β_1 was 0.63, though the true value was 3.0. The Bayes estimator proposed in this paper has much less bias, with a mean of 3.93. When the measurement error is increased to $p_{00} = 0.9$, $p_{11} = 0.6$, the Bayes estimator is still able to recover the parameters governing purchase behavior (the w_i) based on data from simulated intentions (y_i). The Bayes estimator appears to be unbiased in Block 3, with $\beta_0 = 2$, $\beta_1 = -3$, but the standard deviation is more than twice that in Block 2, with $\beta_0 = -2$, $\beta_1 = 3$. The reason for the different performances in Blocks 2 and 3 is that, given the different coefficients, more w_i 's have a value of 0 in Block 3 than in Block 2, and since $p_{00} > p_{11}$ for these blocks, there will be more misclassified y_i 's with Block 3 than Block 2.

The generally higher variance of the Bayes estimator relative to the naive estimator seems to be due to the fact that the probability function for the data y_i assumed with the Bayes estimator is less sensitive to changes in β_1 than is the model used in the naive approach. For the Bayes model, the probability function is:

$$P(y_i = 1 \mid \beta, \mathbf{x}_i, p_{00}, p_{11}) = p_{10} + (p_{11} - p_{10})\Phi(\beta' \mathbf{x}_i), \quad (25)$$

while for the naive estimator, the probability function is simply:

$$P(y_i = 1 \mid \beta, \mathbf{x}_i) = \Phi(\beta' \mathbf{x}_i). \quad (26)$$

The sensitivities to changes in the coefficients are, for the Bayes estimator,

$$\frac{\delta^2}{\delta\beta_k\delta\beta_r} P(y_i = 1 \mid \beta, \mathbf{x}_i, p_{00}, p_{11}) = (p_{11} - p_{10})^2 x_{ik} x_{ir} \dot{\phi}(\beta' \mathbf{x}_i), \quad (27)$$

and for the naive estimator,

$$\frac{\delta^2}{\delta\beta_k\delta\beta_r} P(y_i = 1 \mid \beta, \mathbf{x}_i) = x_{ik} x_{ir} \dot{\phi}(\beta' \mathbf{x}_i). \quad (28)$$

The quantity $p_{11} - p_{10} = p_{11} + p_{00} - 1$ must be less than or equal to 1 in absolute value, equalling 1 only when $p_{11} = p_{00} = 1$, or when $p_{11} = p_{00} = 0$.² The relative insensitivity of

²The latter case indicates a condition in which all respondents claiming an intention to purchase do not in fact purchase, while all respondents denying an intention to purchase do purchase. In this very

the probability function when p_{11} or p_{00} is less than 1 corresponds to a relatively flatter likelihood surface, and hence smaller Fisher's information.

3.1 Hierarchical Bayes Estimation

A further modest simulation study was performed to evaluate the potential gains from using the hierarchical Bayes (HB) approach described in section 2.3. In this study, data from M brands were generated from the model:

$$z_{ik} = \beta'_k \mathbf{x}_i + \epsilon_{ijk}, \quad \epsilon_{ijk} \sim N(0, 1) \quad (29)$$

$$w_{ik} = 1 \text{ if } z_{ik} > 0, 0 \text{ otherwise} \quad (30)$$

$$P(y_{ik} = 1 \mid w_{ik} = 1) = p_{11} \quad (31)$$

$$P(y_{ik} = 0 \mid w_{ik} = 0) = p_{00}, \quad (32)$$

with the β_k distributed normally across brands: $\beta_k \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, $k = 1, \dots, M$. In each simulation trial $\boldsymbol{\Sigma}$ was equal to $\sigma^2 I$ for some scalar σ . Small values of σ indicate homogeneity of β_k across brands, and thus a favorable condition for pooling of data. The coefficients β_k were estimated two ways: using the hierarchical Bayes procedure described in section 2.3, and by a disaggregate estimator in which the Bayes estimator with non-informative priors is applied to each brand separately. Both the hierarchical Bayes and the non-informative Bayes estimates were based on correct values of the p_{ij} probabilities. For all the simulation replications, p_{11} and p_{00} were set to 0.9, and the prior mean $\boldsymbol{\mu}$ was set to $(-2, 3)$. The cross-sectional variance σ , the number of subjects n , and the number of brands M were varied to discover the benefit of hierarchical Bayes methods under different conditions. For each replication, the accuracy was measured by the root mean squared error for estimating the M coefficient vectors β_k : i.e., the RMSE for estimating

untypical situation, a researcher could reverse the intentions data, and apply simple binary regression to the reversed data, to obtain an accurate model for purchase behavior.

β_0 in a given replication was given by $(\frac{1}{M} \sum_{k=1}^M (\beta_{k0} - \beta_{k0}^*)^2)^{1/2}$, where β_{k0}^* denotes the true value for the parameter.

Table 3 lists the estimation results. As can be seen, the gains in estimation accuracy from using the hierarchical Bayes approach are significant. For example, in Block 1, the mean RMSE for the HB estimate of β_1 is 0.736, while the mean RMSE for the non-informative Bayes estimate is 1.051; thus, the pooling achieved by the HB approach reduces estimation error by 43%. As expected, the gains are greatest in cases in which the number of subjects n is small, and when the cross-sectional variation σ is small. Even if the cross-sectional variation in the β_k is fairly large ($\sigma = 0.5$), the gains from using HB appear to be non-negligible. In the limit, if σ is large, or n is large, the HB method will essentially reduce to the simple Bayes estimate, which will be similar to the non-informative Bayes estimate. Thus, for small cross-sectional variation, the HB method will significantly outperform the non-informative Bayes estimate, whereas the two methods should have comparable performance under conditions less favorable to HB.

[Insert Table 3 Here]

3.2 Parameter Estimation: A Summary

In section 3, it was demonstrated that equating intentions with actual behavior can lead to substantially biased inferences, and that, given knowledge of the relationship between intentions and behavior, one can obtain relatively unbiased estimates of the model relating covariates to purchase behavior, by using the technique described in section 2.1. In section 3.1, it was shown that, given intentions data on multiple brands, one may be able to achieve substantial gains in estimation efficiency by pooling the data through hierarchical Bayes methods.

It is worth noting here the assumptions and limitations underlying the methodology. Model (3)–(5) assumes that correlations between the covariates x and stated intentions y are explained strictly through the relationships between the variables x and actual purchase behavior w , and through the relationship between purchase behavior w and

intention y . Thus, it is assumed that the parameters p_{00} and p_{11} relating behavior and intentions are independent of \mathbf{x} . Also, it is assumed that one has fairly informative prior knowledge concerning the values p_{00} and p_{11} . While section 2.2.1 shows that the observable intentions data will be informative about these parameters, in practice it will be difficult to get accurate estimates of the probabilities, unless one has substantial numbers of observations with extremely large, and with extremely small, values of $\beta'x_i$, thus affording a view of the tails of the probability function $P(y_i = 1 | \beta, x_i, p_{00}, p_{11})$ given in equation (7).

4 An Application Involving Personal Computers

Morwitz and Schmittlein (1992) present a study in which members of a large consumer panel were surveyed multiple times concerning their intentions to buy and their actual purchasing of personal computers. In addition, extensive demographic and product usage variables about the panel households were obtained. At the time of the first survey, none of the sample households currently owned a personal computer. During the first survey, respondents were asked the following intention question:

Do you or does anyone in your household plan to acquire a personal computer in the future for use at home?

- Yes, in the next 6 months
- Yes, in 7 to 12 months
- Yes, in 13 to 24 months
- Yes, sometime, but not within 24 months
- No, will not acquire one

One year later, the same sample of respondents indicated whether or not they had purchased since the first survey.

Note that the intention question asked to panel households is not a binary intentions question, but rather is a variation of a standard binary question that contains a purchase

timing element. In order to estimate the model, we classify those respondents who indicated that they intended to purchase within the one year time frame of interest as intenders (i.e., they either intend to buy in 6 months, or in 7 to 12 months). Respondents who indicated that they did not intend to purchase a personal computer and those who indicated that they intend to buy sometime after the first year are classified as non-intenders.

The covariates available to predict purchase probability included household income, profession, geographic region, and current level of PC usage. Due to the confidential nature of the dataset, we have rescaled the data and shuffled labels of the predictor variables; however, these operations should not effect the suitability of the model for the application. The particular set of variables included in the analyses reported here were:

PROFESS = 1 if a professional; 0 otherwise
 STUDENT = 1 if a student; 0 otherwise
 USE = 1 if a household member uses a PC at work or school;
 0 otherwise
 PACIFIC = 1 if house is in pacific region of U.S.; 0 otherwise
 INCOME = household income.

To achieve approximately equal scaling for the variables, income was coded in \$100,000's; e.g., a respondent with income of \$60,000 was coded as 0.6 for the INCOME variable.

The binary intentions data y_i were analyzed under several different specifications for the parameters (p_{00}, p_{11}) : Analysis 1, with $p_{00} = 0.691$, $p_{11} = 0.812$, corresponding to the actual conditional probabilities observed with the given dataset; Analysis 2, with $p_{00} = 1.0$, $p_{11} = 1.0$, corresponding to the naive assumption that stated purchase intention corresponds exactly to purchase behavior, and is thus equivalent to simple probit regression analysis of the intentions data; Analysis 3, with $p_{00} = 0.658$, $p_{11} = 0.907$, obtained by using the values of $q_{00} = 0.962$, $q_{11} = 0.429$ reported in Jamieson and Bass (1989) for purchases of personal computers, and applying formulas (15) and (16). In addition, an Analysis 4 was performed, in which the true actual purchase responses w_i were analyzed using the standard probit model. The Gibbs sampling procedure was run

for 10,000 iterations, and the coefficient estimates were taken as the means of the final 5,000 samples. Table 4 lists the estimated coefficients and standard errors for the different analyses.

[Insert Table 4 Here]

The coefficient estimates based on Analysis 1 of the purchase intentions, with p_{00} and p_{11} chosen based on the corresponding values in the dataset under analysis, are very close to the values obtained from simple probit regression of the purchase responses as presented in Analysis 4. Perhaps one interesting finding is the significant impact of geographic region (PACIFIC) on purchase behavior, even when controlling for income and professional status. One possible explanation is that PACIFIC is acting as a proxy for employment in the computer industry, and that such employees would be more likely than other professionals to purchase a personal computer.

The coefficient estimates based on Analysis 2 of the intentions data, with p_{00} and p_{11} taken to be equal to 1.0, are quite far from the values obtained in Analysis 4, though the order of the magnitudes appears to be correct. Analysis 3, using values for p_{00} and p_{11} based on the intention conversion probabilities observed in Jamieson and Bass (1989) provides estimates that are somewhat close to the estimates of Analysis 4, and substantially less biased than when p_{00} and p_{11} are taken to be 1.

The coefficient estimates from Analysis 2 all appear to be substantially attenuated; i.e., the impact of the covariates on purchase probability is underestimated by the analysis of the purchase intentions. Ignoring the discrepancy between intentions and behavior would lead an analyst to conclude that the covariates have weak effects on purchase probability, when in fact the effects are fairly strong. The phenomenon of measurement error leading to underestimation of regression coefficients is common in econometrics (e.g., Fuller 1989). In the case of ordinary regression, it is measurement error in the independent variables which leads to attenuated parameter estimates, whereas measurement error added to the dependent variable does not lead to bias in OLS estimates of regression coefficients. In the case of binary regression models, measurement error in the independent variables is

known to lead to biased parameter estimates (e.g., Stefanski and Carroll 1985), but here it is seen that error in the response variable leads to bias in the parameter estimates as well.

Finally, as expected, the standard errors of the coefficient estimates for analyses 1 and 3 are considerably higher than those for analyses 2 and 4; since analyses 1 and 3 acknowledge the presence of uncertainty about true purchase behavior, the uncertainty about the parameter estimates is correspondingly greater.

As a simple test of the assumption of conditional independence of y and w given x , a probit regression analysis was run of the y_i vs. (w_i, x_i) . The results, which are presented in Table 5, show that the coefficients for the covariates x are not significant. Non-significance does not prove the model is correct; however these results suggest that the assumption of conditional independence between y and x given w may be reasonable for these data. In usual practice, one will not be able to empirically verify the assumption that the p_{jk} are independent of the covariates x , since data on actual behavior are typically unavailable.

[Insert Table 5 Here]

In summary, it appears that ignoring the discrepancy between purchase intentions and purchase behavior leads to grossly inaccurate estimates for binary regression coefficients, and that taking the discrepancy into account through use of model (3)–(5) can lead to improved accuracy in parameter estimation. Further, significant improvement can be obtained even if the correct parameters p_{00} and p_{11} are not known precisely. In practice, a researcher may wish to fit the model for a variety of different values of p_{00} and p_{11} , in order to assess the sensitivity of the inferences with respect to assumptions about the conversion of intentions to actual purchases.

5 Alternative Models

Equations (3)–(5) describe one possible model for characterizing both the effect of covariates on purchase behavior, as well as the discrepancy between stated intentions and

actual behavior. As an alternative model, one could suppose that z_i and w_i are defined as in (3) and (4), but that y_i , the stated purchase intention, equals 1 if and only if $z_i + \delta_i$ is positive, where $\delta_i \sim N(\mu, \sigma^2)$.³ The parameter σ here measures the correlation between intentions and behavior, and the parameter μ measures the asymmetry of the switching process. The naive model, in which intentions are treated as equivalent to behavior, is characterized by $\mu = \sigma = 0$. Reasonable values for μ and σ^2 could be inferred from studies such as are listed in Table 1,

This alternative specification implies that, for the observable data y_i ,

$$P(Y_i = 1|\mathbf{x}) = \Phi\left(\frac{\beta'\mathbf{x} + \mu}{\sqrt{1 + \sigma^2}}\right), \quad (33)$$

while for the model (3)–(5), the corresponding probability is:

$$P(Y_i = 1|\mathbf{x}) = p_{10} + (p_{11} - p_{10})\Phi(\beta'\mathbf{x}). \quad (34)$$

A choice between the two models could be based on the shape of the empirical probability curve $P(Y = 1|\mathbf{x})$. Log-likelihood ratios and criteria such as AIC could provide a scheme for choosing between different model specifications.

6 Conclusion

The models and analyses of Morrison (1979), Infosino (1986), Manski (1990), and Bemmaor (1995) provide interesting insights into the psychological mechanisms by which purchase intentions deviate from actual purchase behavior. These and other papers have documented the typical magnitude of the discrepancy between intentions and behavior. The present paper takes this discrepancy between intentions and behavior as a given, and develops an estimation procedure for identifying the demographic/psychographic correlates of *behavior* based on data on *intentions*. When data on multiple brands from a

³We thank an anonymous referee for suggesting this interesting model.

common product class is available, the hierarchical Bayes method introduced in this paper can be used to obtain optimal pooling of data across brands. We show, with simulated data and with data from a recent marketing study, that the proposed estimation method can be useful in improving upon the accuracy of estimates of the relationship between the covariates and true purchase behavior. This ability to make better use of intentions data may lead to more accurate market segmentation, and thus improved marketing management decisions. Future research will attempt to extend the current model to include multi-level intentions scales, to explicitly take into account the elapsed time between intent measurement and purchase occasion, and to incorporate latent class and multivariate generalizations.

References

- Adams, F. G. (1974). Commentary on McNeil, 'Federal Programs to Measure Consumer Purchase Expectations'. *Journal of Consumer Research* 1, 11–12.
- Albert, J. and S. Chib (1993). Bayesian Analysis of Binary and Polychotomous Response Data. *Journal of the American Statistical Association* 88, 669–679.
- Allenby, G. and P. Lenk (1994). Modelling Household Purchase Behavior with Logistic Normal Regression. *Journal of the American Statistical Association* 89, 1218–1231.
- Allenby, G. and P. Lenk (1995). Reassessing Brand Loyalty, Price Sensitivity, and Merchandising Effects on Consumer Brand Choice. *Journal of Business and Economic Statistics* 13, 281–289.
- Bemmaor, A. (1995). Predicting Behavior from Intention-to-Buy Measures: The Parametric Case. *Journal of Marketing Research* 32, 176–191.
- Berger, J. (1985). *Statistical Decision Theory and Bayesian Analysis*. New York, NY: Springer Verlag.
- Clawson, C. (1971). How Useful are 90-Day Purchase Probabilities? *Journal of Marketing Research* 35, 43–47.
- Copas, J. (1988). Binary Regression Models for Contaminated Data. *Journal of the Royal Statistical Society, Series B* 50, 225–265.
- Ferber, R. and R. Piskie (1965). Subjective Probabilities and Buying Intentions. *Review of Economics and Statistics* 47, 322–325.
- Fishbein, M. and I. Ajzen (1975). *Belief, Attitude, Intention, and Behavior*. Reading, MA: Addison-Wesley.
- Fuller, W. (1989). *Measurement Error Models*. New York, NY: John Wiley and Sons.
- Gelfand, A. and A. Smith (1990). Sampling Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85, 398–409.

- Geman, S. and D. Geman (1984). Stochastic Relaxation, Gibbs Distributions and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 6, 721-741.
- Gormley, R. (1974). A Note on Seven Brand Rating Scales and Subsequent Purchase. *Journal of Market Research Society* 16, 242-244.
- Granbois, D. and J. Summers (1975). Primary and Secondary Validity of Consumer Purchase Probabilities. *Journal of Consumer Research* 1, 31-38.
- Infosino, W. (1986). Forecasting New Product Sales from Likelihood of Purchase Ratings. *Marketing Science* 5, 372-384.
- Jamieson, L. and F. Bass (1989). Adjusting Stated Intention Measures to Predict Trial Purchase of New Products: A Comparison of Models and Methods. *Journal of Marketing Research* 26, 336-345.
- Juster, F. (1966). Consumer Buying Intentions and Purchase Probability: An Experiment in Survey Design. *Journal of the American Statistical Association* 61, 658-696.
- Kalwani, M. U. and A. J. Silk (1982). On the Reliability and Predictive Validity of Purchase Intention Measures. *Marketing Science* 1, 243-286.
- Lenk, P. J. and A. G. Rao (1990). New Models from Old: Forecasting Product Adoption by Hierarchical Bayes Procedures. *Marketing Science* 9, 42-53.
- Manski, C. (1990). The Use of Intentions Data to Predict Behavior: A Best-Case Analysis. *Journal of the American Statistical Association* 85, 934-940.
- McCulloch, R. and P. E. Rossi (1994). An Exact Likelihood Analysis of the Multinomial Probit Model. *Journal of Econometrics* 64, 207-240.
- McNeil, J. M. (1974). Federal Programs to Measure Consumer Purchase Expectations, 1946-73: A Post-Mortem. *Journal of Consumer Research* 1, 1-10.
- Morrison, D. G. (1979). Purchase Intentions and Purchase Behavior. *Journal of Marketing* 43, 65-74.

- Ⓢ Morwitz, V. G., E. J. Johnson, and D. Schmittlein (1993). Does Measuring Intent Change Behavior? *Journal of Consumer Research* 20, 46–61.
- Ⓢ Morwitz, V. G. and D. Schmittlein (1992). Using Segmentation to Improve Sales Forecasts Based on Purchase Intent: Which ‘Intenders’ Actually Buy. *Journal of Marketing Research* 29, 391–405.
- Odell, P. and A. Feiveson (1966). A Numerical Procedure to Generate a Sample Covariance Matrix. *Journal of the American Statistical Association* 61, 198–203.
- Ⓢ Pickering, J. and B. Isherwood (1974). Purchase Probabilities and Consumer Buying Behavior. *Journal of Market Research Society* 16, 203–226.
- Robert, C. (1995). Convergence Control Methods for Markov Chain Monte Carlo Algorithms. *Statistical Science* 10, 231–253.
- Roberts, G. O. (1992). Convergence Diagnostics of the Gibbs Sampler. In J. Bernardo, J. Berger, A. Dawid, and A. Smith (Eds.), *Bayesian Statistics 4: Proceedings of the Fourth Valencia International Meeting*, pp. 775–782. Oxford University Press.
- Roberts, G. O. and A. F. M. Smith (1993). Bayesian Computation via the Gibbs Sampler and Related Markov chain Monte Carlo Methods. *Journal of the Royal Statistical Society, Series B* 55, 3–23.
- Stefanski, L. A. and R. J. Carroll (1985). Covariate Measurement Error in Logistic Regression. *Annals of Statistics* 13, 1335–1351.
- Tanner, M. (1993). *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions* (2nd ed.). New York, NY: Springer Verlag.
- Tauber, E. M. (1975). Predictive Validity in Consumer Research. *Journal of Advertising Research* 15, 59–64.
- Ⓢ Taylor, J., J. Houlahan, and A. Gabriel (1975). The Purchase Intention Question in New Product Development: A Field Test. *Journal of Marketing*, 90–92.

- Tierney, L. (1994). Markov Chains for Exploring Posterior Distributions (with Discussion). *Annals of Statistics* 4, 1701–1762.
- o Tobin, J. (1959). On the Predictive Value of Consumer Intentions and Attitudes. *Review of Economics and Statistics* 41, 1–11.
- Wakefield, J., A. Smith, A. Racine-Poon, and A. Gelfand (1994). Bayesian Analysis of Linear and Non-Linear Population Models by Using the Gibbs Sampler. *Applied Statistics* 43, 201–221.
- o Warshaw, P. R. (1980). Predicting Purchase and Other Behaviors from General and Contextually Specific Intentions. *Journal of Marketing Research* 17, 26–33.
- Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*. New York, NY: John Wiley and Sons.
- Zellner, A. and C.-K. Min (1995). Gibbs Sampler Convergence Criteria. *Journal of the American Statistical Association* 90, 921–927.

Table 1: Correspondence between purchase intention and purchase behavior observed in different markets and different studies.

| Study | Purchase Probability (%) | | Product |
|---|--------------------------|---------------|-----------------|
| | Intenders | Non-Intenders | |
| Juster 1966 ^a | 50 | 11 | Automobile |
| Jamieson and Bass 1989 ^b | 52.2 | 16.7 | Pump toothpaste |
| | 61.5 | 17.1 | Diet drink mix |
| | 43.5 | 15.6 | Fruit sticks |
| | 2.8 | 4.7 | Stay fresh milk |
| | 56.3 | 12.3 | Salad dressing |
| | 42.9 | 3.8 | Home computer |
| | 12.5 | 11.4 | Cordless phone |
| | 0.0 | 2.7 | Touch lamp |
| | 0.0 | 0.0 | Cordless iron |
| | 0.0 | 1.3 | Shower radio |
| Tauber 1975 ^c | 26 | 12 | Packaged goods |
| Infosino 1986 ^d | 34.5 | 9.7 | Service option |
| Pickering and Isherwood 1974 ^e | 37.5 | 7.6 | Automobile |

^aDerived from Table 3, based on six month probability scale. Intenders defined as those in the "Definite, Probable" intentions class, non-intenders, those in the "No" intentions class.

^bDerived from Table 1. Intenders defined as those who "Definitely/probably will buy", non-intenders, those who "Definitely/probably will not buy".

^cDerived from Table 1. Intenders defined as those who will "Positive purchase intent, solves need", non-intenders, those with "Negative purchase intent".

^dDerived from Figure 1. Intenders defined as those with likelihood rating 7-10, non-intenders, those with likelihood rating 1-6.

^ep. 205, derived from Theil and Kosobud, 1968.

Table 2: Results of simulation study of Bayes estimator with 200 replications.

| Block | n | β_0 β_1 p_{00} p_{11} | | | | Estimates of β_0 | | | | Estimates of β_1 | | | |
|-------|-----|---------------------------------------|----|----|----|------------------------|------|-------|------|------------------------|------|-------|------|
| | | | | | | Bayes | | Naive | | Bayes | | Naive | |
| | | | | | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | 200 | -2 | 3 | .8 | .8 | -2.51 | 0.81 | -0.46 | 0.35 | 3.93 | 1.22 | 0.63 | 0.54 |
| 2 | 200 | -2 | 3 | .9 | .6 | -2.10 | 0.34 | -0.85 | 0.12 | 3.19 | 0.52 | 0.57 | 0.14 |
| 3 | 200 | 2 | -3 | .9 | .6 | 2.15 | 0.82 | -0.09 | 0.10 | -3.27 | 1.21 | -0.41 | 0.09 |
| 4 | 200 | -3 | 4 | .9 | .6 | -2.88 | 0.86 | -0.94 | 0.23 | 3.87 | 1.16 | 0.63 | 0.43 |

Table 3: Results of simulation study of hierarchical Bayes (HB) estimator and disaggregate (DIS) estimator with 100 replications.

| Block | n | M | σ | RMSE, β_0 | | | | RMSE, β_1 | | | |
|-------|-----|-----|----------|-----------------|-------|-------|-------|-----------------|-------|-------|-------|
| | | | | HB | | DIS | | HB | | DIS | |
| | | | | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | 200 | 12 | 0.50 | 0.308 | 0.082 | 0.445 | 0.116 | 0.736 | 0.302 | 1.051 | 0.348 |
| 2 | 100 | 12 | 0.50 | 1.176 | 0.448 | 1.478 | 0.272 | 0.368 | 0.121 | 0.541 | 0.159 |
| 3 | 200 | 12 | 0.25 | 0.618 | 0.196 | 0.911 | 0.272 | 0.224 | 0.057 | 0.313 | 0.081 |
| 4 | 200 | 6 | 0.50 | 0.704 | 0.327 | 1.030 | 0.407 | 0.306 | 0.091 | .484 | .151 |

Table 4: Parameter estimates for Morwitz and Schmittlein (1992) PC dataset. ^a

| Analysis | p_{00} | p_{11} | USE | | INCOME | | PROFESS | | PACIFIC | | STUDENT | |
|----------------|----------|----------|-------|-------|--------|-------|---------|-------|---------|-------|---------|-------|
| | | | Mean | SD | Mean | SD | Mean | SD | Mean | SD | Mean | SD |
| 1 | 0.691 | 0.812 | 1.163 | 0.164 | 0.773 | 0.154 | 1.853 | 0.204 | 0.539 | 0.150 | 0.112 | 0.196 |
| 2 | 1.000 | 1.000 | 0.513 | 0.069 | 0.320 | 0.067 | 0.813 | 0.071 | 0.213 | 0.063 | 0.060 | 0.070 |
| 3 | 0.658 | 0.907 | 1.128 | 0.140 | 0.753 | 0.148 | 1.652 | 0.145 | 0.472 | 0.163 | 0.102 | 0.185 |
| 4 ^b | 1.000 | 1.000 | 1.235 | 0.076 | 0.792 | 0.072 | 1.546 | 0.078 | 0.627 | 0.069 | 0.052 | 0.078 |

^aNumbers in the table represent posterior means and standard deviations of respective regression coefficients.

^bProbit regression of *actual purchase responses*.

Table 5: Probit regression analysis of purchase intention versus purchase behavior and demographic covariates for the Morwitz and Schmittlein (1992) PC dataset.

| Coefficient | Variable | Estimate | Std Err | P-value |
|-------------|----------|----------|---------|---------|
| INTERCEPT | | 0.5212 | 0.131 | 0.0001 |
| PURCHASE | w | -1.3690 | 0.086 | 0.0001 |
| USE | x_1 | 0.0181 | 0.114 | 0.873 |
| INCOME | x_2 | -0.0314 | 0.113 | 0.781 |
| PROFESS | x_3 | -0.0491 | 0.112 | 0.661 |
| PACIFIC | x_4 | -0.0730 | 0.112 | 0.516 |
| STUDENT | x_5 | 0.0234 | 0.111 | 0.833 |