Measuring the Mere Measurement Effect in Non-Experimental Field Settings

Eric T. Anderson
Kellogg School of Management
Northwestern University
eric-anderson@kellogg.northwestern.edu
Phone: (847) 467-6482

Karsten Hansen
Kellogg School of Management
Northwestern University
karsten-hansen@kellogg.northwestern.edu
Phone: (847) 491-7098

Manish Tripathi
Kellogg School of Management
Northwestern University
m-tripathi@kellogg.northwestern.edu
Phone: (847) 491-7061
Fax: (847) 491-2498

[Working Paper]

September 25, 2007

Abstract

The mere measurement effect is a well-documented effect in laboratory studies and has recently been documented in several field surveys. In this paper, we identify two major obstacles to measuring the mere measurement effect in a non-experimental, field setting. These are: Who is asked to participate in a survey? Who agrees to respond to a survey? Firms often target surveys at particular consumers and this induces an effect we refer to as targeting bias. In addition, consumers who respond to a survey may not be representative of the overall population and this may create response bias. In this paper, we present a methodology that simultaneously controls for both targeting and response bias in a non-experimental, field setting. In developing our methodology, we identify four different ways to measure the mere measurement effect. We show how these metrics relate to the amount of information a researcher has about the surveyed customers.

We demonstrate our methodology on a dataset from a direct marketing firm that regularly surveys its customers. The results show that failure to control for targeting and response biases leads to overestimates of the impact of survey response on purchase frequency. Without controlling for these biases, our model predicts a positive mere measurement effect of 15%. After controlling for these biases, we find a negative mere measurement effect; purchase rates decrease by over 17% following completion of a survey. The model also yields estimates of the mere measurement effect for different customer segments. We illustrate this by showing how the mere measurement effect varies with customer tenure.

Key words: mere measurement, question-behavior effect, survey, response bias.

1 Introduction

Firms frequently survey customers to assess product and service offerings, customer satisfaction, and employee performance. It has been well-documented in laboratory settings that asking consumers questions, such as in a survey, can affect subsequent consumer behavior (Sprott et al. 2006). This phenomenon is referred to as the mere measurement effect or the question-behavior effect. Recent studies have shown that the mere measurement effect extends from the lab to field-based survey research (Morwitz, Johnson, and Schmittlein 1993; Dholakia and Morwitz 2002; Chandon, Morwitz, and Reinartz 2004; Chandon, Morwitz, and Reinartz 2005). These studies show that merely asking survey questions can have a substantial impact on subsequent purchase behavior.

The presence of the mere measurement effect in field settings raises the question of whether customer surveys both measure and change consumer behavior. Managers would like to carefully control for the mere measurement effect when they relate survey metrics to subsequent consumer behavior. Academics would like to explore archival survey and purchase data to gain a deeper understanding of the mere measurement effect. Thus, for both managers and academics there is a need for a methodology that enables the analyst to measure the mere measurement effect.

Capturing the mere measurement effect in a field setting poses several challenges. Unlike a laboratory, the researcher must deal with problems that arise in a practical setting. We identify two major obstacles to measuring the mere measurement effect in a non-experimental, field setting. These are: Who is asked to participate in a survey? Who agrees to respond to a survey? Rather than asking questions to a random sample of customers, firms often target surveys at particular consumers (i.e., targeting bias). For example, car manufacturers often survey recent buyers to measure customer satisfaction with their new automobile and the local dealer. When measuring the mere measurement effect, we show that this induces the potential for targeting bias. Survey participation may also be non-random and this may lead to response bias.

The core contribution of this paper is that we develop a methodology that simultaneously controls for both targeting and response bias using non-experimental or observational data, which allows the researcher to quantify the mere measurement effect. A further contribution is that we identify four different metrics for the mere measurement effect. For example, a researcher may be interested in the mere measurement effect on new car buyers and the mere measurement effect on all car buyers. We show that a firm's ability to measure these effects depends crucially on knowing both how and why customers were targeted with a survey (i.e., the targeting policy). Our model also shows that the joint presence of targeting bias and response bias complicates the identification of the mere measurement effect(s). The minimal amount of information required about a firm's targeting policy increases when both response and targeting biases are present.

We apply our methodology to data provided by a direct mail, catalog company that sells products through catalogs and an Internet site. The firm regularly surveys customers and our dataset contains twenty nine different surveys over a four year period. The results demonstrate the importance of controlling for targeting and response bias. Without controlling for these biases, our model predicts that survey response leads to a 15% increase in purchase frequency and 7% increase in customer order size. After controlling for these biases, the model predicts that a survey will decrease purchase frequency by 17% and decrease dollar spending per customer order by 5%.

Our methodology also enables researchers to analyze the mere measurement effect in different customer segments. To illustrate this point, we show how mere measurement varies with customer tenure (i.e., length of time since first purchase). This investigation is motivated by previous research that suggests customer experience may moderate the mere measurement effect (Janiszewski and Chandon 2007). We find a large negative mere measurement effect among customers who have shorter tenure with the firm, but a small negative mere measurement effect among older customers. The explanation for this phenomena is still an open question, although response fluency may help explain the varied effect (Janiszewski and Chandon 2007).

Another contribution of our model and data is that we are able to examine whether surveying customers affects total expenditure with a firm. Our model simultaneously accounts for both purchase frequency and spending per purchase. Together, these allow us to examine whether the mere measurement effect influences firm revenue. In our application, we find that the mere measurement effect decreases purchase frequency but increases spending per purchase. The overall impact on spending with the firm is negative.

1.1 Literature Review

Sherman (1980) first demonstrated the mere measurement effect in a laboratory setting and introduced the concept of "the self-erasing error of prediction" to explain this behavior. Another prevalent explanation for the mere measurement effect is the self-generated validity theory (Feldman and J.G. Lynch 1988). The self-generated validity theory argues that asking a question about a behavior increases the accessibility in memory of pre-existing attitudes and intentions. Thus, measurement does not change preferences, it just highlights pre-existing preferences. Morwitz and

Fitzsimons (2004) use a series of laboratory experiments to disentangle these two theories. Their evidence supports the self-generated validity theory as the mechanism behind the mere measurement effect rather than “the self-erasing error of prediction”. Recent research (Janiszewski and Chandon 2007) has suggested that process redundancy creates a fluency which may be an additional source of the mere measurement effect. Response fluency has been posited as an explanation for empirical evidence of experience as a moderator of the mere measurement effect (Morwitz, Johnson, and Schmittlein 1993). More generally, other researchers in marketing and psychology have studied the underlying consumer processes behind the mere measurement effect; Sprott et al. (2006) offer a thorough summary of this research.

Morwitz, Johnson, and Schmittlein (1993) is the first paper to empirically quantify the mere measurement effect in a field setting. The authors show that measuring intent for automobile purchases increases the likelihood of a future automobile purchase. Because automobiles are expensive products, the results demonstrate that mere measurement can influence consumer decisions of significant dollar value. Fitzsimons and Morwitz (1996) provide an extension of Morwitz, Johnson, and Schmittlein (1993) from the category to the brand level. Their study shows that the relationship between repeat purchase rate and brand choice is positive, and greater for those households that are surveyed versus those households that are not surveyed.

In the data used by Morwitz, Johnson, and Schmittlein (1993) and Fitzsimons and Morwitz (1996) the researcher does not know the specific purchase date, which makes it impossible to measure durational effects. Dholakia and Morwitz (2002) analyze data from a financial services firm that includes all consumer activity over a one year period. This additional information enables them to examine the scope and persistence of the mere measurement effect. Retail customer households enrolled in a client management program were randomly assigned to either an experimental or control group. The experimental group was surveyed over the phone regarding their satisfaction with specific features of the program and the control group was not surveyed. There are two important findings from the study. First, the effects from satisfaction measurement are persistent; that is they are increasing for several months after the measurement, before starting to decline. Second, the likelihood of future purchase/relational behaviors will be higher (lower) when satisfied (dissatisfied) customers respond to a satisfaction survey then when they do not.

Chandon, Morwitz, and Reinartz (2004) extend the empirical mere measurement literature to non-durable goods with repeated purchases. The researchers use purchase data from an online French grocer combined with phone survey data regarding purchase intentions to investigate the mere measurement effect over time. A key finding of the study is that measuring intentions increases the likelihood of repeat purchase incidence and shortens the time until first repeat purchase. The researchers also find that the mere measurement effect decays rapidly after the first three months of the survey.

1.2 Difficulties with Quantifying Mere Measurement

To illustrate the challenges of quantifying the mere measurement effect it is helpful to contrast laboratory settings with non-experimental, field settings. When a survey is administered in the field, all customers can be grouped into three categories: customers who respond (R), customers who are asked but do not respond (NR), and customers who are not asked to participate in a survey (NA). In a laboratory setting, random assignment to test and control groups eliminates targeting bias and if all subjects who participate are exposed to the treatment there is no response bias.

This implies there are no subjects in the NR group and a comparison of behavior in the R and NA group is a valid metric of the mere measurement effect. In a field setting, there are typically three groups of subjects (NA, NR, R) and if there is targeting and response bias then the three groups of customers cannot be directly compared.

Morwitz, Johnson, and Schmittlein (1993) note the possibility that the groups of customers they consider may not be equivalent. In their paper, they compare responders (R) to customers who are not asked (NA). The authors control for observable differences between customer groups by weighing the data based on demographic variables. An important limitation of their methodology is that there may be unobserved factors that explain why customers enter and exit the panel. If these unobserved factors are correlated with purchase behavior, then it is no longer possible to recover the mere measurement effect.

Both Dholakia and Morwitz (2002) and Chandon, Morwitz, and Reinartz (2004) use field experiments, which overcome some of the limitations of Morwitz, Johnson, and Schmittlein (1993). These researchers compare customers who are randomly assigned to a control condition (NA) with customers who respond to a survey (R). Since responders choose to answer a survey, this raises the potential for response bias. Dholakia and Morwitz recognize this limitation and show that NA and R groups are observationally equivalent, except for the fact that the R group has more accounts with the firm prior to the study. The authors control for this difference in their analysis. Chandon, Morwitz, and Reinartz considers new buyers and therefore there are no historical covariates that can be used to compare the NA and R groups. Similar to Morwitz, Johnson, and Schmittlein (1993), a limitation of both of these studies is whether there are unobserved factors that determine response to the survey and are correlated with purchase behavior.

A contribution of our paper is that we derive the specific conditions required for researchers to make valid inferences from non-experimental, field data[1]. Under specific assumptions, observed covariates, such as demographics, can be used to correct for targeting and response bias. However, in general this approach is not sufficient. We document the potential pitfalls in measuring mere measurement and propose a solution.

1.3 Structure of the Paper

The remainder of this paper is organized as follows. In ∮ 2 we examine the biases associated with the calculation of the mere measurement effect. In ∮ 3 we identify mechanisms for correcting these biases. In ∮ 4 we present a formal model for identifying the mere measurement effect in a field-based study, while correcting for biases. We then introduce our data set in ∮ 5. The findings from the model estimation are presented in ∮ 6 and the paper concludes in ∮ 7 with a review of the findings, limitations and opportunities for future work.

2 Identification of the Mere Measurement Effect

We use an example to illustrate potential biases associated with quantifying the mere measurement effect from non-experimental data. Assume the researcher observes survey response and transaction data for a group of customers at a single

[1]We focus the paper on non-experimental data, since this is typically what is available to managers and researchers. However, as pointed out above, field experiments can be used only to control for targeting bias, not for response bias. Since survey response is a customer decision, it is in general not possible to randomize survey responses across customers.

point in time. For example, in a sample of 10,000 customers suppose that the 1,000 customers with incomes over one million dollars are targeted with a survey. Surveying all 1,000 customers is too expensive, so the firm randomly calls 20% of these customers. A total of 200 high income customers are asked to participate and 50 agree to complete the survey. Following the survey, we learn that 2,000 of the 10,000 customers purchased an item. Of these 2,000 customers that purchase an item, 250 are high income customers. Furthermore, of these 250 high income customer that purchase an item, 15 also complete the survey. We want to measure whether survey completion affected the purchase decision of customers.

Let $D_i = 1$ if customer i purchases and is 0 otherwise. From the example above, $D_i = 1$ for 250 high income customers. Let $T_i = 1$ if customer i was asked to participate in the survey and is 0 otherwise. Again, from the example above, $T_i = 1$ for 200 high income customers, and $T_i = 0$ for 800 high income customers[2]. Consumers may not respond to a survey either because they were not asked (NA) or they declined to participate (NR). To formalize this distinction, let $R_i = 1$ when customer i responds to a survey and $R_i = 0$ if a customer declines to participate. Note that R is only defined when T=1. Let $\tilde{R}_i = 1$ if $(T,R) = (1,1)$ and $\tilde{R}_i = 0$ if $(T,R) = (1,0)$ or $T_i = 0$. In the context of the example above, $\tilde{R}_i = 1$ for 50 customers, and $\tilde{R}_i = 0$ for 150 + 800 = 950 high income customers. Now, $\tilde{R}_i = 0$ equals observed non-response and $(T,R) = (1,0)$ equals "true" non-response.

Now consider the purchase decision of consumer i, who has net utility:

$$U_i = \mu + \lambda \tilde{R}_i + \epsilon_i, \tag{1}$$

[2] $T_i = 0$ for 9,000 low income customers as well.

so that $D_i = 1$ if $U_i > 0$. If $\lambda > 0$, the mere measurement effect is present and survey response increases the likelihood of purchase. The purchase probabilities for responders and non-responders are:

$$Pr(D_i = 1|\tilde{R}_i = 1) = \int_{-(\mu+\lambda)}^{\infty} p(\epsilon_i|\tilde{R}_i = 1)d\epsilon_i \tag{2}$$

$$Pr(D_i = 1|\tilde{R}_i = 0) = \int_{-\mu}^{\infty} p(\epsilon_i|\tilde{R}_i = 0)d\epsilon_i \tag{3}$$

A comparison of these probabilities does not typically yield the effect of interest. Using the survey example above, the purchase probability of responders (2) is 15/50 or 30%, and the purchase probability of non-responders (3) is 235/950 or 24.7%. In order for the difference in (2) and (3) to equal the true mere measurement effect, one must also assume that:

$$p(\epsilon_i|\tilde{R}_i = 1) = p(\epsilon_i|\tilde{R}_i = 0). \tag{4}$$

In other words, the conditional distribution of the error terms for survey responders equals that of non-responders. A sufficient condition for this is that survey response, $\tilde{R}_i$, is independent of all unobserved customer effects, ϵ_i. In general, condition (4) is very unlikely to be satisfied. For example, if survey response is correlated with ϵ_i then this results in response bias. Targeting surveys can also lead to violations of (4) and this is the focus of the next section. Later, we derive a methodology for controlling for both types of bias (see section 3). An important contribution of this discussion is that we highlight how targeting and response bias interact in a non-trivial fashion (see section 3.3).

In the following sub-sections of section, we first define and provide examples of targeting policies (section 2.1). We then introduce four metrics of the mere measurement effect (section 2.2). Next, we discuss how the type of information available to the researcher moderates the subset of the types of mere measurement effects that can be quantified (section 2.3). Finally, we conclude the section with a discussion of response bias (section 2.4).

2.1 Targeting Policies

We now illustrate how targeted surveys can lead to biased estimates of the mere measurement effect. Assume that the targeting policy is

$$\Pr(T_i = 1 | Z_i = z_i) = \tau(z_i), \tag{5}$$

where z_i is a vector of customer-specific variables observable by the company. For example, a car manufacturer may target recent automobile buyers, in which case z_i would be a measure of recency of a car purchase. Some examples of targeting policies are

$$\tau(z_i) = \begin{cases} \alpha & \text{if } z_i \in \mathcal{Z},\ \ 0 < \alpha \leq 1, \\ 0 & \text{if } z_i \notin \mathcal{Z}, \end{cases} \tag{6}$$

$$\tau(z_i) = \begin{cases} \alpha_1 & \text{if } z_i \in \mathcal{Z}_1,\ \ 0 < \alpha_1 \leq 1, \\ \alpha_2 & \text{if } z_i \in \mathcal{Z}_2,\ \ 0 < \alpha_2 \leq 1, \\ 0 & \text{if } z_i \notin \mathcal{Z}_1 \cup \mathcal{Z}_2, \end{cases} \tag{7}$$

and

$$\tau(z_i) = \alpha, \qquad \forall z_i. \tag{8}$$

For the targeting policy in (6), customers with observables in the set $\mathcal{Z}$ are targeted with probability α, and the remaining customers are not targeted. A special case of this is $\alpha = 1$, implying that all customers with z_i in the set $\mathcal{Z}$ are targeted. The targeting policy in (7) is a more general stratified targeting policy, where customers in the set $\mathcal{Z}_1$ are targeted with probability α_1, while customers in the set $\mathcal{Z}_2$ are targeted with probability α_2. Finally, in (8) targeting is done completely at random and is independent of z.

We assume that targeting in itself does not cause a change in purchase probability, and thus,

$$\Pr(D = 1|Z = z, T = 1, R = r) = \Pr(D = 1|Z = z, R = r), \qquad r = 1, 0. \tag{9}$$

Survey response is the only factor that can cause the purchase probability to change for a group of customers who are homogeneous with respect to Z. This also implies

$$\begin{aligned} \Pr(D = 1|Z = z, T = 0) &= \Pr(D = 1|Z = z, T = 1, R = 0) \\ &= \Pr(D = 1|Z = z, R = 0). \end{aligned} \tag{10}$$

In other words, if a group of customers share the same observable characteristics, z, and are observed non-responders, $\tilde{R} = 0$, then they have the same purchase probability. In this section, we assume there is no response bias and that R is independent of Z:

$$\Pr(R = 1|Z = z) = \Pr(R = 1). \tag{11}$$

In our later derivation of the full model, we will relax (11).

2.2 Metrics

We now show that there are at least four different ways of characterizing the mere measurement effect. First, one might be interested in the mere measurement effect within specific population groups, such as recent buyers. We define

$$\Delta_z \equiv \Pr(D = 1|Z = z, R = 1) - \Pr(D = 1|Z = z, R = 0) \tag{12}$$

as the mere measurement effect among customers who share the same observable covariates, z. This is the effect measured in Chandon, Morwitz, and Reinartz (2004) who analyze new customers from a supermarket. One might also be interested in the mere measurement effect on the entire population. We define this as

$$\bar{\Delta} = \int \Delta_z p(z) dz. \tag{13}$$

This captures the change in a firm's total market share when everyone in the population responds to a survey. For example, Chandon, Morwitz, and Reinartz (2004) show that the mere measurement effect exists among new consumers but do not consider the effect among all supermarket buyers. When a firm targets surveys, one might be interested in the mere measurement effect within a targeted group of customers. We define this as

$$\bar{\Delta}_{T=1} = \int \Delta_z p(z|T=1)dz. \tag{14}$$

This is the average effect of responding to a survey among the population of consumers who were targeted to receive the survey, which is what is typically measured in a laboratory study. If the firm uses a random targeting policy, we can define the following mere measurement effect:

$$\Delta_k = \int \Delta_z p(z|\tau(Z)=k)dz. \tag{15}$$

Δ_k is the probabilistic version of $\bar{\Delta}_{T=1}$ and is relevant when a firm uses probabilistic targeting, where $\Delta_{100} \equiv \bar{\Delta}_{T=1}$. More formally, Δ_k is the effect for the population of customers with z in the set $\{z : \tau(z) = k\}$. This population has z values implying a targeting probability of size k. While the expression for Δ_k may look awkward, it is common for firms to utilize random targeting. Continuing our previous example, a firm may ask 20% of its high income customers to participate in a survey. Thus, ex-ante each of the 1,000 high income customers has a 20% chance of being asked to participate in a survey. Δ_{20} is the average change in the purchase probability among the 1,000 customers with high income who have a k=20% chance of being targeted for a survey.

2.3 Information Sets & Metrics

In this section, we illustrate how limited information about a firm's targeting policy may lead to biased estimates of the four mere measurement effects. Suppose a researcher compared purchase rates for customers who take the survey to purchase rates for customers who do not take the survey. This is a comparison of purchase rates for customers with $\tilde{R} = 1$ and $\tilde{R} = 0$. Neither of these two populations respond to the survey, but one does it because of choice and the other because it is not given the option of responding. A comparison of responders and non-responders yields the following effect:

$$\tilde{\Delta} = \Pr(D = 1|\tilde{R} = 1) - \Pr(D = 1|\tilde{R} = 0). \tag{16}$$

We can express $\tilde{\Delta}$ as [3]

$$\tilde{\Delta} = \bar{\Delta}_{T=1} + (1-\xi) \int \Pr(D = 1|Z = z, R = 0)\big[p(z|T = 1) - p(z|T = 0)\big]dz \tag{17}$$

Where ξ is the fraction of the targeted population. In general, $\tilde{\Delta}$ will not correspond to any meaningful causal effect. The size of the bias is proportional to the size of the non-targeted population and the degree of dependence between T and z. The problem is that in computing $\tilde{\Delta}$ we are comparing purchase rates for two populations which are different not only due to the survey response, but also due to having different Z's.

There are two special cases under which there will be no targeting bias. The first is when targeting is based on random assignment, which is what happens in a lab setting. The second case is when the variables that impact purchase behavior are distinct from the variables used for targeting. A detailed explanation of the latter case is presented in the Technical Appendix.

[3] Detailed explanation in Appendix A

In general, the group comparison estimator $\tilde{\Delta}$ is biased when a firm targets surveys. A researcher's ability to correct for this bias depends on the available information. When the researcher has full information on who is targeted and the customer's z variables, then all four mere measurement effects can be estimated.

In practice, a researcher may not have full information and we consider five scenarios of limited information:

I. T **observed.** In this case the researcher observes who is targeted but does not observe the z-variables.

II. $(\tau(\mathbf{z}), \mathbf{z})$ **known.** In this case the researcher observes the targeting policy and the z-variables, but does not observe who is targeted.

III. z **known.** In this case the researcher observes only the z variables, and knows that these are the only variables used for targeting, but does not observe the actual targeting policy.

IV. $z_1 \subset z$ **known.** In this case the researcher observes only a subset of z variables.

V. **Nothing is known**. This is the "minimum information" case.

In each of the first three scenarios, a subset of the four mere measurement effects can be recovered. In scenarios 4 and 5, none of the mere measurement effects are identified. These results are summarized in Table 1 and detailed derivations are presented in the Technical Appendix.

In scenario I, the researcher knows which customers were targeted but does not know the z variables used to target. Continuing with our previous example, a researcher may be told which of the 200 customers were contacted for the survey.

In addition, the researcher learns the 50 survey responders (R) and the 150 non-responders (NR). But, the researcher is not told how the 200 customers were selected from the overall population. In this case, the researcher can compute

$$\bar{\Delta}_{T=1} = \Pr(D = 1|T = 1, R = 1) - \Pr(D = 1|T = 1, R = 0), \tag{18}$$

which is an unbiased estimate of the effect of survey response on the targeted population. But, the mean treatment effect in the population, $\bar{\Delta}$, cannot be calculated without knowledge of the Z's.

[INSERT TABLE 1 HERE]

In scenario II, the researcher knows that 1,000 high income customers were targeted for the survey, 20% of these customers were asked to participate, and 50 completed a survey. But, the researcher does not know the 150 customers who were non-responders. This may occur in phone survey research when the market research firm does not keep accurate records of which customers were called. In this case, the researcher cannot compute $\bar{\Delta}_{T=1}$ since one cannot disentangle the event $T = 1, R = 0$ from the event $T = 0$. However, we can compare purchase rates for a group of customers who have a certain *probability* of being targeted and respond, to the purchases rate for a group of customers with the same probability of being targeted and who either were targeted and do not respond, or were not targeted. This is

$$\tilde{\Delta}_k = \Pr(D = 1|\tau(Z) = k, \tilde{R} = 1) - \Pr(D = 1|\tau(Z) = k, \tilde{R} = 0). \tag{19}$$

In the Technical Appendix, we prove that $\tilde{\Delta}_k = \bar{\Delta}_k$. Because the z variables are observed the researcher can also compute the population effect, $\bar{\Delta}$.

In scenario III, the researcher once again observes z variables such as income, age, and other demographics. However, while the researcher observes the responders to the survey, the targeting policy, $\tau(z)$ is unknown to the researcher. Continuing our example, the researcher knows that 50 customers responded to the survey; but the researcher does not know how customers were targeted. What is crucial, however, is that the researcher has the full set of z variables used to target. Because the researcher knows all the variables that are used to target, it is possible to recover $\bar{\Delta}$ and $\bar{\Delta}_z$. Even without knowing the targeting policy, two of the mere measurement effects can be recovered. But, as we discuss later, this result will no longer be true when there is response bias.

Scenarios IV and V illustrate the importance of knowing all the targeting variables, not just a subset of variables. When the researcher knows neither the targeting rules nor the targeting variables, it is clear that none of the mere measurement effects are identified (scenario V) . But, suppose a researcher knows a subset of the targeting variables (scenario IV). In this case, one is still unable to identify any of the mere measurement effects. Full information about the targeting variables is required to accurately estimate any mere measurement effect.

A comparison of the first two scenarios highlights the need to understand how customers were targeted (scenario II) rather than simply who is targeted (scenario I). If the researcher is interested in the effect of survey response on the targeted consumers, then knowing who is targeted is sufficient. However, inferences about the broader population of consumers can only be recovered when the researcher observes how customers were targeted. Thus, in some applications, how targeting is done may be more valuable than who is targeted.

2.4 Response Bias

In this section we consider the effects of response bias. Response bias occurs due to the voluntary nature of surveys: Individuals are free to choose to respond or not, and if responders are systematically different than non-responders, response bias will result. In the social sciences this phenomenon is usually called self-selection (Heckman (1978)).

For ease of exposition in this section, assume that there is no targeting bias. In this case, $R = \tilde{R}$ and the net utility of purchase is given by (1). We wish to identify the causal effect of survey response, R, which is

$$\Delta = \int_{-(\lambda+\mu)} p(\varepsilon)d\varepsilon - \int_{-\mu} p(\varepsilon)d\varepsilon, \tag{20}$$

where $p(\varepsilon)$ is the density of ε. This is the incremental effect of R on the purchase probability if we could run the ideal experiment involving randomization of R across customers. This is also the effect we would recover if the decision to respond is independent of unobservable heterogeneity ε.

The key difference between targeting bias and response bias is that response is a customer decision, which may involve factors/variables that are known only to the customer – not the firm/researcher. Assume that response is determined by

$$\Pr(R = 1|Z = z, \psi) = G(z, \psi), \tag{21}$$

where z are observables and ψ is unobservable to the firm and researcher. For ease of notation and exposition, assume that Z is independent of ε and ψ. If we compare purchase probabilities for responders and non-responders we obtain

$$\tilde{\Delta} = \int_{-(\lambda+\mu)} p(\varepsilon|R = 1)d\varepsilon - \int_{-\mu} p(\varepsilon|R = 0)d\varepsilon. \tag{22}$$

If ψ and ε are correlated then

$$p(\varepsilon|R = 1) \neq p(\varepsilon|R = 0), \tag{23}$$

and in this case we are computing purchase probabilities for two non-comparable populations, and $\tilde{\Delta}$ will not equal Δ. So unless we make strong assumptions, e.g., that unobservable factors in response decisions are uncorrelated with unobservable factors in purchase decisions, response bias will be present. When response bias is present, we will not uncover the correct mere measurement effect in any of the five information scenarios above.

2.5 Summary

If a researcher has full information about who is targeted, how they were targeted, and there is no response bias then four different types of mere measurement effects are identified. Under limited information about the targeting policy and the targeting variables, we have shown that a subset of the four mere measurement effects may be identified. In that analysis, we assumed no response bias. When response bias exists, then ignoring this effect will bias all of the mere measurement effects. In the next section, we show how to simultaneously control for both targeting and response bias.

3 Correcting for Targeting and Response Biases

In the section above we described two key problems in identifying the mere measurement effect in non-experimental settings. These problems arose due to survey targeting by the firm and non-random response by the customer. In this section we propose a methodology that allows a researcher to obtain estimates of the mere measurement effect on customers' purchase frequency, transaction size and total firm revenue. This methodology will provide correct estimates of the mere measurement effect in the first two targeting information scenarios considered in section 2.3. These estimates will be valid regardless of whether or not the firm targets or randomizes surveys, and whether or not customer survey responses are random.

Before we present the full model in the next section, we first outline the main idea in our approach. To do this we briefly revisit the simple model (1) in section 2. There we argued that the model would produce an incorrect estimate of the mere measurement effect whenever

$$p(\epsilon_i|\tilde{R}_i = 1) \neq p(\epsilon_i|\tilde{R}_i = 0). \tag{24}$$

This was due to the model incorrectly specifying that ϵ_i was independent of $\tilde{R}_i$, which will not be the case when targeting and response bias are present. What we need then is a way to "build in" an automatic correction in the model. In other words, we need to account for the different ϵ distributions for the $\tilde{R} = 1$ and $\tilde{R} = 0$ populations. Doing this clearly requires either additional data, additional assumptions or both.

To illustrate our approach, we start by decomposing ϵ_i as

$$\epsilon_i = \beta_i + \varepsilon_i, \tag{25}$$

where β_i is a time-invariant customer specific parameter ("unobservable heterogeneity") and ε_i is a standard error term assumed independent of $\tilde{R}_i$. β_i captures the customer's basic preference for the company's products. We can then express the distributions in (24) as

$$p(\epsilon_i|\tilde{R}_i = j) = \int p_\varepsilon(\epsilon_i - \beta_i) p(\beta_i|\tilde{R}_i = j) d\beta_i, \qquad j = 0, 1, \tag{26}$$

where p_ε is the density of ε_i. If we can find a way to allow the β_i distributions to differ for the $\tilde{R} = 1$ and $\tilde{R} = 0$ populations, we are indirectly correcting the ϵ_i distributions. The assumption implicitly made in the model in (1) is

$$p(\beta_i|\tilde{R}_i = 0) = p(\beta_i|\tilde{R}_i = 1) = p(\beta_i). \tag{27}$$

In this case the probabilities (or likelihood) used to infer the mere measurement effect is

$$\begin{aligned} \Pr(D_i = 1|\tilde{R}_i = 1) &= \int \Pr(D_i = 1|\tilde{R}_i = 1, \beta_i) p(\beta_i) d\beta_i, \\ \Pr(D_i = 1|\tilde{R}_i = 0) &= \int \Pr(D_i = 1|\tilde{R}_i = 0, \beta_i) p(\beta_i) d\beta_i. \end{aligned} \tag{28}$$

This fails to account for the fact that responders may be systematically different than non-responders in terms of their β_i parameter. For example, this will happen if customers with large (or small) β_i parameters are more likely to respond to the survey when targeted, and/or if customers with large (or small) β_i parameters are more likely to be targeted[4]. In any case, this will lead to incorrect inferences about

[4] Note that while targeting cannot directly be related to β_i (since β_i is unobservable), it can be indirectly related to β_i if targeting is done on observable variables that are correlated with β_i.

the mere measurement effect.

We need a mechanism that allows us to construct different β_i distributions for responders and non-responders. To do this we will assume that we have access to *the entire purchase history for each customer*. We can then use this history to infer a β_i parameter for each customer. This automatically induces different β_i distributions for responders and non-responders. In terms of the targeting rule we make the following assumption:

Assumption 3.1. *Let $\mathcal{H}_i$ be customer i's purchase history at the time of targeting, and let Z_i^* be non-purchase related customer specific variables (e.g., customer demographics). Assume that conditional on $Z_i = (\mathcal{H}_i, Z_i^*)$, T_i is independent of β_i.*

This is an assumption about the joint distribution (T_i, Z_i, β_i). Note that Z_i contains all variables that possibly can be used for targeting. The assumption states that conditional on these, β_i is independent of T_i. This assumption rules out the possibility that targeting can be done directly on the basis of β_i. Note carefully what we are assuming here: We are *not* assuming that targeting is independent of unobservable heterogeneity (β_i above). What we are making is a conditional independence assumption. We are assuming that *conditional* on past behavior (and demographics), T_i and β_i are independent. For example, in the context of the earlier example, high income customers were targeted and Z_i only includes income. In the total population of 10,000 consumers, T_i and β_i are likely to be dependent. For example, targeted consumers (i.e., high income consumers) may buy more frequently and spend more. But this is not our assumption. We are assuming that conditional on being a high income buyer, T_i and β_i are independent. In the example, targeting of high income buyers is random and therefore our assumption is valid.

An implication of Assumption 3.1 is

$$\Pr(T_i = 1|Z_i, \beta_i) = \Pr(T_i = 1|Z_i). \tag{29}$$

This turns out to be a convenient result in the case when targeting is observed (scenario I): Conditional on Z the event $T = 1$ does not contain any information about β_i (or any other model parameter), so we can simply condition on T throughout the analysis. In scenario I, Assumption 3.1 is not very restrictive since in this case it amounts to assuming that targeting is not done directly on β_i.

To see how our proposed framework works, consider the simplest case where each customer is observed for two time periods (one pre- and one post-survey) and some customers are targeted for a survey in the second period. Assume that utility in the two time periods is

$$\begin{aligned} U_{i,0} &= \beta_i + \varepsilon_{i,0} \\ U_{i,1} &= \beta_i + \lambda \tilde{R}_i + \varepsilon_{i,1}, \end{aligned} \tag{30}$$

We consider the first three information scenarios from section 2.3. Note that the only purchase behavior variable available for targeting is $D_{i,0}$. Let Z_i^* be other observables used for targeting, e.g., costumer demographics. In the notation of section 2.1 we then have $Z_i = (D_{i,0}, Z_i^*)$.

3.1 Scenario I

In this case we observe who is targeted, but we do not necessarily observe the targeting variables. The likelihood contribution for a targeted customer is then

$$\begin{aligned}\Pr(D_{i1} = d_{i1}, D_{i0} = & d_{i0}, R_i = j | T_i = 1, Z_i^*) = \int \Pr(D_{i1} = d_{i1} | \beta_i, R_i = j) \times \\ & \Pr(R_i = j | T_i = 1, D_{i0} = d_{i0}, \beta_i) \times \\ & \Pr(D_{i0} = d_{i0} | \beta_i) \times \\ & p(\beta_i) d\beta_i,\end{aligned} \tag{31}$$

where for simplicity we assume consumer demographics Z_i^* is independent of β_i[5].

To simplify this expression, assume for the moment that conditional on being targeted, response is random, i.e., R_i is independent of β_i. This simplifies the likelihood to

$$\begin{aligned}\Pr(D_{i1} = d_{i1}, D_{i0} = & d_{i0}, R_i = j | T_i = 1, Z_i^*) = k \int \Pr(D_{i1} = d_{i1} | \beta_i, R_i = j) \times \\ & \Pr(D_{i0} = d_{i0} | \beta_i) \times \\ & p(\beta_i) d\beta_i\end{aligned} \tag{32}$$

Inferences from this likelihood do not suffer from the same problem as the one in (28). This follows from the fact that the β_i distribution that we are integrating over in (32) depends on d_{i0}. This makes all the difference. For example, suppose the targeting rule is that everyone who purchased in the first time period is targeted:

$$T_i = 1 \iff D_{i0} = 1. \tag{33}$$

[5] We can easily relax this assumption and let the β_i distribution depend on Z_i^*.

Targeted customers all make a purchase in the first period, which means that they on average have large β_i's. If we don't correct for this, we will obtain invalid estimates. However, the likelihood already incorporates this: Note that the β_i distribution for targeted customers is

$$p(\beta_i|T_i = 1) = p(\beta_i|D_{i0} = 1) = \frac{\Pr(D_{i0} = 1|\beta_i)p(\beta_i)}{\Pr(D_{i0} = 1)}. \tag{34}$$

But setting $d_{i0} = 1$ we see that this is the exact distribution we are integrating over in (32). This generalizes easily to more elaborate targeting rules and longer purchase histories. As long as targeting is done on past consumer purchase behavior and we are careful to model the dependence between this past behavior and the unobservable heterogeneity parameters in the model, we will obtain the correct estimates of the mere measurement effect. As an aside, note that assumption 3.1 and the assumption that targeting is observed, the likelihood and resulting inferences are robust to the specific targeting policy used. In other words, as long as we observe who is targeted, the targeting rule drops out of the likelihood function[6]. The researcher does not need to know the specific targeting policy used by the firm. However, we show below in scenario II how our approach also can accommodate the case where targeting is unobserved.

When survey response is non-random the likelihood for targeted customers be-

[6]In econometrics jargon, T_i is weakly exogenous.

comes

$$
\begin{aligned}
\Pr(D_{i1} = d_{i1}, D_{i0} = & d_{i0}, R_i = j | T_i = 1, Z_i^*) = \int \Pr(D_{i1} = d_{i1} | \beta_i, R_i = j) \times \\
& \Pr(R_i = j | T_i = 1, D_{i0} = d_{i0}, \beta_i) \times \\
& \Pr(D_{i0} = d_{i0} | \beta_i) \times \\
& p(\beta_i) d\beta_i.
\end{aligned} \tag{35}
$$

To capture the dependence between response and β_i we formulate an auxiliary response model that formally models this dependence. This response model then accounts for the term involving R_i in the likelihood. This corrects the β_i distribution to take response bias into account. Of course, our inferences about the mere measurement effect will in this case hinge on the correct specification of the response model and the estimates will potentially be sensitive to this specification.

Equation 35 is the likelihood contribution for a targeted customer. The likelihood for a non-targeted customer is

$$
\begin{aligned}
\Pr(D_{i1} = d_{i1}, D_{i0} = & d_{i0} | T_i = 0, Z_i^*) = \int \Pr(D_{i1} = d_{i1} | \beta_i, R_i = 0) \times \\
& \Pr(D_{i0} = d_{i0} | \beta_i) \times \\
& p(\beta_i) d\beta_i,
\end{aligned} \tag{36}
$$

where for notational reasons "$R_i = 0$" simply means that R_i does not enter the probability for D_{i1} (recall that the event $R_i = 0$ strictly speaking only is defined when $T_i = 1$). As above this likelihood correctly captures the fact that non-targeted customers have a different β_i distribution than targeted customers. For example, if the targeting rule is (33), non-targeted customers will on average have smaller β_i's.

3.2 Scenario II

When targeting is not observed, but the targeting rule and targeting variables are observed, the likelihood is more complicated. First, if we observe $R_i = 1$ we know that the customer was targeted, and in this case the likelihood is identical to the one above. However, when we observe a non-response we no longer know whether this was due to non-targeting or to targeting combined with true non-response. We can no longer condition on T, and this turns the likelihood into a mixture:

$$
\begin{aligned}
\Pr(D_{i1} = d_{i1}, D_{i0} =& d_{i0}, \tilde{R}_i = 0 | Z_i^*) = \\
& \Pr(T_i = 1 | D_{i0} = d_{i0}, Z_i^*) \times \int \Pr(D_{i1} = d_{i1} | \beta_i, R_i = 0) \times \\
& \Pr(R_i = 0 | T_i = 1, D_{i0} = d_{i0}, \beta_i) \times \\
& \Pr(D_{i0} = d_{i0} | \beta_i) \times \\
& p(\beta_i) d\beta_i + \\
& \Pr(T_i = 0 | D_{i0} = d_{i0}, Z_i^*) \times \int \Pr(D_{i1} = d_{i1} | \beta_i, R_i = 0) \times \\
& \Pr(D_{i0} = d_{i0} | \beta_i) \times \\
& p(\beta_i) d\beta_i
\end{aligned}
\tag{37}
$$

Unlike scenario I, the targeting probability now enters the likelihood. The researcher needs to know the specific targeting rule used by the company in order to compute the mixture weights $\Pr(T_i = j | D_{i0} = d_{i0}, Z_i^*),\ j = 0, 1$. However, subject to this assumption, the likelihood above again corrects the β_i distribution, resulting in valid inferences about the mere measurement effect. Assumption 3.1 is now more restrictive: It is crucial to assume that the researcher knows *all* variables used for targeting. If one or several variables used for targeting are unknown – or known but inadvertently dropped from the analysis – and these variables themselves are correlated with β_i, then T_i is no longer independent of β_i even conditional on the included Z_i variables. This implies that the likelihood above is wrong, and we would obtain incorrect estimates of mere measurement.

3.3 Scenario III

In the absence of response bias, we demonstrated that one could still identify two mere measurement effects when the targeting probabilities were not known (Scenario III). However, when response bias is present, this is no longer the case. When the targeting probabilities are unknown, we cannot form the mixture likelihood above since the mixture weights are unknown. Furthermore, without strong assumptions, it is not possible to estimate these weights. In scenario 3, the only data available is the fraction of responders, i.e., $\Pr(\tilde{R} = 1)$. From this we can estimate $\Pr(R = 1, T = 1)$, but we cannot estimate $\Pr(R = 0, T = 1)$ and $\Pr(T = 1)$.

To summarize, we have provided a methodology for recovering the mere measurement effect in the presence of response bias and targeting bias. We have also shown that knowledge of who is targeted (Scenario I) or knowledge of the targeting policy (Scenario II) are required when both biases are present. In the next section, we develop a full model for our empirical application.

4 Model

We consider the shopping experience of each customer of the firm. We assume that customers are not forward looking with respect to purchase and survey behavior, and that they make decisions over discrete time periods. Thus, in every discrete time period, each customer must make a purchase decision. If a customer decides to purchase, then the customer must determine how much she is going to spend in this purchase occasion. The firm decides which customers to elicit for a survey in a given time period based on particular targeting rules. Customers must decide if they wish to respond to a survey request. This decision to respond to a survey may impact their future behavior with respect to the purchase and spending.

Below, we present a formalized model of the customer shopping experience described above. Since we are interested in total expenditure, we model both the customer purchase decision and customer spending.

We assume customer i faces a purchase decision in every period t. This decision is influenced by several factors, including the customer's inherent preference toward making purchases, recency of last purchase, and price. Our focus is on whether survey response in period t also affects future purchase decisions. We use a probit model model with latent utility at time t, defined as:

$$U_{i,t} = \beta_i X_{i,t} + \lambda_u \tilde{R}_{i,t} + \epsilon^u_{i,t}. \tag{38}$$

In this equation, $X_{i,t}$ represents a set of control variables, including the intercept. The individual specific intercept represents the time-invariant inherent customer preference toward making a purchase or unobservable heterogeneity. $\tilde{R}_{i,t}$ is an indicator variable that equals one if customer i has responded to a survey by time t. For example, if a customer responds to a survey at the beginning of period 10 then $\tilde{R}_{i,t} = 1$ for periods 10 and beyond. We assume the error, $\epsilon^u_{i,t}$, is distributed $N(0, 1)$.

Customer i's level of spending at time t is conditional on whether the customer decided to make a purchase at time t. We can express this behavior with the following, where $E_{i,t}$ is the natural log of the level of expenditure for customer i at time t:

$$E_{i,t} = \begin{cases} \theta_i W_{i,t} + \lambda_e \tilde{R}_{i,t} + \epsilon^y_{i,t} & \text{if } U_{i,t} > 0 \\ NA & \text{otherwise} \end{cases} \tag{39}$$

In this equation, $W_{i,t}$ represents a set of control variables, including the intercept. These control variables can be very similar to the control variables related to the purchase decision, $X_{i,t}$. The individual specific intercept represents the time-invariant inherent customer preference toward expenditure. We assume the error, $\epsilon^y_{i,t}$, is distributed $N(0, 1)$.

We have assumed that each customer i has inherent preferences toward the decision to make a purchase and the amount of expenditure in a given purchase occasion. Let β_{i1} and θ_{i1} correspond to the customer specific intercept in the purchase and spending model. To capture customer i's inherent desire to purchase and spend and assume that they are distributed as follows:

$$\begin{pmatrix} \beta_{i1} \\ \theta_{i1} \end{pmatrix} \sim N(\begin{pmatrix} \mu_\beta \\ \mu_\theta \end{pmatrix}, \begin{pmatrix} \sigma_{\beta\beta} & \sigma_{\theta\beta} \\ \sigma_{\beta\theta} & \sigma_{\theta\theta} \end{pmatrix})$$

Thus, we allow for the inherent preferences with respect to purchase and expenditure to be correlated.

The decision to respond to a survey is assumed to be driven by a latent response utility, R^*. The latent response utility for a customer is only calculated when a customer is targeted to be surveyed. Only when R^* is greater than zero does a customer respond to a survey. We assume that the decision to respond to a survey is a function of a customer's inherent preferences for purchase and expenditure. We have shown above that this unobservable heterogeneity is captured through the individual specific intercepts in equations 38 and 39. Thus, R^* can be expressed using the following probit equation:

$$R^*_{i,t} = \eta_0 + \eta_1\beta_{i1} + \eta_2\theta_{i1} + \epsilon^r_{i,t} \tag{40}$$

We assume the error, $\epsilon^r_{i,t}$, is distributed $N(0,1)$. As discussed in section 3, our inferences about the mere measurement effect will hinge on the correct specification of this response model, and the estimates will potentially be sensitive to this specification.

The model above sufficiently accounts for scenario I discussed in section 3.1 where the researcher observes who is targeted (T). In case II, targeting is not observed, but the targeting rule, τ, and the targeting variables, z, are observed. In order for the model to accommodate scenario II, we treat T as a latent variable that we simulate with accordance to τ.

We use Bayesian techniques to estimate the model. Markov Chain Monte Carlo (MCMC) procedures are used to simulate the posterior distribution of model parameters. We also estimate homogeneous coefficients for the control variables $X_{i,t}$ and $W_{i,t}$. Only the intercepts are customer-specific. The estimation procedure is described in detail in the Technical Appendix.

5 Data

We estimate the model using transaction and survey data provided by a catalog retailer. For confidentiality reasons, we cannot reveal the name of the retailer. But, the retailer sells durable goods that are analogous to CDs and books through catalog and Internet channels. Customer transaction data spans the time period January 1995 to December 1999. For each customer order, we observe the date of the order, the items ordered, the price paid per item, and the unique customer ID of the purchaser.

In February 1997, the retailer began a satisfaction survey that was conducted over the phone. From February 1997 to December 1999, the retailer administered twenty-nine surveys. The retailer surveyed between 100 and 300 customers in each survey, which is a small fraction of the total number of consumers. The survey consisted of seventeen questions and included questions related to product and service satisfaction, overall satisfaction, and future purchase intentions. We use data on all customers who responded to the survey, including date of response. In this paper, we focus on the effect of survey response on future customer behavior. While we do have reported satisfaction scores for customers, we do not use the data in this paper.

The retailer also provided the researchers with the firm's survey targeting policy. For each survey, the company created a list of customers who had purchased any item in the previous 30 to 120 days and who had not bought in the previous 30 days. The list was then randomized and a percentage of customers were contacted. A typical list would have thousands of names but only a few hundred consumers completed a survey. The firm did not provide a clear idea of what percentage of customers were contacted. Thus, with respect to equation 6, we know Z, but are uncertain about the value of α. We conducted a sensitivity analysis around different values of α and the results of the analysis are presented in section 6.1.

We restrict attention to customers who purchased at least one time in the six year period and whose first purchase was after January 1, 1995; we exclude customers who are institutional buyers. From our remaining group of customers, we created a random sample of 19,733 customers. Because we have the complete purchase history for all customers, there are no data censoring issues. We aggregate our data so that one month corresponds to a period. In this application, this aggregation is not restrictive since purchases are infrequent. Descriptive statistics for the customers in the sample are provided below in Table 2.

[INSERT TABLE 2 HERE]

The statistics in Table 2 are over the entire purchase history of the customers in the random sample. Customers have an average interpurchase time of 0.64 years, and spend just over $165 per purchase incident. The mean number of purchases per customer is 0.48 per year. 4.9% of the customers in the sample respond to a survey at some point in their purchase history. While a customer could participate in multiple surveys, this is infrequent; only 0.1% of customers complete multiple surveys.

6 Results

Before presenting the parameter estimates for the model described in section 4, we consider a simple univariate analysis. An advantage of this approach is that it is free of model specification issues but a disadvantage is that we are not able to control for targeting and response bias as well as unobserved customer heterogeneity. Our univariate analysis compares the purchase frequency and spending per purchase incident of the responders with all of the non-responders. This approach provides an estimate of $\tilde{\Delta}$ (see equation 16).

[INSERT TABLE 3 HERE]

As shown in Table 3, survey response leads to a 28.2% increase in purchase frequency and a 8.7% increase in spending per purchase incident.

Next, we estimate three versions of the model described in section 4. First, we estimate a model with no control for targeting or response biases (*Model 1*). This is done by removing heterogeneity from the model (i.e., θ and β are constant across customers) and by excluding the response equation 40 from the estimation. The second model (*Model 2*) adds heterogeneity in β_{i1} and θ_{i1}, but excludes the response equation 40 from the estimation. As discussed in section 4, if survey targeting is solely a function of the inherent purchase preference (β_{i1}) and the inherent spending preference (θ_{i1}), then adding heterogeneity controls for targeting bias. Finally, in *Model 3* we control for both targeting and response biases: we include heterogeneity in θ_{i1} and β_{i1}, incorporate the response equation 40, and account for the random targeting policy. As we don't observe the random targeting policy, we assume that five percent of customers who could be targeted are actually asked to participate in the survey ($\alpha = 0.05$). Later, we conduct sensitivity analysis to show that this assumption is robust.

Table 4 below presents the mean and ninety-five percent probability intervals of the parameter posterior distributions for three models. In *Model 1*, the estimate for λ_u is significant, and the mean of the posterior distribution of λ_u is positive (0.066).

When we introduce heterogeneity (*Model 2*), the direction of λ_u is reversed (-0.098), while λ_e is now significant and negative (-0.043).

[INSERT TABLE 4 HERE]

When we control for targeting and response biases (*Model 3*), λ_u increases slightly (-0.096) relative to *Model 2*. The response equation parameters help to explain why λ_u increases. The inherent preference to make a purchase, β_{i1}, has a significant negative effect on the decision to respond to a survey request (η_1). In other words, more frequent buyers are less likely to respond to the survey. This implies that *Model 2* underestimates the mere measurement effect on purchase frequency.

The results in Table 4 demonstrate that survey response changed customer behavior. But, they provide little information about how much customer behavior changed. Since we observe the targeting policy and the z variables, but do not observe who is targeted (scenario II), we can identify $\bar{\Delta}$, Δ_k, and Δ_z. To examine the magnitude of change, we calculate the mean treatment effect (MTE) in the population with respect to purchase frequency ($\bar{\Delta}_u$), expenditure ($\bar{\Delta}_e$), and the joint effect on both purchase frequency and expenditure ($\bar{\Delta}_{joint}$). The treatment we are interested in is the act of responding to a phone survey. The expressions for the MTEs with respect to purchase and expenditure are:

$$\bar{\Delta}_u = \int [\Phi(\lambda_u + \beta_{i1}) - \Phi(\beta_{i1})] p(\beta_{i1}) d\beta_{i1} \tag{41}$$

$$\bar{\Delta}_e = \int [e^{(\theta_{i1} + \sigma_y^2/2)}(e^{\lambda_e} - 1)] p(\theta_{i1}) d\theta_{i1} \tag{42}$$

Note that $\bar{\Delta}_e$ is calculated with respect to dollars rather than the log dollars, which was the dependent variable in the model. Thus, we must also use the estimated expenditure variance, σ_y^2, in the calculation.

We define $\bar{\Delta}_{joint}$ as the joint effect of purchase frequency and spending. For a firm, this captures the expected change in total revenue. The expression for the joint MTE is:

$$\bar{\Delta}_{joint} = \int [\Phi(\lambda_u + \beta_{i1})e^{(\lambda_e+\theta_{i1}+\sigma_y^2/2)} - \Phi(\beta_{i1})e^{(\theta_{i1}+\sigma_y^2/2)}]p(\beta_{i1}, \theta_{i1})d\beta_{i1}\theta_{i1} \quad (43)$$

Table 5 shows MTEs for each of the three models. Note that the MTE captures the absolute change in purchase probability or dollar spending. To compute a percentage change, we estimate the purchase probability and dollar spending for a model that assumes no survey response and refer to this as *Baseline*. $MTE_{percent}$ gives the percentage change relative to the *Baseline*.

In *Model 1*, survey response increases purchase frequency by 15% and has a positive impact on spending (7%). These results are consistent with our univariate analysis in which we found a substantial increase in purchase frequency and a smaller increase in spending. However, after we control for heterogeneity and response bias the change in purchase frequency is -17% and there is a -4.7% change in spending. For the firm, these percent changes are large and economically meaningful. We conclude that the mere measurement effect changed customer behavior in a manner that is both statistically and economically significant.

An important question is whether survey response creates a revenue opportunity for the firm. If so, then merely asking customers to participate in a survey would be profitable. Surprisingly, we find that survey response has a negative effect on overall revenue for the firm. However, there is empirical evidence in previous research that the mere measurement effect can be negative (Dholakia and Morwitz 2002).

[INSERT TABLE 5 HERE]

As mentioned above, since we observe $(\tau(z), z)$, scenario II, in addition to calculating $\bar{\Delta}$, we can compute Δ_z and Δ_k. Previous research has attempted to identify moderators of the mere measurement effect, such as experience (Morwitz, Johnson, and Schmittlein 1993; Janiszewski and Chandon 2007) and satisfaction (Dholakia and Morwitz 2002). To illustrate the calculation of Δ_z, we examine experience as a z variable. If a customer is with the firm for less than four months prior to responding to a survey, we classify him as having low tenure ($z =$ low tenure). If the customer is with the firm for four or more months, we classify him as having high tenure ($z =$ high tenure). To illustrate the computation of Δ_k, we choose k = 5%. This means that a random 5% of the customers who fit the targeting policy are actually targeted by the firm. Table 6 shows the calculated Δ_z & Δ_k using the λ_u and λ_e estimates from *Model 3* in Table 4.

[INSERT TABLE 6 HERE]

For the shorter tenured customers, the effect of survey response is quite negative. Even when we control for targeting and response there is a -19% decrease in purchase frequency and 23% decrease in the total customer expenditure. Relatively longer tenured customers exhibit a 3% decrease in purchase frequency and total customer expenditure. The explanation for this phenomena is still an open question, although response fluency may help explain the varied effect (Janiszewski and Chandon 2007).

We find effects for Δ_5 that are of similar magnitude to the MTEs calculated in *Model 3* of Table 5. Recall that Δ_5 is the effect of a survey on customers in the target population who have a 5% chance of being surveyed. In our application, surveyed customers purchased in the previous 30-120 days (i.e., recent buyers). Because the population and Δ_5 results do not differ substantially, this suggests that recency of purchase is not an important moderator of the mere measurement effect in this application.

6.1 Robustness

In *Model 3* we assume that $\alpha = 0.05$. To examine the robustness of this assumption, we replicate our findings with varying levels of α. As shown in Figures 1-3, the MTEs are stable across all values of α. A value of $\alpha > 0.80$ implies a phone completion rate of less than 1.5%, which is extremely low and often raise questions of survey validity (i.e., response bias). Therefore large values of α imply response rates that are not realistic. Thus, we conclude that our estimates are robust to alternative assumptions of α.

7 Discussion

We present a new methodology that quantifies the mere measurement effect in a non-experimental, field setting. The methodology helps to overcome two major obstacles to measuring the mere measurement effect in a field setting. These are: Who is asked to participate in a survey? Who agrees to respond to a survey? We label these obstacles as targeting bias and response bias, respectively. The new methodology is applied to data provided by a direct mail, catalog company.

The empirical application demonstrates the importance of controlling for these potential biases. We show that failure to control for both targeting and response bias may lead to inaccurate estimates of the mere measurement effect. But, after controlling for these biases, we find that survey response causes consumers to decrease purchase frequency and decrease the dollar amount spent on each order. This is also a new effect that has not been previously measured in non-experimental field settings.

The methodology is also used to examine the mere measurement effect in different sub-populations. For shorter tenured customers, survey response decreases purchase frequency, spending per order, and total customer spending. Relatively longer tenured customers exhibit smaller decreases in purchase frequency and total customer spending following survey response. These findings demonstrate differential effects of survey response based on length of tenure with the firm. This is also a new effect that has not been previously measured in non-experimental field settings.

Our study highlights several alternative ways to measure the mere measurement effect: the mere measurement effect for customers with the same characteristics (Δ_z), the average mere measurement effect for customers that are targeted for survey ($\bar{\Delta}_{T=1}$), the average mere measurement effect for the entire population ($\bar{\Delta}$), and the average mere measurement effect for the group of customers with targeting probability k, (Δ_k). Laboratory studies typically focus on $\bar{\Delta}_{T=1}$ as the mere measurement effect. Past empirical studies have not discerned between these different types of causal effects. For example, Dholakia and Morwitz (2002) use customers who are active to design their field experiment and match responder and non-responder groups based on observable variables. In our notation they are calculating Δ_z, the mere measurement effect for a subset of the entire population. In contrast, we focus on the average mere measurement effect for the entire population. For practitioners, it is important to understand what kind of mere measurement effect is being calculated, especially if one wants to generalize the results to a broader population.

It has been recognized that the mere measurement effect poses a potential problem for survey research. If measurement itself is affecting customer behavior, then inferences from customer satisfaction surveys may be biased. Our methodology allows future researchers to control for the mere measurement effect in survey research. The methodology may also facilitate future research on mere measurement by broadening the set of empirical applications to non-experimental field settings.

References

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* 31:566–572.

———. 2005. “Do Intentions Really Predict Behavior? Self-Generated Validity Effects in Survey Research.” *Journal of Marketing* 69:1–14.

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* 29:159–167.

Feldman, J.M., and Jr. J.G. Lynch. 1988. “Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior.” *Journal of Applied Psychology* 73:421–435.

Fitzsimons, Gavan J., and Vicki G. Morwitz. 1996. “The Effect of Measuring Intent on Brand-Level Purchase Behavior.” *Journal of Consumer Research* 23:1–11.

Heckman, J.J. 1978. “Dummy Endogenous Variables in a Simultaneous Equations System.” *Econometrica* 46:931–960.

Janiszewski, Chris, and Elise Chandon. 2007. “Transfer-Appropriate Processing, Response Fluency, and the Mere Measurement Effect.” *Journal of Marketing Research* 44:309–323.

Morwitz, Vicki G., and Gavan J. Fitzsimons. 2004. “The Mere-Measurement Effect: Why Does Measuring Intentions Change Actual Behavior?” *Journal of Consumer Psychology* 14:64–74.

Morwitz, Vicki G., Eric Johnson, and David Schmittlein. 1993. "Does Measuring Intent Change Behavior." *Journal of Consumer Research* 20:46–61.

Sherman. 1980. "On the self-erasing nature of errors of prediction." *Journal of Personality and Social Psychology* 39:211–221.

Sprott, David E., Eric R. Spangenberg, Lauren G. Block, Gavan J. Fitzsimons, Vicki G. Morwitz, and Patti Williams. 2006. "The question-behavior effect: What we know and where we go from here." *Social Influence* 1:128–137.

8 Appendix A

Straightforward calculations show that

$$\Pr(D=1|\tilde{R}=0)=\xi\Pr(D=1|T=1,R=0)+(1-\xi)\Pr(D=1|T=0), \quad (44)$$

where

$$\xi=\frac{\Pr(T=1)\Pr(R=0)}{\Pr(T=1)\Pr(R=0)+\Pr(T=0)}. \quad (45)$$

A comparison of responders and non-responders yields the following effect

$$\tilde{\Delta}=\Pr(D=1|\tilde{R}=1)-\Pr(D=1|\tilde{R}=0). \quad (46)$$

The first probability is

$$\Pr(D=1|\tilde{R}=1)=\int\Pr(D=1|Z=z,R=1)p(z|T=1)dz. \quad (47)$$

The second probability is

$$\Pr(D=1|\tilde{R}=0)=\xi\int\Pr(D=1|Z=z,R=0)p(z|T=1)dz+$$
$$(1-\xi)\int\Pr(D=1|Z=z,R=0)p(z|T=0)dz. \quad (48)$$

Substituting in (46) and rearranging gives

$$\tilde{\Delta} = \bar{\Delta}_{T=1} + (1-\xi)\int \Pr(D=1|Z=z,R=0)\big[p(z|T=1) - p(z|T=0)\big]dz \quad (49)$$

9 Appendix B

Scenario	$\tilde{\Delta}$	$\bar{\Delta}_{T=1}$	Δ_k	Δ_z
T observed	-	x	x	-
$(\tau(\mathbf{z}), \mathbf{z})$ known	x	-	x	x
z known	x	-	-	x
$z_1 \subset z$ known	-	-	-	-
Nothing is known	-	-	-	-

Table 1: Identifiable Average Mere Measurement Effects

Sample Size (N)	19,733
Number of Responders	970 (4.9%)
Avg. Interpurchase Time (Yr)	0.64
Avg. Purchase Incidents (PI)/Yr	0.48
Avg. Spending ($)/PI	$165.50
Avg. Spending ($)/Yr	$80.13

Table 2: Descriptive Statistics for the Random Sample

Purchase Incidents(PI)/Year	
Responders	0.61
Non-Responders	0.47
Mean Treatment Effect ($\%MTE_{PI/YR}$) = 28.2%	
Average Spending($)/PI	
Responders	$178.91
Non-Responders	$164.65
Mean Treatment Effect ($\%MTE_{\$/PI}$) = 8.7%	

Table 3: Univariate Survey Analysis

parameter	*Model 1*	*Model 2*	*Model 3*
Purchase Model			
λ_{u1}	0.066 [0.05, 0.082]	-0.098 [-0.117, -0.079]	-0.096 [-0.116, -0.076]
$\beta_{1\mu}$	-1.744 [-1.744, -1.744]	-1.903 [-2.375, -1.431]	-1.898 [-2.37, -1.426]
Expenditure Model			
λ_{e1}	0.068 [0.04, 0.096]	-0.043 [-0.075, -0.011]	-0.048 [-0.081, -0.015]
$\theta_{1\mu}$	5.107 [5.107, 5.107]	5.088 [4.37, 5.806]	5.085 [4.371, 5.799]
Response Model			
η_0	- [-, -]	- [-, -]	0.011 [-0.117, 0.139]
η_1	- [-, -]	- [-, -]	-0.022 [-0.043, -0.001]
η_2	- [-, -]	- [-, -]	0.017 [-0.005, 0.039]

Note: N = 19,733 and α = 0.05

Table 4: Parameter Estimates for the Three Models of Customer Purchase, Expenditure, and Survey Response

Condition	MTE	MTE_b	$MTE_{percent}$
PURCHASE FREQUENCY (λ_u)			
Model 1	0.006	0.041	15.1%
Model 2	-0.007	0.042	-17.5%
Model 3	-0.007	0.043	-17.1%
EXPENDITURE (λ_e)			
Model 1	\$15.93	\$226.84	7.0%
Model 2	\$-10.82	\$258.2	-4.2%
Model 3	\$-12.06	\$256.15	-4.7%
JOINT (PURCHASE AND EXPENDITURE)			
Model 1	\$2.14	\$9.21	23.2%
Model 2	\$-2.25	\$10.75	-21.0%
Model 3	\$-2.27	\$10.82	-21.0%

Table 5: Mean Treatment Effects for Purchase Frequency, Expenditure, and Overall Spending for the Three Models

	$\Delta_{z=LowTenure}$	$\Delta_{z=HighTenure}$	$\Delta_{k=5\%}$
Purchase Frequency (λ_u)	-19.3%	-3.3%	-16.9%
Expenditure (λ_e)	-4.7%	0.0%	-4.7%
Joint (Purchase & Expenditure)	-23.1%	-3.6%	-20.9%

Table 6: Example Identifiable Mere Measurement Effects using Parameters from Model 3

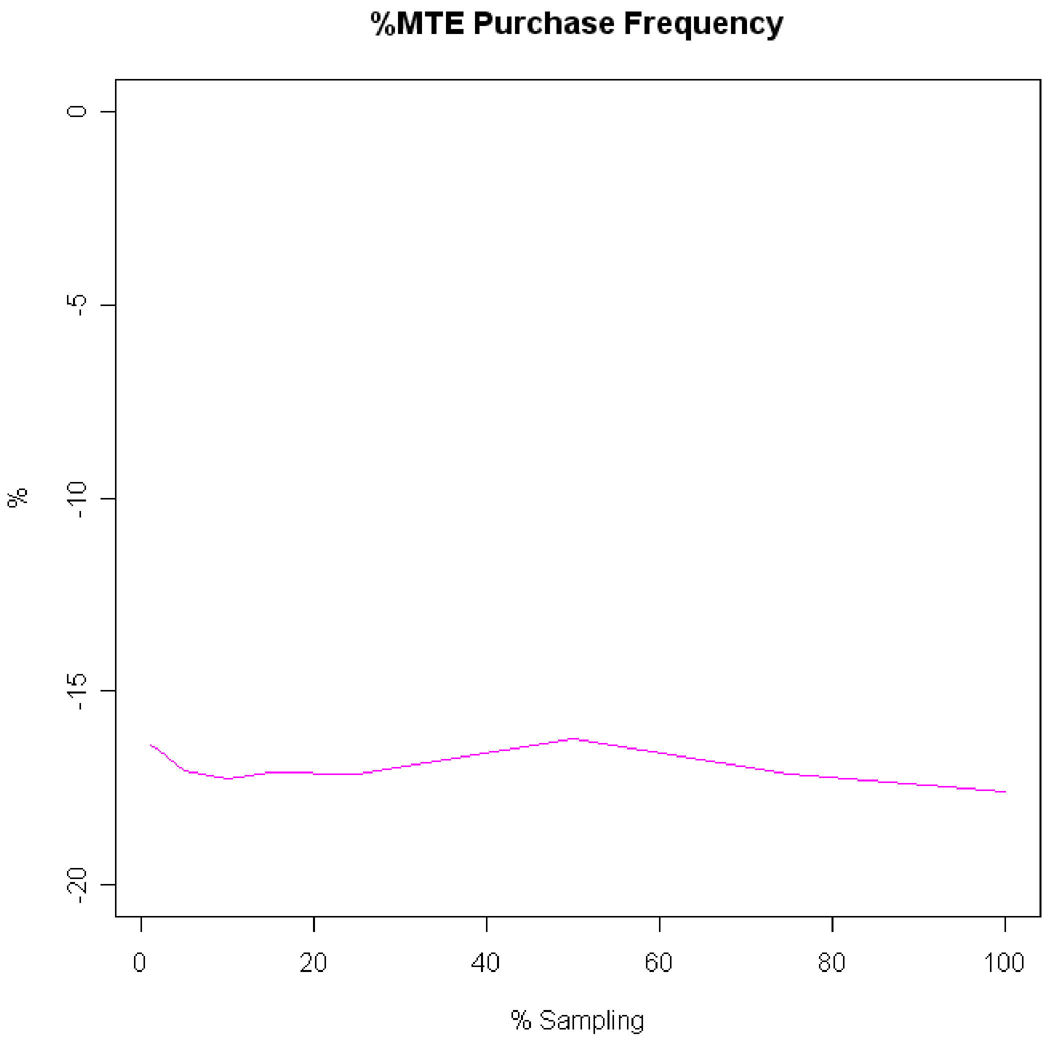

Figure 1: $\%MTE_{PurchaseFrequency}$ as a function of the Random Sampling %

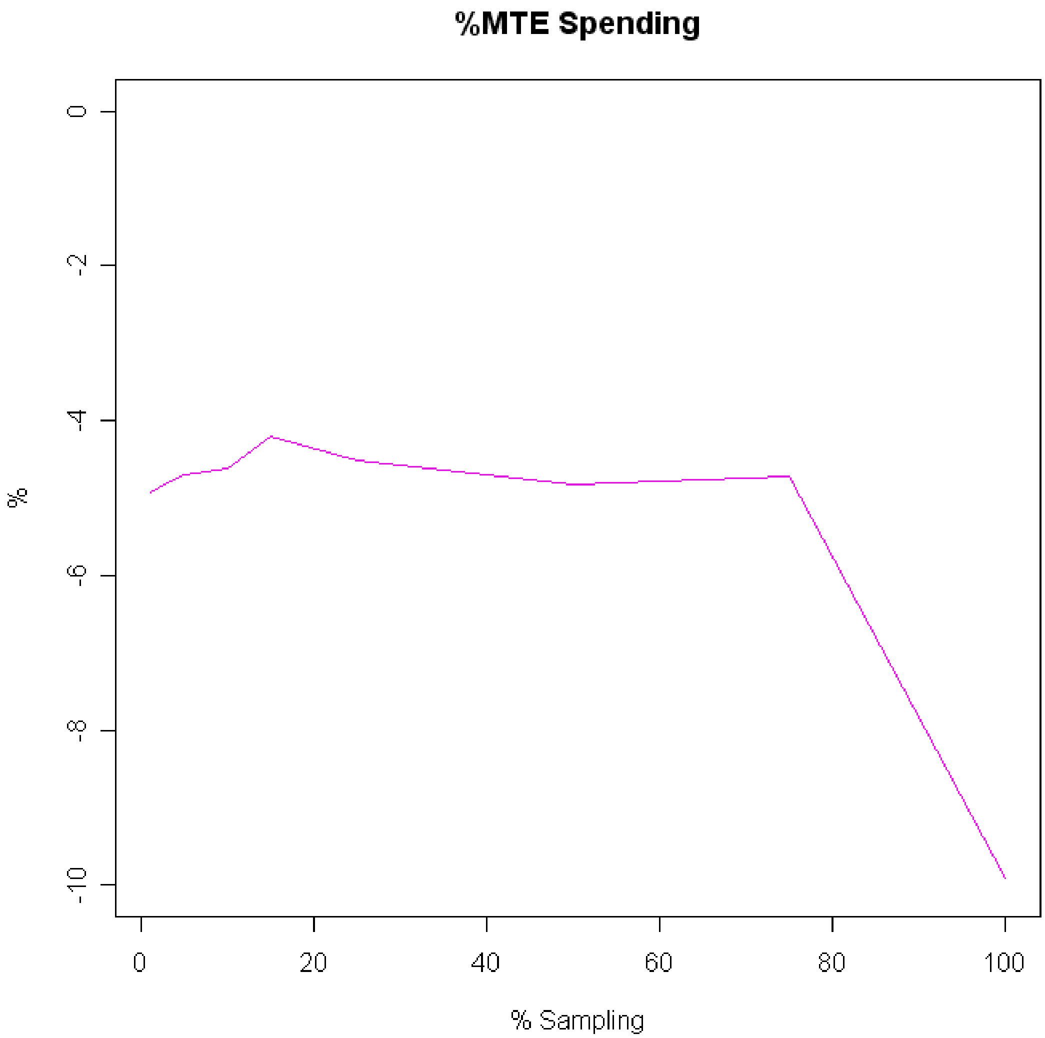

Figure 2: $\%MTE_{Spending}$ as a function of the Random Sampling %

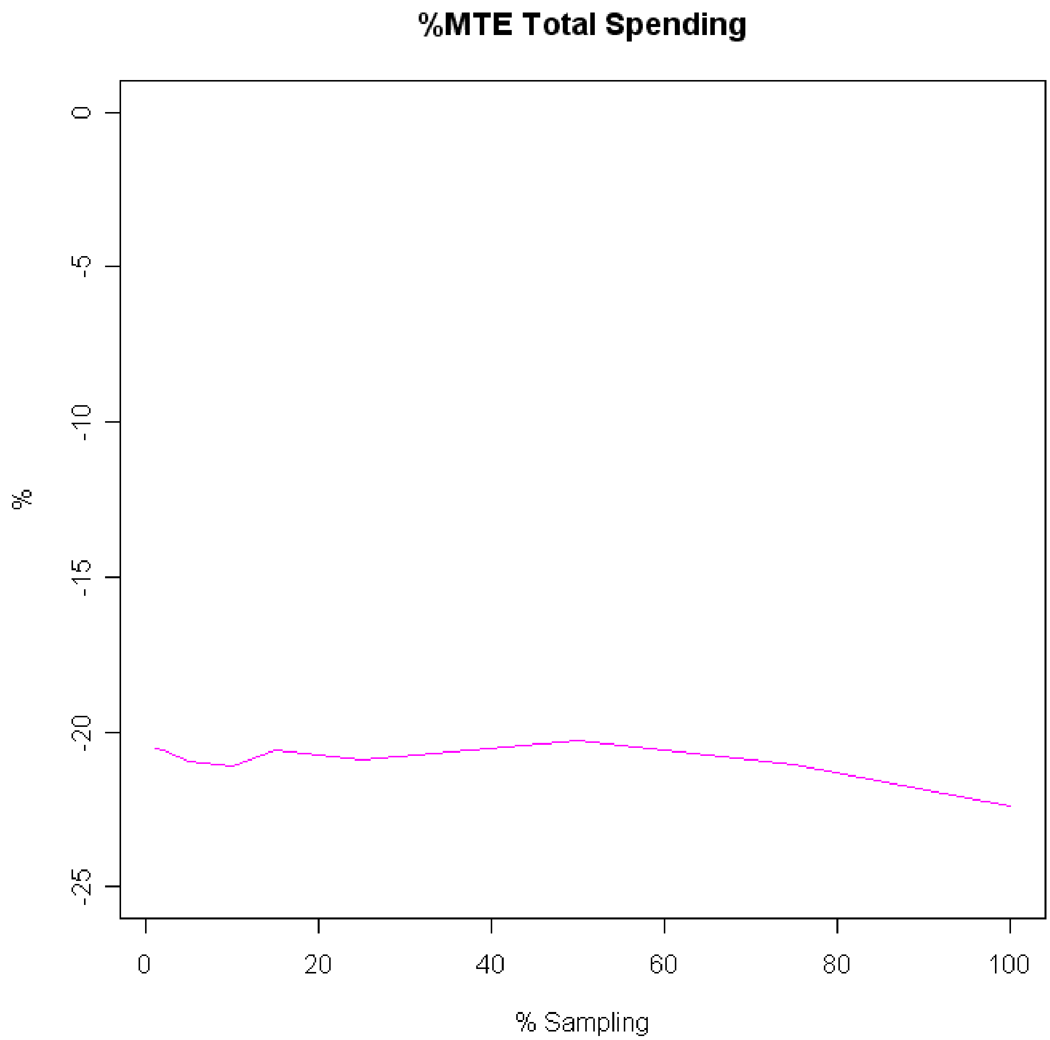

Figure 3: $\%MTE_{TotalRevenue}$ as a function of the Random Sampling %