# Investigating Endogeneity Bias in Conjoint Models

by

Qing Liu
Department of Statistics
Ohio State University
qliu@stat.ohio-state.edu

Thomas Otter
Fisher College of Business
Ohio State University
otter.2@osu.edu

and

Greg M. Allenby
Fisher College of Business
Ohio State University
allenby.1@osu.edu

January, 2006

The authors thank John Hauser, Rich Johnson and Olivier Toubia for helpful comments.

# Investigating Endogeneity Bias in Conjoint Models

Abstract

The use of adaptive designs in conjoint analysis has been shown to lead to an endogeneity bias in part-worth estimates using sampling experiments. In this paper, we re-examine the endogeneity issue in light of the likelihood principle. The likelihood principle asserts that all relevant information in the data about model parameters is contained in the likelihood function. We show that adhering to the likelihood principle leads to analysis where the endogeneity bias becomes irrelevant. The likelihood principle is implicit to Bayesian analysis, and discussion is offered about the role of sampling experiments in Bayesian versus frequentist analysis.

Keywords: Likelihood principle, frequentist analysis, adaptive design, Bayes Theorem

# Investigating Endogeneity Bias in Conjoint Models

## 1. Introduction

The recent paper by Hauser and Toubia (2005) "The Impact of Utility Balance and Endogeneity in Conjoint Analysis" raises an interesting set of issues related to a wide class of marketing models. In their paper, they present evidence that adaptive designs, where answers to early questions in a conjoint interview are used to select later questions, induce biases in the estimated part-worths. While their paper focuses on analysis associated with Sawtooth Software's popular ACA (Adaptive Conjoint Analysis) software, the implications of their analysis reach far beyond conjoint analysis, sequential design problems, and utility balance, touching on important philosophical issues at the core of statistical inference.

In this paper, we re-examine the endogeneity bias identified by Hauser and Toubia (HT), and explain its presence using traditional econometric methods. Bias is an aspect of statistical inference that relies on the notion of a sampling experiment, where hypothetical datasets are used to characterize the performance of an estimator. Sampling experiments are useful to study properties of estimators and other procedures when real data are not available. But, when data are available, we argue that analysis should proceed according to the likelihood principle as originally proposed by R.A. Fisher (1922).

The likelihood principle asserts that the likelihood contains all of the information about model parameters (i.e., conjoint part-worths) in the data. We show that, according to the likelihood principle, endogeneity created by adaptive questioning is not of concern. That is, the analysis conducted by standard software packages, including Sawtooth Software, is correct, so long as all the data collected in the adaptive procedure are included in the analysis. Our discussion of the likelihood principle raises a number of philosophical issues at the core of statistical inference that highlight the difference between classical (i.e., frequentist) and Bayesian philosophies.

Our analysis of endogeneity bias in conjoint models uses HT as a springboard for discussion, and is not meant to criticize their findings. In fact, our analysis covers some of the same ground as their analysis, restating their findings in terms more familiar to the statistics literature. Our examples are sometimes similar to those examined by HT, and sometimes depart from theirs to provide addition insight and analysis. We applaud their interest in analyzing the issue of endogeneity bias because of its importance to both practitioner and academic researchers. Our analysis begins with a review of endogeneity bias in regression model caused by adaptive designs, restating many of the points made by HT. We then introduce the likelihood principle, and examine its implication for data analysis. Our results show that once the data are collected, the issue of endogeneity bias disappears. Discussion is provided about the importance of bias in the evaluation of estimators, and conditions when the issue of endogeneity bias will not disappear. Concluding comments are then offered.

## 2. Endogeneity Bias

Endogeneity bias arises in regression analysis $y = X\beta + \varepsilon$ when the regressors, X, are not independent of the errors, $\varepsilon$. When a specific realization of the regressors, $x'_t$, is selected based on the outcome of previous choices, $y_{<t}$, as is the case for sequential design data, then $x'_t$ is determined from within the system of study and is not independently determined. In this case, the traditional ordinary least squares estimator is biased because the vector space of the regressors is dependent on the error realizations. Regression coefficient bias is defined as the difference between the true value of the estimate and its expected value over hypothetical set of data (D):

$$
\begin{aligned}
E_{D|\beta}\left[\hat{\beta}\right] &= E_{D|\beta}\left[\left(\mathrm{X'X}\right)^{-1}\mathrm{X'y}\right] \\
&= E_{D|\beta}\left[\left(\mathrm{X'X}\right)^{-1}\mathrm{X'}\left(\mathrm{X}\beta+\varepsilon\right)\right] \qquad (1) \\
&= \beta + E_{D|\beta}\left[\left(\mathrm{X'X}\right)^{-1}\mathrm{X'}\varepsilon\right]
\end{aligned}
$$

The presence of endogenous responses biases the regression estimate because the second term in equation (1) is not equal to zero. To see this, consider a simple example where there is just one regressor without an intercept ($y=x\beta+\varepsilon$), and just two observations. The OLS estimator is given by:

$$\hat{\beta} = \frac{x_1 y_1 + x_2 y_2}{x_1^2 + x_2^2} \qquad (2)$$

Then, if $x_2$ is determined by the value of $y_1$, we have:

$$E_{D|\beta}\left[\hat{\beta}\right] = \beta + E_{D|\beta}\left[\frac{x_1\varepsilon_1}{x_1^2+x_2^2} + \frac{x_2\varepsilon_2}{x_1^2+x_2^2}\right] \qquad (3)$$

The expectation of the first term in the brackets is not equal to zero because $y_1$, and hence $\varepsilon_1$, is used to determine $x_2$. Since $\varepsilon_1$ and $x_2$ are not independently determined, $\mathrm{E}[\mathrm{f}(x_2)\varepsilon_1] \neq 0$, and

$$E_{D|\beta}\left[\frac{x_1\varepsilon_1}{x_1^2+x_2^2}\right] \neq E_{D|\beta}\left[\frac{x_1}{x_1^2+x_2^2}\right] E_{D|\beta}\left[\varepsilon_1\right] = 0 \qquad (4)$$

Thus, for regression estimates to be unbiased it must be the case that all error realizations are independent of all values of the independent (X) variables.

The lack of independence between regressors and errors occurs in many models. Models with a lagged dependent variable (e.g., a Koyck lag models, autoregressive models) violate the assumption that the regressors are independent of the error terms. Endogeneity bias also occurs in models with sequential design in which design points are selected based on previous responses. While the specific algorithm used to select design points (e.g., utility balance in ACA) may exacerbate the extent of endogeneity bias documented by frequentist analyses of small sample situations with relatively large error variance, the bias is present whenever the vector of residuals, $\varepsilon$, is not independent of all realizations of the regressors.

Figure 1 illustrates the biasing effect of endogeneity using the simple problem similar to that studied by HT. The analysis involves 1000 replicates of samples consisting of 1000 homogenous respondents each. Each respondent supplies three observations, where the regressor for the third observation is determined as a function of the first two observations. The model is $y_t = \beta_1 x_{1t} + \beta_2 x_{2t} + \varepsilon_t$, t=1,2,3 and $\varepsilon_t \sim$ Normal(0,25). The value of x for the first observation is $x'_1 = (1,0)$, the value of x for the second observation is $x'_2=(0,1)$ and:

$$x_3' = \begin{cases} (1,-1) & if \;\; y_1 y_2 > 0 \\ (1,1) & if \;\; y_1 y_2 \le 0 \end{cases} \tag{5}$$

The design rule in equation (5) is meant to mimic the "utility balance" criteria used by the ACA software. The true values of the regression coefficients are $\beta_1$=1 and $\beta_2$=2. The figure shows that the regression coefficients exhibit positive bias, with E[$\beta_1$] = 1.198 and E[$\beta_2$] = 2.406. Each triangle character in the figure represents the mean of individual-level OLS estimates computed from one sample of 1000 homogenous respondents.

Econometricians recognize the bias present in regression models when functions of lagged dependent variables are included in the model specification, and, as a result, the finite sampling properties of these models are generally unknown (see Judge et.al., p.575). This has led to the use of asymptotics to characterize the sampling properties of estimators in such situations. In particular, the probability limit, or "plim" is used to describe the behavior of estimators as the sample size increases. The probability limit is a formal expression for the consistency of an estimator. An estimator is said to be consistent if the probability limit of obtaining an estimate arbitrarily close to the true parameter value equals 1 in infinite samples, i.e. the probability that an estimate from a sample of size T falls within the interval [$\beta$ - $\varepsilon$, $\beta$ + $\varepsilon$] goes to one as the sample size increases, no matter how small $\varepsilon$:

$$\lim_{T\to\infty} P\left(|\hat{\beta}_T - \beta| < \varepsilon\right) = 1 \tag{6}$$

This result, when coupled with a result known as Slutsky's theorem, can be used to show that the OLS estimator is asymptotically consistent. Slutsky's theorem states that if g(.) is a continuous function and $z_T$ is some random variable that depends on T, then

$$\text{plim } g(z_T) = g(\text{plim } z_T) \tag{7}$$

Thus, from equation (1) we can derive the plim of the regression estimate as the sample size increases as:

$$\begin{aligned} \text{plim}\,\hat{\beta} &= \beta + \text{plim}\left((X'X)^{-1}X'\varepsilon\right) \\ &= \beta + \text{plim}\left((X'X)/T\right)^{-1} \text{plim}\left(X'\varepsilon/T\right) \\ &= \beta \end{aligned} \tag{8}$$

where the last equality holds if plim $((X'X)/T)^{-1}$ converges to a finite-valued matrix (i.e., $\Sigma_{XX}^{-1}$), and plim($X'\varepsilon/T$) converges to zero. The later condition holds as long as $E_D[x_{it}\varepsilon_t]=0$ for any t. Consistency only requires independence between regressor values $x_t$ and their corresponding error term $\varepsilon_t$, and not the entire vector of errors. Thus, the requirements for asymptotic consistency are easier to obtain than the requirements of unbiasedness.

To illustrate the consistency of the OLS estimate in the simulation study, we obtained pooled OLS estimates for each of the replicated data sets, i.e. estimated regression coefficients using all 3000 observations (1000 respondents each supplying three observations) contained in each data set. The pooled estimates are plotted in figure 1 as black diamonds. Clearly, the pooled estimates are not affected by the bias; their mean is $\hat{\beta}' = (1.005, 2.000)$ which is close to the true value of $\beta' = (1, 2)$. Thus, while the simulation study of HT does exhibit an endogeneity bias in small samples, the estimates are asymptotically consistent.

— Figure 1 —

## 3. The Likelihood Principle

The study by HT touches on an important philosophical point about the role of sampling experiments in managerial decision making. Bias is obviously a concern in conjoint analysis where part-worth estimates are used to set prices and guide product formulation. But, managerial decision making doesn't involve hypothetical samples of data by which properties such as bias can be assessed. A more relevant issue facing management is how best to analyze the one dataset they have collected. Thus, while statisticians may be interested in the sampling behavior of an estimator across multiple samples of data, practitioners often require information derived from just one sample.

A principle of statistical inference first introduced by Fisher (1922) is that the likelihood function contains all the information in the data about the model parameters. The likelihood principle implies that, if two likelihoods for a parameter $\beta$ are proportional, then we should make the same inference for $\beta$ regardless of which likelihood we use. This principle has a number of significant and far-reaching implications. Berger and Wolpert (1984) provide an excellent review of this topic and provide many interesting examples.

An example of the application of the likelihood principle involves sequential sampling. Suppose we observe the number of successes Z in $n$ independent Bernoulli trials with success probability $\theta$. The likelihood for the data is

$$\pi_1\left(z \mid \theta\right) = \binom{n}{z} \theta^z \left(1-\theta\right)^{n-z} \tag{9}$$

Now, suppose that instead of holding fixed the number of trials, we were to decide to sample until we obtained z successes and then observed the realization of N, the number of trials to the $z^{th}$ success. In this scenario, N has a negative binomial distribution so the model is

$$\pi_2\left(n \mid \theta\right) = \binom{n-1}{z-1} \theta^z \left(1-\theta\right)^{n-z} \tag{10}$$

and we have that $\pi_1(z \mid \theta) \propto \pi_2(n \mid \theta)$. Despite the fact that the sampling scheme and the dependent variable are different, the likelihood principle states that we should ignore this and make the same inference about $\theta$ in both cases.

The likelihood function for our example above should acknowledge that $x_3$, the third design point, is determined from within the system of study and is therefore endogenous:

$$\begin{aligned} \pi\left(y_1, y_2, x_3, y_3 \mid \beta, \sigma^2\right) = {} & \pi_1\left(y_1, y_2 \mid \beta, \sigma^2\right) \\ & \times \pi_2\left(x_3 \mid y_1, y_2, \beta, \sigma^2\right) \\ & \times \pi_3\left(y_3 \mid x_3, \beta, \sigma^2\right) \end{aligned} \tag{11}$$

The first factor on the right side of equation (12) corresponds to the first two observations where the design points $x_1$ and $x_2$ are determined beforehand. The likelihood for these observations corresponds to that found in a standard regression analysis:

$$\pi_1\left(y_1, y_2 \mid \beta, \sigma^2\right) = \frac{1}{2\pi\sigma^2} \exp\left[\frac{-1}{2\sigma^2}\left((y_1 - x_1'\beta)^2 + (y_2 - x_2'\beta)^2\right)\right] \tag{12}$$

Similarly, the third factor, which conditions on the realized design point $x_3$, is the same as found in standard analysis:

$$\pi_3\left(y_3 \mid x_3, \beta, \sigma^2\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{-1}{2\sigma^2}\left((y_3 - x_3'\beta)^2\right)\right] \tag{13}$$

The effect of endogeneity is therefore isolated in the second factor, $\pi_2$.

The likelihood for $x_3$ is determined by equation (5), where a positive product of $y_1$ and $y_2$ leads to the selection of $x_3'$=(1,-1) and a negative product yields $x_3'$ = (1,1). Given $y_1$ and $y_2$, the selection of $x_3$ is deterministic, or:

$$\pi_2\left(x_3 \mid y_1, y_2, \beta, \sigma^2\right) = \pi_2\left(x_3 \mid y_1, y_2\right) = 1 \tag{14}$$

Thus, while the likelihood for design point $x_3$ is unconditionally dependent on the model parameters $\beta$ and $\sigma^2$, the likelihood is independent of these parameters given $y_1$ and $y_2$. This makes intuitive sense since we learn about $\beta$ and $\sigma^2$ from the respondent, not the mechanism used to select the design point. This mechanism affects the amount that we learn through issues related to design efficiency, but not what we learn.

The likelihood principle implies that, given the data, the selection mechanisms such as ACA's utility balance is irrelevant to analysis. The likelihood of model parameters is unchanged if $\pi_2$ is included. That is, analyses using standard methods for these data are correct.

*Heterogeneity and Selection Bias*

In section 6 of their paper, HT suggest that the presence of heterogeneity might compound the biasing effects of utility balance by also biasing the hyper-parameters of the model that describe the mean part-worths (i.e., the mean of the random-effects distribution). The likelihood principle again implies that, conditional on the data, the use of utility balance is irrelevant to the analysis. In the presence of consumer heterogeneity, a respondent's contribution to the likelihood is

$$\ell\left(\beta_i\right) = \pi\left(D_i \mid \beta_i\right) \tag{15}$$

and, by conditional independence, the sample likelihood is

$$\ell\left(\{\beta_i\}_{i=1}^{N}\right) = \prod_i \pi\left(D_i \mid \beta_i\right) \tag{16}$$

Equation (14) demonstrates that adaptive questioning does not change the likelihood for any of the "i" respondents. The total sample likelihood is therefore not affected. In the setting of a Bayesian hierarchical model with heterogeneity, a random-effects distribution for $\beta_i$ is introduced along with a prior distribution for the hyper-parameters, $\tau$:

$$\pi\left(\{\beta_i\}_{i=1}^{N}, \tau \mid \{D_i\}_{i=1}^{N}\right) \propto \prod_i \pi\left(D_i \mid \beta_i\right) \times \pi\left(\beta_i \mid \tau\right) \times \pi\left(\tau\right) \tag{17}$$

This corresponds to a data generating process where the $\beta_i$ are modeled as draws from the random-effects distribution $\pi(\beta_i \mid \tau)$ and the data is generated from the likelihood $\pi(D_i \mid \beta_i)$. Since ACA's adaptive questioning does not change the likelihood, it is not relevant to the analysis, whether heterogeneity is present or absent. Moreover, posterior estimates of the hyper-parameters $\tau$ will converge to the true data generating values as the sample size increases, even if each consumer provides a limited amount of information and the individual-level parameters, $\beta_i$, experience shrinkage toward the hyper-parameter $\tau$. This result is due to exchangeability of consumers implied by the random-effects distribution – adaptive questioning does not affect the exchangeability of the response vectors obtained from each consumer.

We investigate the impact of endogenous covariates in a hierarchical Bayes model by adding heterogeneity to our earlier example. That is, instead of 1000 identical respondents, we assume that the part-worths for the respondents follow a random-effects distribution:

$$\beta_i \sim Normal\left(\bar{\beta} = \begin{pmatrix} 2 \\ 1 \end{pmatrix}, V_\beta = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right) \quad \text{i = 1, …, 1000} \tag{18}$$

where i indexes the respondents. Diffuse prior distributions, centered on the true values, are assumed for the hyper-parameters:

$$\pi\left(\sigma^2\right)=\frac{\upsilon_0 s_0^2}{\chi^2_{\upsilon_0}} \qquad \text{with } \upsilon_0 = 5 \text{ and } s_0^2 = 25 \tag{19}$$

$$\pi\left(V_\beta\right)=IW\left(\upsilon_0,V_0\right) \quad \text{with } \upsilon_0 = 10 \text{ and } V_0 = 10I \tag{20}$$

Figure 2 compares Bayesian analysis of $\bar{\beta}$ for exogenous versus endogenous covariates. Time series plots of the draws of $\bar{\beta}$ for these two sets of data show little difference. The posterior means for the plot with exogenous $x_3$ is $\bar{\beta}'$ = (2.00, 0.96) with a posterior standard deviation of (0.12, 0.11). For $x_3$ endogenous, the posterior mean of $\bar{\beta}'$ = (2.01, 0.96) with a posterior standard deviation of (0.12, 0.12). These results agree with the implications of the likelihood principle – analysis that conditions on the observed data is not affected by endogenously determined covariates. Analysis with fewer respondents (e.g., 100) yield similar results, although the precision of the parameter estimates is less. Managerial decisions based on the analysis of the data are also unaffected.

— Figure 2 —

## 4. Discussion

Our analysis shows that the presence of endogenously determined covariates leads to small-sample bias in conjoint part-worths that diminishes as the sample size increases. The likelihood principle, on the other hand, asserts that likelihood-based analysis should ignore the fact that the covariates are endogenously determined. In this section we discuss whether endogeneity bias is relevant to analysis and managerial decision-making. Marketing data, after all, is characterized by short panel sizes where properties such as consistency may have little relevance in a specific research study.

Consider two types of managers – Manager A at a firm like Sawtooth Software that develops statistical procedures for use by others, and Manager B at a user firm who makes inferences about part-worths in specific instances. Manager A could be worried about endogeneity bias because he or she is constructing an instrument that will be used repeatedly. However, we argue that endogeneity bias should be of less concern to Manager B, particularly if data have already been collected. When data are already available, there is no such thing as a sampling experiment – there is just one set of data that is used to make inferences about part-worths and other model parameters.

Prior to data collection, Manager B may be concerned about the manner in which the data will be collected, and the anticipated method used to analyze these data. While data collection and data analysis are related tasks, and managerial decisions regarding them are often made jointly in light of commercially available software, they are not necessarily related. As discussed by HT and others, utility balance may be pursued for reasons that are difficult to quantify within a statistical model. Forcing respondents to select from among alternatives with nearly equal value may avoid scaling problems that occur when one alternative dominates the rest, and may encourage respondents to more carefully evaluate the alternatives. These aspects are not reflected in the standard linear models used in traditional conjoint analysis, and therefore do not enter into traditional analysis. Yet, utility balance may lead to improvements in data collection that improve the overall quality of part-worth estimates.

*Sampling Properties*

The distinction between analysis based on one realized, versus many hypothetical, datasets is at the core of statistical inference, and there are two points of view – Bayes and frequentist. Bayesian analysis conditions on the specific data confronting the analyst and adheres to the likelihood principle. Hypothesis testing, for example, is conducted as Pr(H|D) where H denotes the hypothesis and D denotes the data. In contrast, frequentist testing conditions on the hypothesis, not the data. The p-value, for example, is computed as the probability of observing the realized test statistic, or a test statistic greater in magnitude, given the null hypothesis. Thus, the p-value includes as evidence data that are not observed – data that lead to more extreme outcomes of the test statistic. This aspect of frequentist inference violates the likelihood principle. Thus, while the likelihood principle may appear to be a reasonable tenant for statistical analysis, it is not universally accepted.

Bias is one measure of the performance of an estimator, and while managers should be concerned about bias prior to data collection, bias should not be used as a "litmus test" for selecting an estimator. Bias is an aspect of statistical risk, defined as the expected loss from incorrectly estimating a respondent's true part-worths (β). For squared error loss we have:

$$\begin{aligned} Risk(\beta) &= E_{D|\beta}\left[\left(\hat{\beta}-\beta\right)^2\right] \\ &= E_{D|\beta}\left[\left(\left(\hat{\beta}-E_{D|\beta}\left[\hat{\beta}\right]\right)+\left(E_{D|\beta}\left[\hat{\beta}\right]-\beta\right)\right)^2\right] \\ &= E_{D|\beta}\left[\left(\hat{\beta}-E_{D|\beta}\left[\hat{\beta}\right]\right)^2+\left(E_{D|\beta}\left[\hat{\beta}\right]-\beta\right)^2+2\left(\hat{\beta}-E_{D|\beta}\left[\hat{\beta}\right]\right)\left(E_{D|\beta}\left[\hat{\beta}\right]-\beta\right)\right] \\ &= Var\left(\hat{\beta}\right)+Bias\left(\hat{\beta}\right)^2 \end{aligned} \tag{21}$$

where the cross-product (third) term is zero because the expectation of its first factor in parentheses is zero. Bias can be traded off against variance to obtain lower risk, as it is in ridge regression, and is not usually pursued as a goal in and of itself. Many biased estimators have excellent sampling-theory properties.

Bayesian estimators ($\tilde{\beta}$), for example, minimize expected loss with respect to the posterior distribution (see Rossi, Allenby and McCulloch, p.17-18):

$$\min_{\tilde{\beta}}\left\{E_{\beta|D}\left[L\left(\tilde{\beta},\beta\right)\right]=\int L\left(\tilde{\beta},\beta\right)\pi\left(\beta\mid D\right)d\beta\right\} \tag{22}$$

where L(.) is the loss (e.g., MSE) associated with using $\tilde{\beta}$ to estimate β. Sampling properties of $\tilde{\beta}$ can be studied across multiple realizations of the data:

$$\begin{aligned} E_D\left[E_{\beta|D}\left[L\left(\tilde{\beta},\beta\right)\right]\right] &= \int\int L\left(\tilde{\beta},\beta\right)\,\pi\left(\beta\,|\,D\right)\,\pi\left(D\right)\,d\beta\ dD \\ &= \int\int L\left(\tilde{\beta},\beta\right)\,\pi\left(D\,|\,\beta\right)\,\pi\left(\beta\right)\,dD\ d\beta \\ &= E_\beta\left[E_{D|\beta}\left[L\left(\tilde{\beta},\beta\right)\right]\right] \\ &= E_\beta\left[Risk\left(\beta\right)\right] \end{aligned} \tag{23}$$

The last equality shows that Bayes estimators have the property of minimizing expected risk, where the expectation is taken with respect to the prior distribution. Moreover, while Bayesian estimators are biased by the presence of the prior distribution, π(β), they often outperform other estimators by successfully trading off increased bias for lower variance. Theoretically, bias (i.e., $E_{D|\beta}[\tilde{\beta}-\beta]$ ) is of diminished relevance to Bayesians because its computation requires knowledge of the unobserved true value of β.

*Conditions When Endogeneity Bias Will Matter*

When an adaptive procedure is used to learn about respondent part-worths so that an informative design can be determined, and then these data are excluded from analysis, the likelihood for the endogenously determined covariates is no-longer redundant. Consider, for example, the likelihood for the data in the example above, but assume that $y_1$ and $y_2$ are not used to estimate the part-worths, β. The likelihood for this situation can be derived from equation (12) by integrating out the unused data:

$$\begin{aligned} \pi\left(x_3,y_3\,|\,\beta,\sigma^2\right) &= \int \pi\left(y_1,y_2,x_3,y_3\,|\,\beta,\sigma^2\right) d\left(y_1,y_2\right) \\ &= \pi_2\left(x_3\,|\,\beta,\sigma^2\right)\times\pi_3\left(y_3\,|\,x_3,\beta,\sigma^2\right) \end{aligned} \tag{24}$$

and the likelihood of $x_3$ is now a function of $\beta$ and $\sigma^2$ because $y_1$ and $y_2$ are not available. This occurs, for example, when self-explicated data are used for design purposes, but are not used to estimate $\beta$. If past observations are used to learn about model parameters, and design points are picked based on the acquired knowledge, then any analysis that proceeds without incorporating that knowledge is incomplete and the likelihood function mis-specified.

For the case of utility balance, consider what would happen if enough data were initially collected to make precise inferences about a respondent's part-worths. The utility balance algorithm would then generate candidate design points with responses close to zero for all choices. The likelihood $\pi_3$ in equation (12) would then have two modes, one at zero and one at the true parameter value. Without access to the earlier data or the original designer's knowledge, inferences about the part-worths would be biased downward toward zero.

More generally, analysis in marketing often relies on sequential analyses where respondents in earlier surveys are re-sampled. Examples include the management of customer lists in direct marketing, and panel studies where follow-up questions are posed to respondents. In these settings, individuals respond to offers and questions determined from previous analyses that are designed to be profitable and/or informative. The likelihood for stimuli selection (e.g., $\pi_2$) needs to be included in these analyses to avoid model mis-specification.

## 5. Concluding Comments

The likelihood principle is implicit to the Bayesian approach to statistics where the posterior distribution is derived from the prior distribution and the likelihood. Bayesian analysis conditions on the data to draw inferences about unobservable parameters in the analysis. In a conjoint analysis, it provides an answer to the question "Given the data at hand, what do I know about the part-worths?" Our view is that the answer to this question is more managerially relevant than the corresponding frequentist question concerning performance of an estimator across multiple datasets.

Sampling experiments are useful for understanding statistical properties such as bias when data have not yet been collected. They are not useful, however, once data are available for analysis. We demonstrate that the endogeneity bias identified by HT need not be relevant for analysis once data have been collected. This implies that a manager at a user firm (i.e., Manager B), who wants to make inferences about part-worths based on a specific data set, should not worry about endogeneity bias so long as all the data are present in the analysis.

Whether one should condition on the data and adhere to the likelihood principle, as in Bayesian analysis, or conduct sampling experiments, as in a frequentist analysis, is at the philosophical core of statistical inference. The Bayes-frequentist debate is sometimes dismissed as irrelevant because one obtains about the same answer in both cases. The issue of endogeneity bias raised by HT provides a counter-example to this view, where philosophical principles of inference play an important role in conducting analysis.

## References


Berger, James O. and Robert L. Wolpert (1984) The Likelihood Principle, Institute of Mathematical Statistics LectureNotes – Monograph Series.

Fisher, R.A.(1922) "On the Mathematical Foundations of Theoretical Statistics," *Philos. Transactions of the Royal Society*, London, Series A, 222, 309-368.

Hauser, John R. and Olivier Toubia (2005) "The Impact of Utility Balance and Endogeneity in Conjoint Analysis," *Marketing Science*, 24, 498-507.

Judge, George G., R. Carter Hill, William E. Griffiths, Helmut Lutkepohl, and Tsoung-Chao Lee (1988) Introduction to the Theory and Practice of Econometrics, 2nd Edition, John Wiley & Sons.

Rossi, Peter E., Greg M. Allenby and Robert McCulloch (2005) Bayesian Statistics and Marketing, John Wiley & Sons.

Figure 1
Expected Value of OLS Estimate Across 1000 Replications

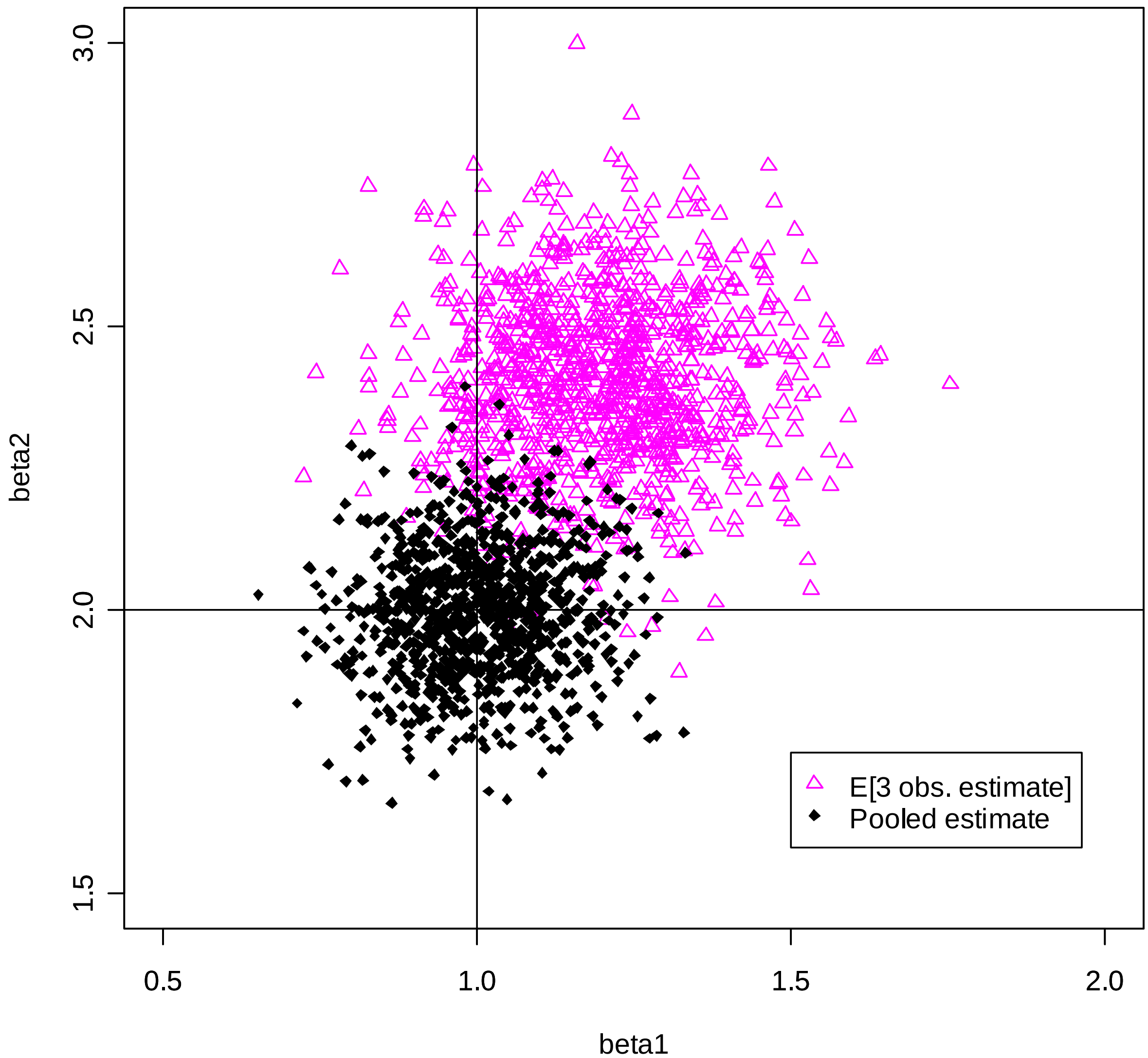

Figure 2
MCMC Draws for Exogenous and Endogenous Covariates
1000 Respondents