Survey Research Methods (2007)
Vol.1 , No.3 , pp. 145-154
ISSN 1864-3361
http://www.surveymethods.org

Model Based Survey Design Using Logits: Estimating Lost Statistical Power from Random Alternative Sampling

Jay Bhattacharya
Stanford Medical School

Baoping Shang
RAND, Santa Monica

McFadden's random alternative sampling conditional logit estimator permits researchers and survey designers to estimate a random utility choice model observing information about a subset of available alternatives. We quantify the extent to which a small sample size and a reduction in the number of sample alternatives lead to bias and loss of statistical power. The sample size must be small *and* choice probabilities must be weakly correlated with choice characteristics for there to be substantial bias and low power. Finally, we find that there is a sharply decreasing marginal gain from increasing the number of sampled alternatives. We provide an empirical example on the choice of health insurance plans that verifies our conclusions.
Keywords: Random Alternative Sampling Conditional Logit, Statistical power, Monte Carlo simulation

Introduction

Economists often study what choices people make among a fixed set of alternatives, each with its own set of pertinent characteristics such as price and quality. The conditional logit model is perhaps the most popular econometric model used to analyze how changes in characteristics affect choice probabilities. Given the well-known maximum utility interpretation of the conditional logit model, such popularity is justified.[1] However, the traditional conditional logit model poses substantial data requirements. For each alternative available to each consumer, the analyst must know all of its characteristics. In a survey context, collecting such complete data on all available alternatives may be too costly.

McFadden (1973, 1978) provides an appealing way to circumvent the data requirements of a traditional conditional logit model. Based only on information about the alternative chosen by the consumer and a random subset of the non-chosen alternatives, it is possible to consistently estimate model parameters. The cost of this approach is a loss of statistical power, which is unsurprising since less information is used by the econometrician in estimation. However, we are unaware of any study to date that has investigated how much statistical power is lost. Nor has there been any investigation of bias and statistical power of this random alternative sampling conditional logit (RASCL) estimator when sample sizes are small. Because it is common for economists to analyze consumer choices with less than infinite samples and with information about only a small subset of the available alternatives, these are important gaps in the literature. Using Monte Carlo methods, this paper fills these gaps. We find that in many common situations, the RASCL estimator loses surprisingly little statistical power relative to the full information conditional logit estimator.

To make this argument, we organize the paper as follows. We start, in Section 2, with a brief non-technical introduction to random alternative sampling and a consideration of the practical problems that arise in applying random alternative sampling to the problem of survey design. In Sections 3 and 4, we formally describe random alternative sampling and the design of our Monte Carlo exercise. In Section 5, we convey the results of the Monte Carlo exercise. In Section 6, we describe an empirical example on the choice of health plans by experts in managed care. In that section, we estimate the loss in statistical power that would have occurred had the surveyors (including one of the authors of this paper) used random alternative sampling. Finally, Section 7 concludes.

Random Alternative Sampling in Survey Research

At first sight, the random alternative sampling might seem an unlikely tool to use in the context of survey research. While it would certainly be less expensive to collect information about a random subset of alternatives than it would be to collect information about every alternative, nevertheless there are some serious practical difficulties with random alternative sampling. First, a researcher would need to know about the existence (though not any other details) of all the alternatives available to every survey respondent in order to randomize over those alternatives. Finding out about even the existence of the alternatives may be prohibitively expensive. In that case, of course, it would also be prohibitively expensive to

Contact information: Jay Bhattacharya, Center for Primary Care and Outcomes Research, Stanford Medical School, 117 Encina Commons, Stanford, CA 94305-6019 (jay@stanford.edu) and Baoping Shang, RAND, 1700 Main Street, Santa Monica, CA 90401 (bpshang@gmail.com)

[1] See McFadden (1973), Maddala (1983) or any recent advanced econometrics textbook for a discussion of this point.

collect information about all the alternatives. However, there are some important situations when collecting information about a subset of alternatives would not be expensive. There are strategies researchers can use to reduce these costs that we will discuss shortly. Second, if the set of alternatives is small and all survey respondents face a common set of alternatives, it will be likely that each alternative will have been picked by at least one respondent. In that case, there is no gain from random alternative sampling – to estimate a logit choice model every alternative must be investigated.

Perhaps the most important situation where random alternative sampling might be useful for survey designers is when the number of alternatives is much larger than the number of respondents. For example, Train (1987) use random alternative sampling to analyze consumer choice among local telephony plans where consumers pick from the same set of more than a million alternative plans. Unless there are special circumstances, it would be impossible to collect information about all those plans.[2]

A second situation where random alternative sampling might be useful is when each respondent faces a different set of alternatives. Even if the number of alternatives faced by each respondent is small, the total number of available alternatives may be quite large, making it prohibitively expensive to collect detailed information about all the alternatives. This is a very common situation faced by labor economists and health economists who are analyzing worker choice from a menu of fringe benefits such as health insurance. For example, Dowd (1995) and Chernew (1998) analyze the choice of health plans by consumers who, because they work for different employers, each must pick from a different set of plans. In those studies, the authors had available to them full information about the details of every available plan (deductible amounts, copayment rates, limits to coverage, and so on). If random alternative sampling had been used instead, the same logit choice parameters could have been estimated at lower cost.

One of the authors of this paper has been involved in two survey collection efforts where the cost of collecting information about a complete set of available alternatives was acutely felt. In the data collection effort that led to the paper by Studdert (2002), the largest costs of surveying involved contacting the employers of the surveyed workers to find out the characteristics of the available plans. Collecting detailed information about plan characteristics would have entailed extensive efforts on the part of contacted employers and hence to non-response. For that study, the authors decided instead to reduce the number of plan characteristics analyzed, which greatly limited the set of research question that could be addressed using these data. While the survey designers considered random alternative sampling, fears that using this technique would greatly compromise statistical power in this small scale study led to rejecting the technique. In this paper, we reanalyze these Studdert (2002) data using RASCL to examine whether these fears were well founded.

The Health Care for Communities (HCC) dataset also involved administering a survey to workers and then contacting the employers of these workers to learn details about the health plans available to workers – see Sturm (1999). In this second survey collection effort involving one of the authors of this paper, the survey designers decided explicitly to not ask employers about any of the alternative plans available to workers to limit the response burden placed on employers. As a result of this decision, while the HCC is useful for many other purposes, it cannot be used to analyze health plan choices by workers.

Though it is beyond the scope of this paper to comprehensively discuss practical strategies for implementing random alternative sampling in a survey setting – we are primarily interested here in the statistical power properties of the RASCL estimator against a full information estimator – we discuss a few such strategies here. As we noted above, an important practical problem is that the researcher much know what all the alternatives are in order to randomly pick from among them. In our experience, it is often possible to solve this problem by asking survey respondents which alternatives are available to them. This poses only a small additional burden on the respondent relative to asking about the characteristics of all the available alternatives (many of which the respondent will know nothing about). Then, a researcher who wants to analyze respondent choices can contact the providers of a random subset of these alternatives to garner the necessary information about alternative characteristics.

What if survey respondents cannot be relied upon to accurately report available alternatives? In that case, it is often possible to garner this information from other sources. In the examples we discuss above, employers know which health plans are available to their workers even if workers do not. Another important example is a study of what school characteristics cause parents to pick one school for their children over the others available to them. Parents may not be able to enumerate a complete list of public and private schools available in a city, but local city officials will certainly have this information. These same officials, though, may not have readily attainable information about detailed characteristics about the school in an easily attainable form. Researchers who want to use random alternative sampling can contact the city officials to get the universe of available schools and then contact the schools themselves to get detailed information about each sampled school.

We finish this section by noting one important caveat about using random alternative sampling in a survey context. Surveys collected using the technique require researchers to use logit methods to analyze respondent choice. The logit choice model works in this setting because it has built in it an assumption that relative choice probabilities do not change when irrelevant alternatives are removed. Though we do not

[2] In that study, the cost of collecting information about the alternative plans was actually nil. The telephone plans differed on the basis of a small number of characteristics, but millions of different combinations of these characteristics were available to consumers. The authors categorized the available plans by simply listing all the possible combinations of the characteristics. They justified the use of random alternative sampling by arguing that limitations in computing power made estimating a full information multinomial logit model prohibitively expensive. Such an argument would be less persuasive today than it was in 1987.

elaborate on the issue here, random alternative sampling may also help in situations where the independence of irrelevant alternatives (IIA) assumption is violated, such as in a nested logit model. It would be straightforward to organize a survey sampling scheme around a nested logit. The basic idea is that within nests, the IIA property applies, although across nests it does not. A survey designer would do random alternative sampling *within* each nest. The trade-off would be that the surveyor would have to sample a larger random set of alternatives – at least one choice within each nest. In some applications, this might be worth it. Other choice models, such as the multinomial probit (McFadden 1989) require full information about all the characteristics of all the alternatives, and so are not suitable for use with surveys where random alternative sampling is used.

Random Alternative Sampling Conditional Logit Estimator

We start the discussion with the full information conditional logit (FICL) estimator, mainly to introduce our notation. The setting is the well-known random utility model. Let i be the index over the consumers in the dataset, let M be the number of alternatives available to each consumer, and let j be an index over the alternatives,[3] X_{ij} is a vector that characterizes each alternative j for person i, β is a vector of parameters to be estimated, and ε_{ij} is an error term that represents the determinants of choice that are unobserved by the econometrician. Let consumer utility be a linear function of X_{ij}, and suppose that consumer i chooses by maximizing utility over all available alternatives. Suppose further that ε_{ij} follows an independent and identically distributed type I extreme value distribution, then the probability of choosing j' in a FICL model is given by:

$$P[j = j'] = \frac{\exp(\beta' X_{ij'})}{\sum_{j=1}^{M} \exp(\beta' X_{ij})} \tag{1}$$

With (1) in hand, it is easy to estimate β using maximum likelihood methods, as long as we observe consumer choices and X_{ij} for each alternative j.

We turn next to the RASCL estimator. The data requirements for the RASCL estimator include the characteristics of the chosen alternative, as well as the characteristics of a suitably picked random subset of the non-chosen alternatives.

Let A be the set of alternatives that we gather X_{ij} information about, and let j' continue to denote the chosen alternative (among the M alternatives). RASCL requires $j' \in A$. In addition, let j_k, $k = 1 \ldots K$, be a set of K other alternatives randomly picked by the econometrician to include in A. Thus, using RASCL, we need to collect information about X_{ij} from only $K + 1 \leq M$ alternatives. Let $B \supseteq A$ represent the set of all alternatives available to the consumer.

Following McFadden (1978), we define $\pi(A|j)$ as the conditional probability of constructing a particular set A, given that the chosen alternative is j. We calculate the joint probability of the consumer choosing a particular choice, j', from A, and the econometrician constructing a particular random set of alternatives A:

$$\pi(j', A) = \pi(j'|A)P[A] = \pi(A|j')P[j = j'] \tag{2}$$

Since A consists of information about the chosen alternative and K non-chosen alternatives and since the chosen alternative is always in A by definition, $\pi(A|j) = 0$ for $j \notin A$. By Bayes' rule, the conditional probability of j' being chosen, given the randomly picked subset of $K + 1$ alternatives in A, is:

$$\pi(j'|A) = \frac{\pi(A|j')P[j = j']}{\pi(A|j')P[j = j'] + \sum_{k=1}^{K} \pi(A|j_k)P[j = j_k]} \tag{3}$$

McFadden (1978) shows that any predetermined algorithm to randomly pick K non-chosen alternatives to construct A yields consistent estimates of β. For instance, suppose that $\pi(A|j) > 0$ for all j.[4] Plugging (1) into (3) yields:

$$\pi(j'|A) = \frac{\exp(\beta' X_{ij'} + \ln(\pi(A|j')))}{\exp(\beta' X_{ij'} + \ln(\pi(A|j'))) + \sum_{k=1}^{K} \exp(\beta' X_{ij_k} + \ln(\pi(A|j_k)))} \tag{4}$$

In this paper, we impose an additional assumption (called the uniform conditioning property) to simplify the exposition:[5]

$$\pi(A|j') = \pi(A|j) \; \forall j \in A \tag{5}$$

We can derive the probability of picking j' out of **A** by plugging (1) and (5) into (3) to obtain:

$$\pi(j'|A) = \frac{\exp(\beta' X_{ij'})}{\exp(\beta' X_{ij'}) + \sum_{k=1}^{K} \exp(\beta' X_{ij_k})} \tag{6}$$

[3] To streamline the discussion, we assume here that each consumer faces the same number of alternatives. This assumption can be relaxed and the conditional logit model extended to handle varying numbers of choices for the consumers. See Amemiya (1985), ch. 9 for a discussion of this case.

[4] This assumption is known as the positive conditioning property (see Haab 2002 for a discussion of this nomenclature).

[5] In a previous version of this paper, we estimated the statistical power loss from the RASCL estimator in which we assume only the positive conditioning property - equation (4). For those estimates, we assumed the probability of selecting a plan into **A** was proportional to the popularity of the plan in the population. The results from those Monte Carlo experiments showed qualitatively similar power loss when compared with experiments in which we assume the uniform conditioning property. As one might expect, imposing the additional uniform conditioning property assumption tends to raise statistical power. Results with the positive conditioning property are available upon request from the authors.

A simple algorithm to implement the uniform conditioning property involves randomly picking K non-chosen alternatives, j_k, $k = 1 \ldots K$, to include in **A**. Given the choice by the consumer of alternative j', all of the other alternatives are equally likely to be picked for inclusion in A.[6]

Using (6), it is easy to estimate β by maximum likelihood methods. Since equation (6) contains information only on alternatives j' and the randomly picked (by the econometrician) K non-chosen alternatives, we do not need information about any other alternatives to obtain consistent estimates of β. Note that since M does not enter (6), the asymptotic variance of β will depend only on K (the number of non-chosen alternatives about which the econometrician observes attributes X_{ij}), not on M (the number of alternatives available to person i). However, relative to the FICL model in equation (1), the estimate of β using (6) should be less efficient (since it is based on less information), though from econometric theory alone it is not clear how inefficient it will be in finite samples.

Monte Carlo Evaluation

While it is encouraging to know that we can estimate the relationship between characteristics and choice with information from a random subset of the alternatives, it is not the end of the story. Finding a consistent estimate for β would be useless if the estimate was so inefficient that we would not be able to perform powerful hypothesis tests. It seems intuitively true that we will lose some efficiency and power using RASCAL since we estimate β on a subset of the information needed for FICL. The main aim of the Monte Carlo experiment is to characterize how much efficiency is lost under a wide variety of conditions.

The first step in conducting a Monte Carlo experiment is to specify the data generating process. In this case, the process we use is suggested by the random utility interpretation of the conditional logit model. To focus the study, we suppose that there are two covariates, X_{ij} and Z_{ij}, determining choice.[7] Suppose there are N individuals making choices among M alternatives, which are characterized by X_{ij} and Z_{ij}:

$$\begin{aligned} X_{ij} &\sim \text{Uniform}[0,1] \\ Z^*_{ij} &\sim Uniform[0,1] \\ Z_{ij} &= \begin{cases} 0 \text{ if } Z^*_{ij} \leq 0.75 \\ 1 \text{ if } Z^*_{ij} > 0.75 \end{cases} \end{aligned} \tag{7}$$

While everyone draws from the same distribution of characteristics, one consequence of (7) is that no two people choose from the same set of alternatives. Everyone in the dataset, however, does share the same values of coefficients (as is the case in any conditional logit estimation). The differences in the utility functions across people and alternatives arise from the idiosyncratic portion of utility, ε_{ij}, which is drawn from a Type I Extreme Value distribution that is independent of the X_{ij} and Z_{ij} draw. To generate the ε_{ij} draws, we first draw a number, z, from a Uniform [0,1] distribution. Then we set:

$$\varepsilon_{ij} = F^{-1}[z] = -\ln(-\ln z) \tag{8}$$

where F is the cumulative distribution function for a type I extreme value random variable.

Using the random draws on X_{ij}, Z_{ij} and ε_{ij}, we calculate each person's utility from each alternative:

$$U_{ij} = X_{ij}\beta + Z_{ij}\alpha + \varepsilon_{ij} \quad i = 1 \ldots N, j = 1 \ldots M \tag{9}$$

As usual, we designate the alternative with the maximum utility as j', the chosen alternative. From the design of the experiment, we have a value for α and β.

Next, we throw away all the information that we would normally not have as analysts. Thus, we estimate the FICL model using information on X_{ij} and Z_{ij} for each alternative and j' – we ignore ε_{ij} and U_{ij}, except as revealed through choice. We next calculate the RASCL estimators using the same dataset as in the FICL analysis, except we randomly pick K of the non-chosen alternatives (j_k) for each person in the dataset. Then, using only information on j' and the K non-chosen alternatives, we estimate α and β by maximizing the likelihood function implied by (6). In the RASCL case, each person in the dataset faces the same set of alternatives as they do in the FICL case. The only difference is that when estimating RASCL, we ignore information on the non-picked, non-chosen alternatives, while when estimating FICL we use these data. Thus, any differences in the precision of the α and β estimates in these analyses should arise from not using the full information set typically available in a conditional logit analysis.

With this data generating process, there are no problems caused by misspecification of the econometric model. In fact, the only correctly specified econometric model we could use, given the data generating process we assume, is a conditional logit model (and modifications of it). Any error in estimating α and β using equation (1) is due entirely to the error introduced by sampling. Any increase in the standard error of the estimate of β using equations (6) must be due to the additional sampling error introduced by random sampling from the non-chosen alternatives.

To limit the influence of random sampling variation, we repeat the experiment over 1.000 trials. That is, we generate 1.000 random datasets for each combination of parameters we evaluate, and estimate both models.

Given the data generating process, there are only a few parameters that affect the outcome of the experiment. In this study, α takes a fixed value of 0.1, and we vary the number of individuals in the dataset (N), the number of choices each hypothetical person faces (M), number of selected non-chosen alternatives (K) and the strength of the association between the covariate X_{ij} and the probability of choice (β). We allow

[6] That is, equal probability sampling over the alternatives without replacement.

[7] The results (not shown) are similar if only one covariate is included. These results are available on request from the authors.

β to take 5 values: 0.1, 0.2, 0.3, 0.5 and 0.9.[8] For these values of β, the covariates explain 0.16%, 0.32%, 0.57%, 1.36%, and 4.05% of the variance in U. We allow M to be 3, 4, 5, 10, 20 and 50; and we allow K to be 1, 2, 3, 4, 9, 19 and 49.[9] Our strategy, for any particular fixed combination of M, K and β, is to estimate the minimum sample size needed to achieve statistical power of 80% and 50%. We use a binary search algorithm to locate the smallest sample size for a given statistical power.

Results

In this section, we present the results of our experiment. Sections 5.1 and 5.2 discuss the small sample bias and statistical power of the RASCL model.

Small Sample Bias

For both models – FICL and RASCL – we use a maximum likelihood estimator of β. For each of these estimators, there are standard asymptotic results that assure us of the consistency (and efficiency relative to the information set) of the β estimates. What is unclear from these asymptotic results is how applicable they are in small samples. It is certainly possible that, despite the theorems about consistency, it takes a larger value of N for the partial information estimators to attain a small bias in the estimated β than it does for the full information estimators.

Table 1shows the small sample bias of sample size 50 and 500 for various combinations of M, K, and β. The first column is the total number of alternatives (M), and the second column is the number of non-chosen alternatives included in the estimation (K). For example, when $N = 50$ and $\beta = 0.1$, the small sample bias of the RASCL estimator is -0.0281 (the third column) or -28.1% of the true value of β. In that case, the FICL model is also similarly biased – the row with $M = 3$ and $K = 2$ corresponds to the FICL model. Generally, our results indicate that small sample bias varies with neither M nor K, though it decreases with sample size and β. As sample size increases from 50 to 500, the small sample bias, averaged across all combinations of M and K, decreases from 14.80% to 3.81% when $\beta = 0.1$; from 3.53% to 0.62% when $\beta = 0.5$; and from 3.50% to 0.46% when $\beta = 0.9$. The small sample bias can be as high as 34% when $N = 50$ and $\beta = 0.1$. As N increases to 500 or β increases to 0.5 or 0.9, the small sample bias is well within 10% in most of the scenarios.

However, since these numbers are averaged over 1.000 trials, they might be misleading – in real life, we only have access to a single trial. By averaging over so many trials, we allow the cases of positive bias to cancel the cases of negative bias. This concern will be addressed in the next section by examining the estimation precision or estimation efficiency in β over the trials given the experimental parameters.

Statistical Power

Depending on the experimental parameters, the relative loss in statistical power[10] moving from full information model to partial information model with $K = 1$ ranges from 14% to 38% and this loss increases with M. This is consistent with the intuition that when M is larger and K held fixed, more information is "thrown away" by the RASCL estimator. For example, when $\beta = 0.1$ and $N = 4.000$, the absolute statistical power for RASCL ($M = 4$, $K = 1$) is 0.511 and for the FICL model ($M = 4$ and $K = 3$) is 0.689; the relative statistical power of the RASCL model is 0.742 which is a loss of 25.8% relative to the statistical power of FICL.

Figure 1 shows how statistical power compares across different combinations of M and K, holding fixed β. For all sample sizes, the absolute statistical power does not seem to change with K. For example, M4K1[11] and M50K1 have almost identical statistical power over the entire range of sample size N. This should not be surprising since M does not enter the RASCL likelihood function at all – see (6)– so (holding K fixed) statistical power should not vary with M.. Conditional on sample size, however, FICL with $M = 50$ (M50K49) has much larger absolute statistical power than FICL with $M = 4$ (M4K3). These results demonstrate that, conditional on sample size and β, absolute statistical power increases with K. The larger power loss associated with larger M for the RASCL estimator is due to more statistical power in the corresponding full information model, and does not imply that, for the same K, survey designs with larger M need larger sample sizes to achieve the same levels of statistical power than those with smaller M. Large M only means there is more information available and if we do not fully use it, the relative loss or opportunity cost will be high. We can increase statistical power by two ways, one is to increase sample size and the other is to increase the number of non-chosen alternatives included. The gain in statistical power from increasing sample size does not depend on M, but larger M offers more room to push the boundary up.

Table 2 shows decreasing marginal gain in statistical power from increasing the number of non-chosen alternatives included. The pattern does not seem to vary with sample size. It is also true that, conditional on statistical power for the full information model and M, relative power loss does not depend on β (for the range of β's we examine).[12]

To better assess the trade-offs between sample size (N) and number of non-chosen alternatives included in the estimation (K), we compare the minimum sample sizes needed to achieve certain levels of statistical power between the partial

[8] The range of β values we consider in this study spans only the range where there is a weak correlation between the explanatory variable and choice. In experiments (not reported here) where X and U are strongly correlated, both estimators exhibit high levels of statistical power even with small sample sizes.

[9] Of course, we exclude cases where $K > M$.

[10] Statistical power is defined as the percentage of trials that reject a hypothesis of zero effect of Z_{ij} on choice (that is, a hypothesis that $\beta = 0$ with the alternative $\beta > 0$).

[11] Total number of alternatives is 4 and number of non-chosen alternatives included in the estimation is 1.

[12] We use relative power loss here because we cannot set the increment in sample size too small due to computational burden, and therefore are unable to hold statistical power exactly the same across specifications.

information models and the full information model. Necessary sample size to achieve a fixed level of power is a more intuitive measure of the costs of adopting a RASCL estimator over a FICL estimator.

Table 2 shows the minimum sample sizes necessary to achieve 50% and 80% statistical power for each combination of β, M and K. These minimum sample sizes decrease with β. It is relatively easier to achieve a desired level of statistical power with a relatively small sample size when X_{ij} is strongly correlated with choice probabilities. The results also show that the relative increase in required sample size increases with M and does not seem to depend on β. This is consistent with our finding that there is more power loss associated with larger M when moving from full information model to partial information model. The relative increase in sample size moving from the case $K = M - 1$ (the FICL estimator) to $K = 1$ (the least information model) range from 20% to 130% depending on the specification. With additional information on the costs of data collection, the results can be used to assess the trade-off between sample size (N) and number of non-chosen alternatives included in the estimation (K) to reach the most cost-effective survey design (minimal costs to achieve a given level of statistical power).

Health Plan Choice: An Empirical Example

For our empirical example, we reanalyze data from the Studdert (2002) study of the health plan choices of managed care experts. Based upon computerized literature searches using Medline, the investigators identified the set of 20 universities in the United States that had produced the greatest number of published research papers on the topic of managed care. From these same searches, the investigators identified the set of researchers at each university responsible for those studies. The researchers also randomly chose a control set of law, math, and philosophy professors at the same university who had never published on managed care.

The researchers contacted the human resources division within each university to find out the set of health insurance plan choices available to university employees at each university. Based on these contacts, the investigators found key details about each health plan choice, including the plan type – whether the plan offers a health maintenance organization (HMO), preferred provider organization (PPO), point of service plan (POS), a fee-for-service plan (FFS), or catastrophic plan – and the premium charged for choosing a plan (which typically differs for families and single people).

Finally, Studdert and his colleagues contacted by email each managed care expert and control professor to find out which plan each had selected. Studdert (2002) provide more detail about this dataset. Their main finding was that, with the exception of junior social scientists, most managed care experts were less likely than controls choose a managed care plan. By far, the most expensive part of the survey involved getting information from the human resources divisions about the features of each plan at each university. Here, we reanalyze these data to find the effect of premiums on plan choice, and we ask what effect using a RASCL estimator rather than a FIML estimator would have on the findings.

After dropping individuals with missing information, the sample contains 416 individuals. An individual can face as many as five types of health plans: HMO, PPO, POS, FFS, and Catastrophic. Among the 416 individuals, 166 have two choices, 103 have three choices and 147 have four choices. The plan-level characteristics that we analyze are premiums for individual coverage and for family coverage.

In FICL estimation, we use premium information on all the choices each individual was facing. In RASCL estimation, for individuals with more than two choices, we only include the chosen type and a randomly picked type among the non-chosen types. In our specification, we interact premiums with whether an individual is in the control group or in the experimental group. Our results are shown in Table 1.

We find higher family premiums, but not premiums for single people, reduce the likelihood of plan choice. Managed care experts and control professors do not differ statistically in their demand elasticity. The RASCL estimates are typically close to the FICL estimates, though as one might expect, the standard error estimates from the RASCL estimator are slightly larger than those from the FICL estimator. Consequently, the RASCL and FICL estimates produce similar results for statistical inference.

To further evaluate how the standard error estimates from RASCL compare with the standard error estimates from FICL, we drew 10.000 bootstrap samples with replacement and repeated the FICL and RASCL estimation. For every bootstrap replication and for every coefficient estimate, the 95% confidence interval from RASCL overlaps with the 95% confidence interval from FICL.

Conclusion

If data were costless to collect, it would be ideal to have information about all available choices to analyze the effect of choice characteristics on choice probabilities. Even though McFadden's random alternative sampling conditional logit (RASCL) estimator permits researchers and survey designers to limit these potentially onerous data requirements, there are costs. Theoretically at least, having information about only a few alternatives can lead to loss of precision and small sample bias. Of course, data collection is not costless and researchers often face difficult choices about resource allocation in the conduct of research. Our results provide a quantification of this trade-off.

We find that small sample bias and loss of statistical power become important concerns with the use of the RASCL estimator only when *both* sample size is small *and* choice probabilities and choice characteristics are only weakly correlated. Furthermore, we find that the performance of the RASCL estimator does not degrade when the number of alternatives rises. For example, if a researcher solicits information about choice characteristics about only two alternatives, one might think that more information is thrown away when fifty alternatives are available than when there are only three. This is true but only because there is more information contained in

Table 3: Health Plan Choice Parameter Estimates

Covariate	FICL	RASCL
Individual Premium	0.0043	0.0037
	(0.0057)	(0.0057)
Individual Premium*Experimental Group	0.0023	0.0038
	(0.0073)	(0.0075)
Family Premium	-0.0040^{*}	-0.0039^{*}
	(0.0023)	(0.0023)
Family Premium*Experimental Group	−0.0003	−0.0010
	(0.0030)	(0.0030)

Standard errors in parentheses.
*indicates the coefficient is statistically significant at $p < 0.10$.

the full information model when the number of alternatives is large (holding sample size fixed). Our primary finding is that there is a sharply decreasing marginal gain from increasing the number of sampled alternatives.

Applying random alternative sampling may thus be an attractive option for survey designers who are interested in analyzing the determinants of a choice when the costs of collecting information about all alternatives is prohibitively expensive (and when that choice can be fruitfully modeled with a logit in the tradition of McFadden (1978)).

Acknowledgements

We thank Roland Sturm and Naihua Duan for their comments and encouragement. We also thank the editors of this journal and the referees for very helpful suggestions. Bhattacharya's work on this project was funded by the National Institute of Mental Health. Shang's work was funded by the Robert Woods Johnson Foundation. All errors are ours.

References

Amemiya, T. (1985). *Advanced econometrics*. Cambridge, MA: Harvard University Press.

Chernew, M., & Scanlon, D. P. (1998). Health plan report cards and insurance choice. *Inquiry*, *35*(1), 9-22.

Dowd, B., & Feldman, R. (1995). Premium elasticities of health plan choice. *Inquiry*, *31*(4), 438-444.

Haab, T., & McConnell, K. E. (2002). *Valuing environmental and natural resources: The econometrics of non-market valuation*. Northampton, MA: Edward Elgar.

Maddala, G. S. (1983). *Limited-dependent and qualitative variables in econometrics*. Cambridge, UK: Cambridge University Press.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in econometrics.* Academic Press.

McFadden, D. (1978). Modeling the choice of residential location. In A. Kar-lquist, L. Lundqvist, F. Snickars, & J. Weibull (Eds.), *Spatial interaction theory and planning models* (p. 75-96). Amsterdam: North-Holland.

McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica*, *57*(6), 995-1026.

Studdert, D., Bhattacharya, J., Warren, B., Schoenbaum, M., & Escarce, J. J. (2002). Health insurance choices of managed care experts. *Medical Care*, *40*(6), 375-386.

Sturm, R., Gresenz, C., Sherbourne, C., Bhattacharya, J., Farley, D., Young, A. S., et al. (1999). The design of healthcare for communities: A study of health care delivery for alcohol, drug abuse, and mental health conditions. *Inquiry*, *36*(2), 221-233.

Train, K. E., McFadden, D. L., & Ben-Akiva, M. (1987). The demand for local telephone service: A fully discrete model of residential calling patterns and service choices. *Rand Journal of Economics*, *18*(1), 109-123.

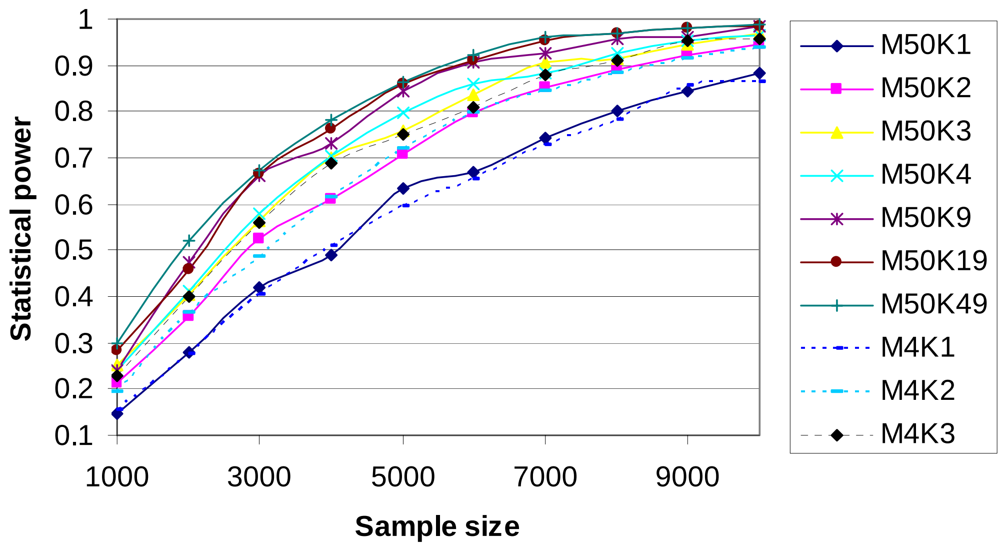
Table 1: Small Sample Bias: Conditional on Same Sample Size for Each

		N=50						N=500					
		$\beta = 0.1$ Covariates explain 0.16% of Var(U)		$\beta = 0.5$ Covariates explain 1.36% of Var(U)		$\beta = 0.9$ Covariates explain 4.05% of Var(U)		$\beta = 0.1$ Covariates explain 0.16% of Var(U)		$\beta = 0.5$ Covariates explain 1.36% of Var(U)		$\beta = 0.9$ Covariates explain 4.05% of Var(U)	
M	K	Bias	% Bias	Bias	% Bias	Bias	% Bias	Bias	% Bias	Bias	% Bias	Bias	% Bias
3	1	−0.0281	-28.05	0.0282	5.64	0.1445	16.06	0.0060	5.99	0.0083	1.66	0.0117	1.30
3	2	−0.0285	-28.54	−0.0082	-1.63	0.0228	2.53	−0.0031	-3.07	0.0049	0.99	0.0024	0.26
4	1	0.0245	24.48	0.0400	7.99	0.0683	7.59	−0.0015	-1.45	0.0056	1.12	0.0048	0.54
4	2	0.0033	3.28	0.0092	1.84	0.0384	4.26	0.0028	2.82	0.0040	0.80	0.0134	1.49
4	3	0.0076	7.59	−0.0058	-1.16	0.0209	2.32	−0.0018	-1.82	0.0023	0.46	0.0038	0.42
5	1	0.0217	21.68	0.0723	14.46	0.0859	9.54	−0.0015	-1.52	0.0016	0.33	0.0115	1.28
5	2	0.0057	5.67	−0.0055	-1.11	0.0216	2.40	0.0033	3.34	−0.0035	-0.70	−0.0066	-0.74
5	3	−0.0062	-6.19	0.0066	1.32	0.0196	2.18	0.0041	4.13	−0.0039	-0.78	−0.0023	-0.26
5	4	−0.0103	-10.32	−0.0172	-3.44	−0.0003	-0.03	0.0019	1.89	0.0002	0.04	−0.0023	-0.26
10	1	0.0152	15.23	0.0362	7.25	0.0813	9.03	0.0027	2.70	0.0013	0.27	−0.0022	-0.24
10	2	−0.0255	-25.50	0.0078	1.57	0.0549	6.11	−0.0043	-4.33	−0.0001	-0.02	0.0029	0.32
10	3	−0.0071	-7.10	−0.0146	-2.91	0.0025	0.27	−0.0048	-4.81	0.0016	0.31	0.0030	0.33
10	4	−0.0323	-32.33	−0.0100	-2.01	0.0179	1.99	−0.0076	-7.64	0.0014	0.27	0.0008	0.09
10	9	−0.0187	-18.72	0.0036	0.71	0.0042	0.47	−0.0089	-8.87	−0.0028	-0.57	0.0005	0.05
20	1	0.0027	2.67	0.0542	10.84	0.0357	3.97	−0.0067	-6.68	0.0044	0.88	0.0090	1.00
20	2	0.0027	2.72	0.0095	1.89	0.0307	3.41	−0.0065	-6.55	−0.0008	-0.16	0.0045	0.50
20	3	0.0106	10.58	0.0113	2.27	0.0181	2.01	−0.0007	-0.74	−0.0012	-0.24	−0.0043	-0.48
20	4	−0.0149	-14.93	−0.0114	-2.28	0.0054	0.60	−0.0034	-3.36	−0.0037	-0.75	0.0005	0.05
20	9	−0.0045	-4.53	0.0046	0.91	−0.0001	-0.01	−0.0045	-4.48	0.0047	0.94	0.0023	0.26
20	19	−0.0163	-16.29	−0.0138	-2.77	−0.0165	-1.83	−0.0003	-0.28	−0.0021	-0.43	0.0018	0.20
50	1	−0.0038	-3.79	0.0293	5.86	0.0656	7.29	0.0016	1.61	0.0064	1.27	−0.0030	-0.34
50	2	−0.0166	-16.58	0.0165	3.29	0.0382	4.24	−0.0018	-1.79	0.0052	1.03	0.0017	0.19
50	3	−0.0016	-1.63	0.0219	4.39	0.0305	3.39	−0.0059	-5.92	0.0013	0.25	0.0094	1.05
50	4	−0.0342	-34.18	−0.0129	-2.58	0.0151	1.68	−0.0022	-2.17	0.0036	0.72	0.0037	0.42
50	9	−0.0122	-12.25	−0.0150	-2.99	0.0034	0.38	−0.0088	-8.81	−0.0024	-0.49	0.0002	0.02
50	19	−0.0323	-32.32	−0.0079	-1.58	0.0036	0.40	−0.0055	-5.47	−0.0033	-0.66	0.0005	0.05
50	49	−0.0125	-12.52	0.0031	0.62	−0.0039	-0.43	0.0006	0.64	−0.0033	-0.67	0.0014	0.16

Table 2: Trade-offs Between Number of Non-chosen Alternative Included (K) and Sample Size (N) in Achieving Given Levels of Statistical Power

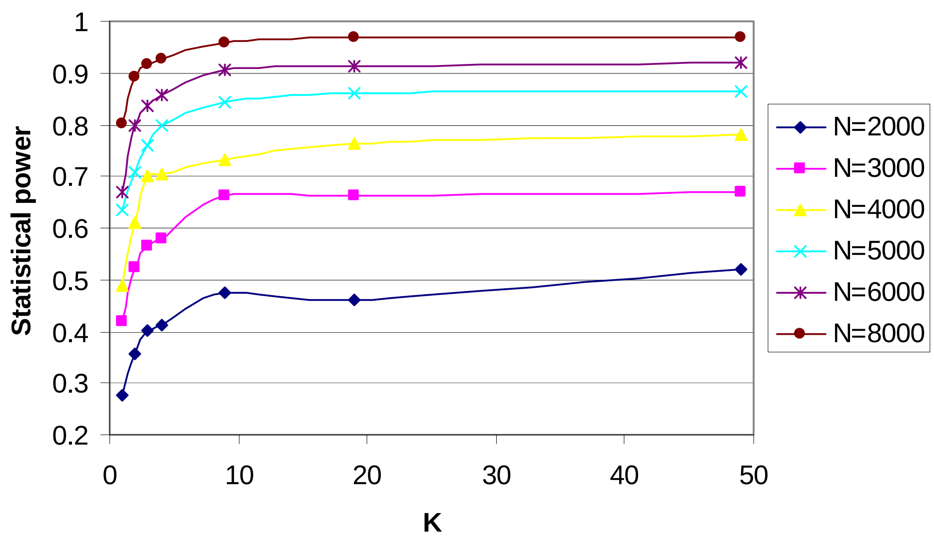
		$\beta = 0.1$ Covariates explain 0.16% of Var(U)				$\beta = 0.2$ Covariates explain 0.32% of Var(U)				$\beta = 0.3$ Covariates explain 0.57% of Var(U)				$\beta = 0.5$ Covariates explain 1.36% of Var(U)			
		SP= 80%		SP= 50%		SP= 80%		SP= 50%		SP= 80%		SP= 50%		SP= 80%		SP= 50%	
M	K	N	Relative Increase	N	Relative Increase	N	Relative Increase	N	Relative Increase	N	Relative Increase	N	Relative Increase	N	Relative Increase	N	Relative Increase
3	1	8562	1.356	3995	1.215	2018	1.299	965	1.359	904	1.361	454	1.397	318	1.347	160	1.404
3	2	6312	1.000	3289	1.000	1554	1.000	710	1.000	664	1.000	325	1.000	236	1.000	114	1.000
4	1	8562	1.621	4022	1.509	2073	1.525	1045	1.591	889	1.492	458	1.678	325	1.484	160	1.553
4	2	5718	1.083	3120	1.171	1471	1.082	728	1.108	652	1.094	325	1.190	250	1.142	118	1.146
4	3	5281	1.000	2665	1.000	1359	1.000	657	1.000	596	1.000	273	1.000	219	1.000	103	1.000
5	1	7906	1.496	3940	1.622	2029	1.561	957	1.630	893	1.633	392	1.420	325	1.738	154	1.638
5	2	6128	1.160	2984	1.228	1493	1.148	755	1.286	704	1.287	325	1.178	239	1.278	113	1.202
5	3	5581	1.056	2601	1.071	1340	1.031	640	1.090	573	1.048	280	1.014	195	1.043	104	1.106
5	4	5283	1.000	2429	1.000	1300	1.000	587	1.000	547	1.000	276	1.000	187	1.000	94	1.000
10	1	8562	1.762	3640	1.677	2062	1.858	968	1.826	882	1.917	466	1.966	314	1.975	164	2.103
10	2	6210	1.278	3093	1.425	1506	1.357	725	1.368	678	1.474	332	1.401	235	1.478	115	1.474
10	3	5390	1.109	2628	1.211	1318	1.187	651	1.228	541	1.176	291	1.228	199	1.252	99	1.269
10	4	5171	1.064	2382	1.097	1318	1.187	640	1.208	541	1.176	255	1.076	195	1.226	89	1.141
10	9	4859	1.000	2171	1.000	1110	1.000	530	1.000	460	1.000	237	1.000	159	1.000	78	1.000
20	1	8343	1.880	3968	1.828	2062	1.883	979	2.057	874	1.991	407	1.893	322	2.025	154	2.110
20	2	6320	1.424	2997	1.380	1493	1.363	684	1.437	689	1.569	325	1.512	210	1.321	111	1.521
20	3	5390	1.215	2628	1.211	1362	1.244	640	1.345	552	1.257	284	1.321	199	1.252	89	1.219
20	4	4924	1.110	2327	1.072	1231	1.124	590	1.239	541	1.232	262	1.219	180	1.132	91	1.247
20	9	4528	1.021	2190	1.009	1121	1.024	531	1.116	458	1.043	240	1.116	162	1.019	75	1.027
20	19	4437	1.000	2171	1.000	1095	1.000	476	1.000	439	1.000	215	1.000	159	1.000	73	1.000
50	1	8343	2.054	4241	2.222	2073	2.100	1023	2.177	848	1.932	466	2.273	310	2.138	153	2.186
50	2	6265	1.542	2984	1.563	1515	1.535	706	1.502	637	1.451	306	1.493	221	1.524	102	1.457
50	3	5527	1.361	2601	1.362	1318	1.335	640	1.362	563	1.282	269	1.312	184	1.269	99	1.414
50	4	4843	1.192	2546	1.334	1231	1.247	640	1.362	537	1.223	266	1.298	180	1.241	89	1.271
50	9	4638	1.142	2245	1.176	1143	1.158	525	1.117	462	1.052	232	1.132	162	1.117	75	1.071
50	19	4351	1.071	2163	1.133	1001	1.014	492	1.047	444	1.011	203	0.990	147	1.014	73	1.043
50	49	4062	1.000	1909	1.000	987	1.000	470	1.000	439	1.000	205	1.000	145	1.000	70	1.000

Figure 1. Statistical Power and the Number of non-Chosen Plans



Note: $\beta = 0.1$ (covariates explain 0.16% of Var(U)) in this graph. The graphs are qualitatively similar for other values of β.

Figure 2. Statistical Power Loss and Sample Size



Note: $\beta = 0.1$ (covariates explain 0.16% of Var(U)) and M=50 in this graph. The graph is qualitatively similar for other values of β and M.