ELSEVIER

Journal of Econometrics 89 (1999) 131–157

JOURNAL OF Econometrics

A model of health plan choice: Inferring preferences and perceptions from a combination of revealed preference and attitudinal data

Katherine M. Harris[a], Michael P. Keane[b,*]

[a] *Institute for Health Services Research, Tulane University Medical Center, New Orleans, LA 70112, USA*
[b] *Department of Economics, University of Minnesota, Minneapolis, MN 55455, USA*

Abstract

In this paper, we model the elderly's choice among health plans using data from a 1988 study of the Minneapolis–St. Paul Medicare health plan market. We show how attitudinal data can be combined with revealed preference data to provide more reliable estimates of consumers' preferences for and perceptions of the attributes of choice alternatives. We develop an extended heterogeneous logit model that incorporates information about attribute importance and allows for heterogeneity in tastes for observed and unobserved attributes. Our results indicate that the inclusion of attitudinal information produces a substantial improvement in model fit and in the substantive interpretation of estimated parameters. © 1999 Elsevier Science S.A. All rights reserved.

JEL classification: C35; I10

Keywords: Discrete choice; Attitudinal data; Simulation estimation; Multinomial logit

1. Introduction

In this paper, we model the health plan choices of elderly Medicare beneficiaries in the Minneapolis–St. Paul area, using revealed preference and attitudinal data collected by the Health Care Financing Administration in 1988. Medicare, the federal health insurance program for the elderly, leaves many services

* Corresponding author. E-mail: mkeane@atlas.socsci.umn.edu

PII: S0304-4076(98)00058-X

uncovered and requires substantial cost sharing in the form of deductibles and coinsurance. As a result, many elderly individuals supplement Medicare benefits by purchasing fee-for-service supplements, known as 'medigap' plans, or by joining health maintenance organizations (HMOs). The 1988 Twin Cities Medicare data contained information on the set of health plan options in which the respondents were enrolled – Medicare alone or Medicare plus one of the medigap or HMO options. Thus, we observe the elderly's revealed preference for health plan combinations with different features.

The Twin Cities Medicare data also contain information on how the elderly valued various features of health care plans. Essentially, they were asked to provide ratings of how important each of several plan features were to them in deciding on a plan. Our main objective in this paper is to show how such attitudinal data can fruitfully be combined with revealed preference data to provide more reliable estimates of consumers' preferences for the attributes of choice alternatives.

A substantial literature exists on the use of revealed preference data to infer how people value attributes of choice options. Much of this research has made use of conditional logit models (see, e.g., McFadden, 1974; Hensher and Johnson, 1980; Ben-Akiva and Lerman, 1985), and focused on estimating how people value observable attributes of choice alternatives. In particular, Feldman et al. (1989) used this approach to model choice among employer-sponsored health plans.

In the marketing literature, however, there has been a recognition that, in many contexts, crucial attributes of alternatives may be unobserved and/or inherently difficult to measure. This has led to the development of so-called 'market-mapping' methods (see, e.g., Elrod, 1988, Elrod, 1991,), in which positions of choice alternatives along unobserved attribute dimensions are inferred from the error structure in choice models. The basic idea behind market-mapping methods is that similarity (dissimilarity) of choice alternatives in terms of unmeasured attributes will show up as positive (negative) correlation of the errors associated with those alternatives. Such correlations are identified from switching behavior. For example, in a 3 alternative case, suppose that if the price of plan 1 is increased the market share of plan 2 increases substantially while that of 3 is little changed. Such a violation of the independence of irrelevant alternatives (IIA) property of the conditional logit model may arise if 1 and 2 are very similar in terms of an unobserved attribute for which consumers have heterogeneous tastes, while 3 is quite dissimilar.

A problem with market-mapping methods is that assignment of names to the unobserved attributes relies purely on subjective judgment. In some contexts, such as predicting the change in market shares that will result from a price change, this is irrelevant (i.e., the names assigned to the unobserved attributes have no bearing on how the error structure influences switching probabilities). In contexts where learning how consumers value the unobserved attributes is itself of key interest, however, proper identification of the unobserved attributes is crucial.

We argue that attitudinal data can be useful in identifying the unobserved attributes that lead to IIA violations. For example, suppose that consumers' perceptions of the quality of care provided by a health plan are difficult to measure, but that consumers are asked how much they value quality of care. If, in the above 3 alternative scenario, consumers who choose plan 1 or 2 tend to be those who say they place a great weight on quality in choosing a plan, while those who choose 3 tend to place little weight on quality, this would be evidence that the unobserved attribute leading to the IIA violation is perceived plan quality. It is exactly this type of information that we provide a framework for exploiting.[1]

Our framework is based on a heterogeneous conditional logit model that allows for the existence of unobserved attributes of alternatives, and for consumer heterogeneity in tastes for both observed and unobserved attributes. We assume taste parameters are normally distributed in the population. Most importantly, we extend the heterogeneous logit by allowing the distribution of taste parameters to differ conditional on consumers' stated preferences for attributes. This enables us to identify the specific unobserved attributes that generate IIA violations. More generally, we find that our incorporation of the attitudinal data leads to a substantial improvement in choice model fit, and more precise estimates of all choice model parameters.

Since we assume that the taste parameters in the heterogeneous logit model are normally distributed, the model generates choice probabilities that are multivariate normal integrals, and estimation requires the use of simulation techniques. Despite this difficulty, the heterogeneous logit model has been used several times previously in marketing (see, e.g., Elrod, 1988; Erdem, 1996; Revelt and Train, 1996). Although the advent of simulation techniques has made estimation of multinomial probit models feasible, heterogeneous logit models are rapidly becoming popular because they appear to have some important ease-of-use advantages. Despite this, the published Monte-Carlo work on small sample performance of simulation-based estimators has dealt exclusively with probit models (see Keane, 1994; Geweke et al., 1994, 1997; Lee, 1997). Our Monte-Carlo analysis of heterogeneous logit models fills this gap.

[1] Attitudinal data can also aid identification in other important ways. For instance, important attributes of choice alternatives are often collinear in market data, making it impossible to infer from revealed preference data how consumers value the separate attributes. Adamowicz et al. (1994) argue that a key advantage of stated preference data over revealed preference data arises because stated preference choice experiments can always be designed so that attributes of interest are not collinear. Alternatively, we show that attitudinal data can be combined with market data to obtain some separate information on consumer valuations of attributes that are collinear in the market.

The outline of the paper is as follows. In Section 2, we describe our extended heterogeneous logit model. In Section 3, we present Monte-Carlo results on the performance of a simulated maximum likelihood estimator for the model. In Section 4, we describe the Twin Cities Medicare data and present our empirical results. Section 5 concludes.

2. An heterogeneous logit model incorporating stated preference weight data

In this section, we present an extended heterogeneous logit model that combines revealed and stated preference data. We begin with the simple heterogeneous (conditional) logit model

$$U_{ij} = X_j\beta_i + \varepsilon_{ij}, \quad j = 1,\dots,J, i = 1,\dots,N, \tag{1}$$

$$d_{ij} = \begin{cases} 1 & \text{if } U_{ij} > U_{im} \quad \forall m \neq j, \\ 0 & \text{otherwise,} \end{cases}$$

where U_{ij} is the utility that person i obtains given choice of option j, X_j is a $1 \times K$ matrix of attributes of alternative j, β_i is the $K \times 1$ vector of utility weights that consumer i applies to the attributes, and ε_{ij} is a standard extreme value error term. The ε_{ij} are assumed independent across alternatives and to have equal variance for all alternatives. Suppose β_{ik}, the utility weight that consumer i assigns to attribute k, is given by

$$\beta_{ik} = \beta_k + \sigma_k\mu_{ik} \quad \textit{for } k = 1,\dots,K, i = 1,\dots,N, \tag{2}$$

where $\mu_{ik} \sim$ N(0,1), so that $\beta_{ik} \sim$ N(β_k,σ_k). Then choice probabilities take the form

$$P(d_{ij} = 1|X, \beta, \sigma) = \int_{\mu_i} \frac{\exp\{X_j(\beta + \sigma\mu_i)\}}{\sum_{m=1}^{J} \exp\{X_m(\beta + \sigma\mu_i)\}} f(\mu_i)\, d\mu_i, \tag{3}$$

where $X \equiv (X_1,\dots,X_J)$ is the $J \times K$ matrix of attributes of all alternatives, $\beta \equiv (\beta_1,\dots,\beta_K)'$ denotes the $K \times 1$ vector of population mean utility weights, $\mu_i \equiv (\mu_{i1},\dots,\mu_{iK})'$ denotes the $K \times 1$ vector of individual components of utility weights, σ denotes the $K \times K$ matrix with standard deviations of the utility weights along the diagonal, and $P(\bullet|\bullet)$ denotes conditional probability.

If J is greater than 3, it is not feasible to evaluate such integrals analytically, and simulation techniques are needed. Observe that the integral in (3) is easier to simulate than a multivariate normal integral, because a simple frequency simulator will be smooth. That is, if we obtain D random draws for the μ_i, denoted

μ_i^d for $d = 1, \dots, D$, and form the simple frequency simulator

$$\hat{P}(d_{ij} = 1 | X, \beta, \sigma) = D^{-1} \sum_{d=1}^{D} \frac{\exp\{X_j(\beta + \sigma\mu_i^d)\}}{\sum_{m=1}^{J} \exp\{X_m(\beta + \sigma\mu_i^d)\}}, \tag{4}$$

the result is a smooth function of the model parameters (β_k, σ_k), $k = 1, \dots, K$. The extreme value error term plays the role of a kernel smoother, as described in McFadden (1989). This simplicity with which choice probabilities may be simulated is a key attraction of heterogeneous logit.

The heterogeneous logit model can easily allow for unobserved attributes of alternatives, as in Elrod (1988). Let there be P unobserved (i.e., not included in X) attributes of alternatives, and let $A_j \equiv (A_{j1}, \dots, A_{jP})$ denote the vector of levels of the unobserved attributes possessed by alternative j. Further, let $W_i \equiv (W_{i1}, \dots, W_{iP})'$ denote the utility weights that consumer i assigns to the attributes, and assume that $W_i \sim \mathrm{N}(W, \Sigma)$, where Σ is a diagonal matrix. Then the heterogeneous logit model can be rewritten as

$$U_{ij} = X_j\beta_i + A_jW_i + \varepsilon_{ij} \quad j = 1, \dots, J, \ i = 1, \dots, N. \tag{5}$$

Model (5) requires some restrictions for identification. First, the scale of the A_{jp}'s and the variance of the W_{ip}'s are not separately identified, so we assume without loss of generality that $W_i \sim \mathrm{N}(W, I_{\mathrm{P}})$. Second, note that only relative attribute positions affect choices and can be identified. So we restrict $A_{jp} = 0$ for $j = 1, \dots, p$ and $p = 1, \dots, P$ without loss of generality. For example, in the $P = 2$ case we restrict $A_{11} = A_{12} = 0$, so that alternative 1 is placed at the origin of the 'market-map' that contains the positions of the alternatives on the two unobserved attribute dimensions. But, since only relative attribute positions affect choices, choices are invariant to rotation of this map about the origin. Thus, the additional restriction $A_{22} = 0$, which forces alternative 2 to lie on the horizontal axis of the two-dimensional market-map, is also necessary. Third, observe that at most $J - 1$ unobserved attributes may be allowed. With $P = J - 1$, one obtains a model in which the error term is the sum of an extreme value error and a normal error, and the normal error term has an unrestricted covariance matrix whose lower triangular Cholesky matrix is $A \equiv \{A_{jp}\}$. If $P < J - 1$, it is equivalent to imposing a lower dimensional factor structure on that covariance matrix.

The reason we can restrict alternative 1 to be at the origin of the two-dimensional market-map (i.e., it has 'zero' levels of all unobserved attributes), and alternative 2 to lie along the horizontal axis (i.e., it has a 'zero' level of attribute 2), and so on, without loss of generality, is that the true identity of the unobserved attributes is not known. Market-map methods suffer from the same fundamental non-identification as factor analysis. It is up to the subjective judgment of the researcher to rotate (ex post) the estimated market-map to a position that looks 'reasonable' and then assign names to the attributes (factors).

One benefit of using attitudinal data is that it can solve this identification problem. We allow the distribution of preference weights to differ across individuals who respond differently to questions about how they weigh various attributes in their decisions. Let S_{ik} be the response of person i to a categorical question in which they are asked how much they weigh observed attribute k in choosing a health plan. In the Twin Cities Medicare data, there are 3 ordered response categories. Let S^*_{ip} be a similar measure of how much a person i weighs an unobserved attribute which we identify as the pth unobserved attribute. In general, we could allow the distribution of preference weights for an attribute to differ in an unrestricted way across groups of people who provide different answers when asked how much they value that attribute. Instead, in order to achieve a more parsimonious representation, we simply let the means of the preference weight distributions vary linearly with the responses to the attribute importance questions

$$\beta_{ik} = \beta_{0k} + \beta_{1k} S_{ik} + \sigma_k \mu_{ik}, \quad k = 1,\ldots,K, \tag{6}$$

$$W_{ip} = W_{0p} + W_{1p} S^*_{ip} + \sigma^*_p v_{ip}, \quad p = 1,\ldots,P, \tag{7}$$

where $\mu_{ik} \sim$ N(0,1) and $v_{ip} \sim$ N(0,1). Combining Eqs. (6) and (7) with Eq. (5), we obtain the unrestricted version of our extended heterogeneous logit model

$$U_{ij} = \sum_{k=1}^{K} (\beta_{0k} + \beta_{1k} S_{ik} + \sigma_k \mu_{ik}) X_{jk} + \sum_{p=1}^{P} (W_{0p} + W_{1p} S^*_{ip} + \sigma^*_p v_{ip}) A_{jp} + \varepsilon_{ij}, \quad j = 1,\ldots,J, \ i = 1,\ldots,N. \tag{8}$$

We now discuss identification issues in this model. First, observe that we must again restrict $A_{1p} = 0$ for $p = 1,\ldots,P$ so that alternative 1 is placed at the origin of the market-map. However, a crucial distinction between the extended heterogeneous logit model (8) and the simple heterogeneous logit model (5) is that no additional restrictions on A are necessary to achieve rotational invariance in Eq. (8). For instance, if $P = 2$, the restriction that $A_{21} = 0$ is required in model (5). But in Eq. (8) this restriction would be *with* loss of generality, because it says that alternatives 1 and 2 have the same level of attribute 2. Such a restriction is innocuous if attribute 2 is just a nameless common factor, but in Eq. (8) the identity of attribute 2 is determined by the importance measure S^*_{i2}. Thus, the information contained in the importance measures breaks the fundamental non-identification problem that exists in factor analytic market-mapping. Of course, if $W_{1p} = 0$ (i.e., the consumers' responses to the importance questions are not informative about their positions in the heterogeneity distribution) then this argument breaks down.

Second, note that the scale of the A_{jp} for $j=2,\dots,J$ is not identified separately from that of W_{0p},W_{1p} and σ_p^*. As a scale normalization we restrict $\sigma_p^*=1$ for $p=1,\dots,P$. This is analogous to the $\Sigma = I_P$ restriction in Eq. (5).

Third, if alternative specific intercepts are included in X then obviously the W_{0p} are not identified. Also, since there are no attribute importance questions corresponding to the intercepts, one obviously has that $\beta_{1k}=0$ for each k corresponding to an intercept. Note, however, that it is standard in market-map analysis to omit explicit alternative specific intercepts. Rather, each alternative has an implicit intercept, which is interpreted as the mean preference in the population for the unobserved attributes associated with that alternative. The implicit intercept for alternative j is the population mean of the second term in Eq. (8), which is $\sum_{p=1}^{P}(W_{0p}+W_{1p}\overline{S_{ip}^*})A_{jp}$. Observe that these implicit intercepts lie in the space spanned by the unobserved attributes A. This is highly desirable for two reasons. First, it rationalizes the intercepts as arising from unobserved attributes of alternatives, rather than treating them as a black-box. Second, it enables one to forecast demand for a new alternative. If intercepts are unrestricted, one cannot determine the intercept value for a new alternative. But if intercepts are constrained to lie in the space spanned by A, then, provided that the researcher can describe the position of the new alternative along the unobserved attribute dimensions, the intercept for the alternative can be determined. Thus, we will only consider models in which X does not contain alternative specific constants.

A fourth and rather subtle problem is that one can always find transformations of the A_j,W_0, and W_1 vectors that leave the first and second moments of the second term in Eq. (8) unchanged (examples of such transformations are available from the authors on request).[2] The only effect of such transformations is to alter the shape of the error distribution, which is determined by how the normal v_{ip} and the multinomial S_{ip}^* are mixed in the second term of Eq. (8). Thus, identification of the model with W_{0p} and W_{1p} for $p=1,\dots,P$ and A_{jp} for $j=2,\dots,J$ and $p=1,\dots,P$ all free would be extremely tenuous, since it would hinge on the information in higher moments of the error distribution.

As a practical matter, a reasonable way to deal with this problem is to restrict the $W_{0p}=0$ for $p=1,\dots,P$ and to simultaneously normalize the S_{ip}^* to have positive means for all p. Then, the contribution of any unobserved attribute p to the intercept for an alternative j is constrained to be proportional to the product

[2] For example, with independent importance weights, the variance of the second term in Eq. (8) is given by $\sum_{p=1}^{P}(W_{1p}^2\mathrm{Var}(S_{ip}^*)+1)A_{jp}^2$ and the covariance between those terms associated with two alternatives j and m is $\sum_{p=1}^{P}(W_{1p}^2\mathrm{Var}(S_{ip}^*)+1)A_{jp}A_{mp}$ where both formulas assume the restriction that $\sigma_p^*=1$ for $p=1,\dots,P$ has been imposed. In this case, transformations of the A_j, W_0, and W_1 vectors that leave unchanged the implicit intercepts for all alternatives, along with all these variances and covariances, are rather simple to see.

of the level of that attribute possessed by the alternative, A_{jp}, and the mean importance that consumers say they assign to that attribute, $\overline{S^*_{ip}}$. This constraint is quite appealing in terms of model interpretation, provided the S^*_{ip} for all attributes are measured on the same scale. Then, alternative j will tend to have a larger positive intercept to the extent that it has larger positive levels of attributes that the mean individual weights heavily.

Fifth, note that if two attributes in X are perfectly collinear, then the β_{0k} associated with one attribute must be fixed at zero, but the separate β_{1k} associated with each attribute can be identified provided there is not perfect collinearity in the S_{ik} for the two attributes. Note also that if two unobserved attributes are collinear in the market, the utility weights consumers assign to each will be separately identified provided that the S^*_{ip} for those two attributes are not perfectly collinear. Thus, it is possible to forecast demand for a new alternative whose position along these two attribute dimensions does not lie along the line observed in the market.

Sixth, in Eq. (8) one does not need the restriction $P \leqslant J-1$. One can estimate a market-map that is higher dimensional than the number of alternatives in the choice set. This feature is likely to be very important in situations in which there are only a small number of alternatives.

In the next section, we turn to a Monte-Carlo analysis of our extended heterogeneous logit model. However, preliminary Monte-Carlo investigation revealed that the likelihood surface is very flat along ridges in the parameters σ_k for $k=1,\dots,K$. To deal with this problem we impose the additional constraint $\sigma_k = \sigma\beta_{1k}$ for $k=1,\dots,K$. This constraint has an appealing interpretation. It says that the attribute importance measures S_{ik} convey the same amount of information about consumer i's relative position in the population distribution of importance weights for all attributes k in the following sense: a one unit increase in S_{ik} implies a σ^{-1} standard deviation increase in consumer i's preference for attribute k, for all attributes k.

With all the restrictions discussed above imposed, the final model we use in the Monte-Carlo analysis is

$$U_{ij} = \sum_{k=1}^{K} (\beta_{0k} + \beta_{1k}S_{ik} + \sigma\beta_{1k}\mu_{ik})X_{jk} + \sum_{p=1}^{P} (W_{1p}S^*_{ip} + v_{ip})A_{jp} + \varepsilon_{ij}, \quad j=1,\dots,J,\ i=1,\dots,N. \tag{9}$$

In the empirical analysis, we find that the likelihood surface is also very flat along ridges in the parameters W_{1p} for $p=1,\dots,P$. For this reason, we apply the same 'equal information' principle as above to the stated importance weights for the *unobserved* attributes by setting $W_{1p} = \sigma^{-1}$ for $p=1,\dots,P$. This restriction appears quite reasonable in the context of our data, because the attribute importance weights are all measured in the same way, and on the same scale.

3. Monte-carlo analysis of a simulated maximum likelihood estimator

3.1. Experimental design

In order to explore the properties of the extended heterogeneous logit in realistic settings, we conduct a series of Monte-Carlo experiments. The first step is to generate artificial data that resemble available market data. We chose attribute levels for each of the five alternatives to represent the types of plans available to elderly consumers in the Medicare supplemental health plan market. Plan attributes in the artificial data are shown in Table 1. The first plan, called the 'Basic Plan,' represents the Medicare entitlement available to all Social Security Recipients. The choice set also includes two types of HMO plans, and two types of 'medigap' plans. The utility in each alternative is derived from three observable attributes and two unobserved attributes.

The two unobserved attributes in the artificial data correspond to perceived quality and convenience. The three observed attributes are constructed to resemble realistic configurations of premiums, coinsurance rates, and provider choice. Consistent with the real world, the premiums and coinsurance for the Basic Plan and the two HMO plans do not vary across individuals. For medigap plans, the degree of individual variation in coinsurance rates in market data depends largely on whether survey participants belong to groups where members choose among a fixed set of health plans, as well as the degree of regulation in the market. Thus, it is important to understand how the degree of coinsurance variation affects inferences. For this reason, we specify two types of coinsurance variation. In the first case, there is no individual variation in coinsurance rates (variance of X_3 'low'). In the second case, coinsurance rates take on three values which are distributed multinomially as shown in Table 1 (variance of X_3 'high').

We generate ordinal attribute importance data for each of the five plan attributes by drawing from a five-variate multinomial distribution. In one case we assume independence. In the other case we mimic the empirical correlations among the attribute importance weights that we observe in the Twin Cities Medicare data.[3]

[3] Specifically, we first draw continuous random vectors $(z_{i1},\dots,z_{iK}; z_{i1}^*,\dots,z_{iP}^*) \sim N(0,\Sigma)$. The attribute importance responses are divided into five discrete categories by setting $S_{ik}=m$ if $b^m \leqslant z_{ik} < b^{m+1}$ and by setting $S_{ip}^*=m$ if $b^m \leqslant z_{ip}^* < b^{m+1}$, where $b^1 = -\infty$, $b^2 = -1.29$, $b^3 = -0.53$, $b^1 = 0.53$, $b^5 = 1.29$, and $b^6 = \infty$ for $m = 1,\dots,5$, $k = 1,\dots,K$ and $p = 1,\dots,P$. For independent importance weights we set $\Sigma = I$. In this case the probabilities that the S_{ik} and S_{ip}^* take on the values 1 through 5 are 0.10, 0.20, 0.40, 0.20 and 0.10, respectively. To generate correlated weights, we set $\mathrm{Vec}(\Sigma) = (1.00, 0.00, 1.00, 0.50, 0.00, 1.00, 0.00, 0.50, 0.25, 1.00, 0.25, 0.25, 0.25, 0.50, 1.00)$, which is based on empirical correlations of the importance weights in the Twin Cities Medicare data.

Table 1
Health plan attribute values in Monte-Carlo analysis

Health plan: $j = 1,\dots,J$	Premium: x_{j1}	Provider choice: x_{j2}	Coinsurance: x_{j3}		Quality of care: A_{j1}	Convenience: A_{j2}
Basic plan ($j = 1$)	0	1	0.30		0.0	0.0
HMO1 ($j = 2$)	0.10	0	0.00		− 0.25	1.00
HMO2 ($j = 3$)	0.20	0	0.00		0.75	0.50
Medigap1 ($j = 4$)	$x_{41} = 0.5 + 0.25v_4$ where $v_{i4} \sim \mathrm{N}(0,1)$	1	Case 1: 'Low' variance $x_{43} = 0.3$	Case 2: 'High' variance $x_{43} \sim \mathrm{MN}(0.1,0.2,0.3)$ where $\mathrm{p}(x_{43} = 0.1) = 0.25$ $\mathrm{p}(x_{43} = 0.2) = 0.25$ $\mathrm{p}(x_{43} = 0.3) = 0.50$	− 0.50	0.75
Medigap2 ($j = 5$)	$x_{51} = 0.7 + 0.25v_5$ where $v_{i5} \sim \mathrm{N}(0,1)$	1	Case 1: 'Low' variance $x_{53} = 0.1$	Case 2: 'High' variance $x_{53} \sim \mathrm{MN}(0.1,0.2,0.3)$ where $\mathrm{p}(x_{53} = 0.1) = 0.50$ $\mathrm{p}(x_{53} = 0.2) = 0.25$ $\mathrm{p}(x_{53} = 0.3) = 0.25$	1.00	− 0.75

We examine the effect of realistic configurations of data on inference in the extended heterogeneous logit model using a full $2 \times 2 \times 2 \times 2$ (i.e., two 'treatment' levels for each of four interventions) factorial design for a total of 16 experiments. Specifically, we vary the sample size ($N = 2000$ or 5000), the level of variation in the coinsurance rate (variance X_3 'low' or 'high,' as discussed above), the presence or absence of correlation in the attribute importance weights, and the number of draws used to approximate choice probabilities ($D = 200$ or 400). For each of the experiments, we generate 20 artificial data sets and evaluate the performance of the simulated maximum likelihood estimator by comparing estimated model parameters to the true values.

3.2. Description of Monte-Carlo results

In Tables 2 and 3 we report results for data sets constructed to have attribute importance weights with a covariance structure similar to that observed in the Twin Cities Medicare data. The two tables report on the properties of estimates obtained using 200 and 400 draws to simulate choice probabilities, respectively.

Each table contains four blocks of results, corresponding to data sets generated with sample sizes of either 5000 or 2000 and the variance of X_3 set to either 'high' or 'low.' Within each block, the first column lists the mean estimate of each parameter across the 20 experiments, the second column lists the empirical standard deviation of the estimates across experiments, and the third column lists the mean of the estimated asymptotic standard errors across experiments.

We now describe some key features of the Monte-Carlo results, focusing on those aspects that will be useful to consider when interpreting our empirical results. First, note that in all cases reported in Tables 2 and 3 the mean parameter estimates are reasonably close to the true values. In particular, the mean estimates for the β parameters, which capture preferences for observed attributes, are generally quite close to the true values. But larger estimated biases are generally observed for the W, σ, and A parameters, which are related to preferences for unobserved attributes and to the taste heterogeneity distribution. In many instances these estimated biases are significant at the 5% level, as indicated by the asterisks in the tables.

Observe that many cases of significant bias appear when $N = 5000$, while few appear when $N = 2000$. This is not surprising. Simulated maximum-likelihood estimates are only consistent and asymptotically normal (with a limiting distribution properly centered at zero) if the number of draws is increased with sample size at a sufficient rate so that $DN^{-1/2} \rightarrow \infty$ as $N \rightarrow \infty$ (Lee, 1995). Thus, if we hold the number of draws fixed while increasing sample size, there should come a point where significant bias appears.

An interesting result is that the estimates of W_{11} and W_{12} are always biased upward. When this bias is significant, the associated A parameters are always biased downward towards zero, typically by about 10–20%. Importantly, the

Table 2
Monte-Carlo results, 200 draws, correlated weights

Parameter	True	$N = 5000$, $\sigma_{x3} =$ High			$N = 2000$, $\sigma_{x3} =$ High			$N = 5000$, $\sigma_{x3} =$ Low			$N = 2000$, $\sigma_{x3} =$ Low		
		$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$
β_{01}	− 1.00	− 0.940	0.254	0.224	− 1.042	0.362	0.386	− 0.940	0.204	0.225	− 1.154	0.472	0.413
β_{02}	2.00	1.935	0.216	0.264	2.061	0.547	0.434	1.876	0.354	0.340	2.104	0.770	0.614
β_{03}	− 2.00	− 2.170	0.614	0.678	− 2.002	1.556	1.174	− 1.991	1.296	1.250	− 1.933	2.460	2.195
β_{11}	− 0.50	− 0.489	0.060	0.064	− 0.511	0.121	0.107	− 0.472*	0.056	0.064	− 0.530	0.149	0.117
β_{12}	1.00	0.979	0.126	0.121	1.043	0.304	0.204	0.956	0.107	0.123	1.132	0.348	0.237
β_{13}	− 2.00	− 1.847*	0.318	0.266	− 1.989	0.620	0.448	− 1.828*	0.283	0.269	− 2.142	0.634	0.503
W_{11}	1.00	1.198*	0.315	0.399	1.096	0.357	0.516	1.156*	0.283	0.424	1.329	1.010	2.493
W_{12}	1.00	1.145*	0.295	0.289	1.097	0.447	0.577	1.175*	0.345	0.413	1.192	0.661	1.082
σ	1.00	0.799*	0.412	0.490	0.792	0.520	1.217	0.722*	0.366	0.486	0.936	0.542	0.827
A_{21}	− 0.25	− 0.206*	0.086	0.129	− 0.278	0.267	0.258	− 0.244	0.119	0.137	− 0.264	0.380	0.332
A_{31}	0.75	0.666	0.209	0.236	0.842	0.447	0.396	0.697	0.232	0.262	0.930	0.502	0.493
A_{41}	− 0.50	− 0.415*	0.133	0.163	− 0.500	0.201	0.281	− 0.412*	0.150	0.174	− 0.601	0.344	0.368
A_{51}	1.00	0.886	0.263	0.302	1.082	0.507	0.481	0.912	0.269	0.342	1.207	0.639	0.631
A_{22}	1.00	0.889*	0.221	0.230	1.055	0.391	0.431	0.879*	0.231	0.263	1.064	0.430	0.526
A_{32}	0.50	0.440*	0.118	0.151	0.525	0.285	0.276	0.411*	0.131	0.159	0.555	0.336	0.328
A_{42}	0.75	0.666	0.184	0.181	− 0.767	0.278	0.327	0.647	0.183	0.206	0.808	0.354	0.410
A_{52}	− 0.75	− 0.687	0.215	0.202	− 0.785	0.374	0.351	− 0.658*	0.187	0.226	− 0.815	0.418	0.435

Note: $\bar{\hat{\theta}} = \frac{1}{20}\sum_{r=1}^{20} \hat{\theta}_r$, s.d. $= \sqrt{\frac{1}{19}\sum_{r=1}^{20} (\hat{\theta}_r - \bar{\hat{\theta}})^2}$, and $\overline{\text{ASE}} = \frac{1}{20}\sum_{r=1}^{20} \text{ASE}_r$ where r denotes replication, $\hat{\theta}_r$ denotes the estimate on replication r, and ASE_r is the estimated standard error from replication r. An asterisk indicates $|t| > 2.09$ where $t = \sqrt{20}(\bar{\hat{\theta}} - \theta_{\text{true}})\text{s.d.}^{-1}$ is the t-statistic for the estimated bias.

Table 3
Monte-Carlo results, 400 draws, correlated weights

Parameter	True	$N = 5000$, $\sigma x_3 =$ High			$N = 2000$, $\sigma x_3 =$ High			$N = 5000$, $\sigma_{x3} =$ Low			$N = 2000$, $\sigma_{x3} =$ Low		
		$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$	$\bar{\hat{\theta}}$	s.d.	$\overline{\text{ASE}}$
β_{01}	− 1.00	− 0.963	0.243	0.228	− 1.032	0.368	0.377	− 0.950	0.227	0.229	− 1.127	0.460	0.408
β_{02}	2.00	1.947	0.203	0.272	2.048	0.536	0.452	1.865	0.359	0.341	2.075	0.721	0.601
β_{03}	− 2.00	− 2.170	0.607	0.686	− 2.032	1.505	1.151	− 1.949	1.194	1.256	− 1.965	2.387	2.170
β_{11}	− 0.50	− 0.497	0.064	0.065	− 0.507	0.125	0.108	− 0.477	0.057	0.066	− 0.522	0.146	0.117
β_{12}	1.00	1.001	0.124	0.128	1.024	0.305	0.210	0.967	0.112	0.127	1.096	0.313	0.234
β_{13}	− 2.00	− 1.886	0.313	0.278	− 1.956	0.627	0.459	− 1.842*	0.283	0.275	− 2.090	0.614	0.504
W_{11}	1.00	1.266	0.643	0.761	1.181	0.602	1.039	1.216*	0.478	0.655	1.407	1.254	4.331
W_{12}	1.00	1.099*	0.201	0.225	1.126	0.477	0.658	1.152*	0.263	0.317	1.155	0.538	0.862
σ	1.00	0.841	0.433	0.537	0.797	0.522	1.115	0.782*	0.363	0.694	0.941	0.493	0.987
A_{21}	− 0.25	− 0.208	0.096	0.125	− 0.295	0.293	0.263	− 0.253	0.137	0.142	− 0.297	0.346	0.333
A_{31}	0.75	0.681	0.231	0.246	0.831	0.453	0.429	0.707	0.268	0.275	0.922	0.479	0.532
A_{41}	− 0.50	− 0.432	0.155	0.173	− 0.488	0.212	0.300	− 0.437	0.182	0.185	− 0.589	0.329	0.399
A_{51}	1.00	0.904	0.288	0.317	1.068	0.518	0.518	0.926	0.316	0.361	1.197	0.623	0.694
A_{22}	1.00	0.919	0.186	0.222	1.022	0.394	0.436	0.892*	0.219	0.256	1.033	0.398	0.493
A_{32}	0.50	0.458	0.107	0.150	0.507	0.300	0.276	0.415*	0.131	0.157	0.523	0.312	0.313
A_{42}	0.75	0.688	0.159	0.175	0.741	0.279	0.332	0.657*	0.170	0.201	0.784	0.315	0.392
A_{52}	− 0.75	− 0.709	0.174	0.195	− 0.756	0.351	0.358	− 0.670*	0.176	0.219	− 0.799	0.387	0.416

Note: $\bar{\hat{\theta}} = \frac{1}{20}\sum_{r=1}^{20} \hat{\theta}_r$, s.d. $= \sqrt{\frac{1}{19}\sum_{r=1}^{20} (\hat{\theta}_r - \bar{\hat{\theta}})^2}$, and $\overline{\text{ASE}} = \frac{1}{20}\sum_{r=1}^{20} \text{ASE}_r$ where r denotes replication, $\hat{\theta}_r$ denotes the estimate on replication r, and ASE_r is the estimated standard error from replication r. An asterisk indicates $|t| > 2.09$ where $t = \sqrt{20}(\bar{\hat{\theta}} - \theta_{\text{true}})\text{s.d.}^{-1}$ is the t-statistic for the estimated bias.

model does an excellent job of sorting out *relative* positions of alternatives on the unobserved attribute dimensions, and of determining the mean contributions of the unobserved attributes to utility – the $W_{11}A_{jl}$ products.[4] We therefore conclude that the observed bias in the W and A parameters is not a serious problem.

For purposes of statistical inference, it is desirable that empirical and asymptotic standard errors be close. For the β parameters, the agreement is generally good when $N = 5000$, but asymptotic standard errors understate empirical standard errors by about 25% on average when $N = 2000$. This illustrates the well-known phenomenon that quite large sample sizes are usually necessary before asymptotics 'kick in' for discrete choice models. An important result is that, for the A parameters, the asymptotic standard errors usually exceed the empirical standard errors by about 15% even when $N = 5000$. Thus, the asymptotic standard errors overstate the degree of uncertainty about relative positions of the alternatives on the unobserved attribute dimensions.

We are motivated to compare results when the variance of X_3 is 'low' vs. 'high' because of the findings in Keane (1992) which suggest that it becomes very difficult to identify covariance parameters in multinomial probit models when there is little variation in attributes across respondents. We wish to see if the same type of phenomenon appears in heterogeneous logit models. Indeed, we find that, when $N = 2000$, setting the variance of X_3 to 'low' leads to substantial increases in empirical standard errors for *all* model parameters (see the last block of columns in Tables 2 and 3). But this problem is not too severe when $N = 5000$.

A comparison of Tables 2 and 3 reveals the effect of increasing simulation size. Notice that, in the $N = 5000$ and variance of X_3 'high' case, all but one of the significant biases that appeared with 200 draws disappear when 400 draws are used. The remaining case of significant bias is for W_{12}. But the magnitude of the bias is only about 10% of the true value, and the model seems to compensate for this by shifting the A_{j2} parameters towards 0, thereby maintaining precision in the estimated mean utility contributions of the unobserved attributes.

We do not report our results for the data sets constructed to have attribute importance weights that are independent across attributes (these are available on request). Such a structure is not realistic (e.g., we would expect respondents who place a relatively high weight on low premiums to also value low coinsurance rates), but it enables us to learn about how much information is lost when

---

[4] Referring to Eq. (9), we see that the tendency to overestimate the W and underestimate the A means that the model underestimates the degree of heterogeneity in preferences for the latent attributes. This is because the degree of heterogeneity in preferences for the unobserved attributes is determined by the length of the A vectors. Similarly, observe that σ tends to be biased towards 0. This means that the model also tends to underestimate the degree of heterogeneity in preferences for the observed attributes.

the importance weights are correlated. Indeed, with independent attribute important weights the empirical standard errors drop substantially for many model parameters. The asymptotic standard errors reflect this gain in precision well for the β, W, and σ parameters. But again, asymptotic standard errors overstate empirical standard errors for the A parameters. Other aspects of the results are similar to those we have discussed.

In summary, we conclude that the SML estimator does a good job of uncovering the parameters of heterogeneous logit models of the type we are interested in estimating. Significant biases do appear for many model parameters. But in all cases we examined, biases appear to be modest as a percentage of the true parameter values, and biases in combinations of parameters often appear to cancel out, leading to accurate estimates of economic quantities of interest.

4. Empirical results

4.1. Description of the data

The data for our empirical analysis come from a 1988 survey questionnaire administered to elderly Medicare beneficiaries (non-Medicaid eligible) living in the seven county Minneapolis–St. Paul (Twin Cities) metropolitan statistical area. The analysis is restricted to the subgroup of sample members ($N = 1274$) reporting enrollment in an individual market plan. The average age of the sample is 74, median income is $15,154, 60% are female, and 34% are in poor health.

Sample members are assumed to choose among five types of plans: (1) basic Medicare entitlement without any form of supplemental coverage, (2) medigap plans without drug coverage, (3) medigap plans with drug coverage, (4) a type of HMO called an independent practice organization (IPA), and (5) network HMOs. The fractions of the sample enrolled in each option are 12.8%, 12.6%, 16.5%, 21.7% and 36.4%, respectively. The observable attributes for the five plans are displayed in Table 4.[5]

[5] The attributes of the basic Medicare entitlement, of the two types of HMO plans, and some of the non-financial, non-coverage features of medigap plans do not vary across individuals and were obtained from widely distributed informational and marketing materials. Coverage and cost sharing characteristics of the medigap plans do vary across respondents, and the data contained some limited information about the characteristics of medigap plans actually held by survey respondents. This information was sufficient to classify medigap plans into two main types: those that cover outpatient prescription drugs and those that do not. Medigap premium information was obtained from a 1990 study of the medigap market conducted by the Minnesota Department of Commerce.

Table 4
Values corresponding to observable attributes in 1988 Twin Cities Medicare data

Attribute	Basic Medicare	Medicare + Medigap w/o Drugs	Medicare + Medigap w/ Drugs	IPA	Network
Monthly Premium[a]	\$28	Age 65–69 \$71 70–74 \$75 75 + \$82	Age 65–69 \$95 70–74 \$102 75 + \$109	\$53	\$40
Covers drugs (1,0)	0	0	1	0	1
Covers preventative care (1,0)	0	0	0	1	1
Permits provider choice (1,0)	1	1	1	1	0
Submit claims (1,0)	1	1	1	0	0

[a] All premiums include \$28 for voluntary Medicare Part B coverage.

The Twin Cities data did not contain information about two attributes that earlier studies suggest are important to elderly decision makers in choosing a health plan (Dowd et al., 1994; Walker, 1990; Marquis et al., 1985). The first is detailed information about the cost sharing requirements of the medigap plans. Moreover, the retrospective health status and service use data needed to construct a reasonable proxy for expected out-of-pocket spending for each alternative was also unavailable. Second, the data do not contain measures of the quality of care provided by each of the five plans. This is not surprising because, to date, it is not clear that valid measures of quality of care from the perspective of consumers even exist. Thus, we treat cost sharing and quality of care as unobserved attributes in the analysis.

The survey respondents were asked to rate the importance of 10 health plan attributes in the context of choosing a health plan. Table 5 lists the proportion responding at each level of importance for each attribute. We assign values of 1, 2 and 3 to the three ascending levels of importance. We create a single measure of the importance of the overall quality of care by summing responses to the three questions on referral to specialists, quality of care, and not being rushed from hospital, and dividing by 3. Similarly, we created a single measure of the importance of provider choice by averaging the importance of physician choice and hospital choice items.

4.2. Estimation results

In Table 6, we report estimates of our extended heterogeneous logit model of the health plan choices of elderly Twin Cities Medicare beneficiaries. We set

Table 5
Proportion of sample members ($n = 1274$) responding at each level of attribute importance.[a]

Attribute	Have to have...		Like to have...		...Doesn't matter	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
Lowest premiums	293	23.0	754	59.2	227	17.8
Drug coverage	281	22.1	767	60.2	226	17.7
Preventive care	408	32.0	705	55.3	161	12.6
Choice of hospital	329	25.8	769	60.4	176	13.8
Choice of physician	439	34.5	705	55.3	130	10.2
Minimum paperwork	481	37.8	677	53.1	116	9.1
Highest quality of care	561	44.0	657	51.6	56	4.0
Referral to specialists	525	41.2	690	54.2	59	4.6
Not being rushed from hospital	418	32.8	718	56.4	138	10.8
Lowest cost sharing	395	31.0	761	59.7	118	9.3

[a] The attribute importance question in the survey of Twin Cities Medicare beneficiaries was worded as follows: If you had to change your health plan for any reason, which of the following features would a health plan have to have at a minimum for you to consider it? For each feature that I read, please tell me if you would have to have, like to have or doesn't matter to consider a plan. 'Have to to have' is coded 3, 'Like to have' is coded 2, and 'Doesn't matter' is coded 1.

$D = 200$ to implement the SML procedure. We report parameter estimates for four versions of the model: (1) the 'Full Model' that includes two latent attributes – quality of care and cost sharing, (2) a 'Restricted Model' that imposes some restrictions on the relative positions of the alternatives on the latent attribute dimensions, (3) a model that excludes the quality of care latent attribute, and (4) a model that excludes both latent attributes.

We begin by describing the estimates of the Full Model. Observe that the intercept in the preference weight equation for premium is 0.014, while the slope coefficient is -0.007. Together, these parameters determine the mean coefficients on premium among different groups of respondents. Specifically, among respondents whose importance weights for low premiums are 1, 2 and 3, the predicted mean coefficients on premium are 0.007, 0.000 and -0.007, respectively. Thus, only among the group of respondents who place the highest importance on low premiums is the point estimate for the mean coefficient on premium negative. To evaluate the implied magnitudes of premium effects, we report in Table 7 simulations of the effects of a \$20 increase in the monthly premium for each plan. The model predicts little change in market share for any plan.

Drug coverage is an additional aspect of cost that we include as an observed attribute, and the estimates imply that it is an important consideration for many elderly beneficiaries. For example, according to the simulations in Table 7, if the

Table 6
Extended heterogeneous logit model of the health plan choices of Twin Cities Medicare beneficiaries

Parameter label	θ	Full model	Restricted model	Single latent attribute	No latent attributes
Constant in preference equation					
Premium	β_{01}	0.014	0.017	0.0187	0.017**
		(0.011)	(0.012)	(0.011)	(0.006)
Drug coverage	β_{02}	0.057	– 0.200	0.088	– 0.470
		(0.912)	(0.740)	(0.790)	(0.446)
Preventive care and submit claims	β_{03}	1.887**	1.770**	1.162**	– 0.013
		(0.498)	(0.501)	(0.437)	(0.320)
Provider choice	β_{04}	– 0.395	– 1.278	– 1.122	– 2.445**
		(1.081)	(1.054)	(1.055)	(0.810)
Slope in preference equation					
Premium	β_{11}	– 0.007**	– 0.009**	– 0.009**	– 0.009**
		(0.003)	(0.003)	(0.003)	(0.003)
Drug coverage	β_{12}	0.384**	0.547**	0.554**	0.406**
		(0.145)	(0.191)	(0.181)	(0.150)
Preventive care	β_{13}	0.766**	0.881**	0.805**	0.574**
		(0.202)	(0.233)	(0.193)	(0.146)
Provider choice	β_{14}	1.430**	1.809**	1.912**	1.060**
		(0.489)	(0.619)	(0.512)	(0.342)
Submit claims	β_{15}	– 0.274**	– 0.292	– 0.123	0.054
		(0.130)	(0.151)	(0.130)	(0.099)
Heterogeneity in preferences					
All attributes	σ	0.372	0.729*	0.983**	0.250
		(0.873)	(0.325)	(0.237)	(0.596)
Latent attribute 1: *cost sharing*					
Medigap w/o drugs	A_{21}	– 0.270	– 0.559*	– 0.050	
		(0.664)	(0.332)	(0.200)	
Medigap w/drugs	A_{31}	– 0.355	– 0.777*	– 0.512	
		(0.859)	(0.465)	(0.344)	
IPA	A_{41}	– 0.414	– 0.772*	– 0.951**	
		(1.013)	(0.452)	(0.444)	
Network	A_{51}	– 0.271	Restricted to A_{41}	Restricted to A_{41}	
		(0.638)			
Latent attribute 2: *quality of care*					
Medigap w/o drugs	A_{22}	0.269	0.615		
		(0.644)	(0.386)		
Medigap w/ drugs	A_{32}	0.261	Restricted to A_{22}		
		(0.640)			
IPA	A_{42}	– 0.0811	– 0.196		
		(0.241)	(0.260)		
Network	A_{52}	0.161	0.508		
		(0.422)	(0.417)		
Likelihood		– 1834.002	– 1838.197	– 1854.909	– 1876.642

Note: Standard errors are in parentheses. Double asterisks indicate that an estimated parameter is significant at the 5% level. A single asterisk indicates significance at the 10% level.

Table 7
Changes in predicted choice probabilities resulting from changes in observable health plan attributes

Health Plan	Changes	Basic Medicare	Medigap w/o drugs	Medigap w/ drugs	IPA	Network
	Baseline probabilities	0.091	0.094	0.124	0.256	0.436
Basic Medicare	Add $20.00 to premium	– 0.002	0.000	0.000	0.000	0.002
Medigap w/o drugs	" "	0.000	0.001	0.000	0.000	– 0.001
Medigap w/ drugs	" "	0.000	– 0.001	0.002	– 0.001	– 0.001
IPA	" "	0.000	0.000	– 0.001	0.001	0.000
Network	" "	0.001	0.000	0.000	0.001	– 0.002
Basic Medicare	Add drug coverage	0.086	– 0.012	– 0.015	– 0.034	– 0.024
Medigap w/o drugs	Add drug coverage	– 0.014	0.088	– 0.018	– 0.036	– 0.020
Medigap w/ drugs	Remove drug coverage	0.002	0.006	– 0.066	0.008	0.050
IPA	Add drug coverage	– 0.024	– 0.023	– 0.033	0.161	– 0.081
Network	Remove drug coverage	0.025	0.023	0.036	0.068	– 0.152
Basic Medicare	Remove provider choice	– 0.082	0.006	0.009	0.011	0.056
Medigap w/o drugs	Remove provider choice	0.005	– 0.086	0.013	0.013	0.054
Medigap w/ drugs	Remove provider choice	0.008	0.013	– 0.113	0.024	0.068
IPA	Remove provider choice	0.023	0.027	0.039	– 0.230	0.141
Network	Add provider choice	– 0.076	– 0.077	– 0.103	– 0.216	0.472

IPA plan were to add drug coverage, its market share would increase by 16.1 percentage points.[6] Among the non-cost related observed attributes, the estimates imply that provider choice is a very important concern for many elderly beneficiaries, as can also be seen in Table 7. For example, if the IPA plan were to remove provider choice, the model predicts its market share would decrease by 23 percentage points.[6]

[6] Another aspect of cost is preventive care coverage, but that and the claim submission requirement are collinear across the five plans (i.e., all the plans that cover preventive care do not require claims). This means we can only estimate an intercept term in the preference weight equation for one of these two attributes. The estimate of that intercept is positive (1.887), but we cannot determined the extent to which this means respondents generally like to have coverage of preventive care, dislike having to submit claims, or some combination of the two.

Now consider the unobserved latent attributes. The first is the degree of cost sharing required in each plan. We restrict $A_{11}=0$, so that Medicare is at the origin of the map. The negative point estimates of A_{21} through A_{51} imply that all four supplemental plans are perceived as having *greater* cost sharing requirements than basic Medicare (the coefficients are negative because greater cost sharing is a 'bad'). The IPA plan is perceived as requiring the most cost sharing, since A_{41} is the largest negative element in the first column of A.[7]

These results indicate that elderly Medicare beneficiaries have inaccurate perceptions of the cost sharing requirements of the various plans. While Medicare is perceived as requiring the least cost sharing, it actually has the greatest cost sharing. This finding is consistent with earlier studies indicating that the elderly have a poor understanding of Medicare entitlement and the supplemental health plan market (Davidson et al., 1992; McCall et al., 1986; Cafferata, 1984). Note, however, that none of the parameters A_{21} through A_{51} is individually significant based on the asymptotic standard errors. But in light of our Monte-Carlo analysis in Section 3, which indicated that standard errors for the elements of A tend to be rather severely biased upward, we view these standard errors with some skepticism. We return to this point below.

The second latent attribute is quality of care. We again put basic Medicare at the origin of the map by setting $A_{12}=0$. The point estimates of A_{22} through A_{52} imply that the two medigap plans are perceived as the highest quality, followed by the network HMO, then basic Medicare, and then the IPA plan. Again, however, none of the parameters A_{22} through A_{52} is individually significant based on the asymptotic standard errors.

Finally, note that parameter σ, which is (proportional to) the standard deviation of the preference weight distribution, is not significant according to its asymptotic standard error. This implies that there is no significant unobserved heterogeneity in tastes for observed and unobserved plan attributes (after the observed heterogeneity accounted for by the importance weights is controlled for). However, as was the case with the elements of A, our Monte-Carlo results in Section 3 indicated that standard errors for the parameters capturing the degree of heterogeneity in preferences tend to be biased upward. We turn now to a further investigation of this issue.

---

[7] Recall from the discussion following Eq. (9) that the utility weight on unobserved attributes is restricted to equal $\sigma^{-1}S$. Thus, for example, the point estimates imply a mean utility contribution of $-0.414(1/0.372)S = -1.113S$ for cost sharing for the IPA plan among respondents who assign an importance weight of S to having low cost sharing. Since this is negative, the probability of enrolling only in basic Medicare rather than the IPA plan increases as respondents report being more concerned with low cost sharing. For instance, we have performed a simulation of the model in which each individual's importance weight on cost sharing is increased by an amount equal to 50% of the mean importance weight in the population. We find that the market share of Medicare would increase from 9.1% to 21.0% under this scenario.

In the 'restricted model' in Table 6, two reasonable restrictions are placed on the A matrix. These are that the two medigap plans have the same position on the quality of care dimension and that the two HMOs have the same position on the cost sharing dimension. These restrictions have little effect on the estimates of the parameters in the preference weight equations. However, they cause the estimate of σ to increase substantially and become significant. The magnitudes of the elements of A also increase substantially, all the elements of A corresponding to cost sharing become significant at the 10% level, and A_{22} comes close to that level. Note that the reduction in σ^{-1} tends to counteract the increases in the magnitudes of the elements of A, so that the products $\sigma^{-1}A_{jp}$, which determine mean utility contributions, change little.[8]

The fifth column of Table 6 contains estimates for a model in which the quality of care latent attribute is dropped. Although none of the elements A_{22}, A_{42} or A_{52} is individually significant in the model of column 4, the (simulated) χ^2 (3) statistic for the restriction is 33.42 compared to a 5% critical value of 7.81. Finally, the last column of Table 6 reports a model with both latent attributes dropped, which again causes a substantial deterioration in the log-likelihood.

These findings are consistent with Monte-Carlo results in Section 3. That is, the likelihood is rather flat along ridges in the σ and A parameters, so that σ^{-1} A_{jp} products are pinned down well, as are relative positions of the alternatives along the A dimensions, while absolute magnitudes of σ and the A_{jp} are difficult to pin down. This accounts for the fact that restrictions on the relative positions of alternatives along the A dimensions tend to be rejected, even when the relevant individual elements of A have very large asymptotic standard errors. Thus, we feel that we can have considerable faith in our estimates of relative positions of alternatives along the A dimensions and of the $\sigma^{-1}A_{jp}$ products that determine mean utility contributions.

Our results when we omit the latent attributes of cost sharing and quality of care from the model illustrate well how failure to control for latent attributes can lead to severe bias in estimates of the preference weights for the observed attributes. The most striking change is for the provider choice variable. According to the model in the last column of Table 6 the intercept and slope in the preference weight equation for provider choice imply that most elderly beneficiaries actually dislike having provider choice. This occurs because the only plan to restrict provider choice is the network HMO, and this plan is perceived as having relatively high quality of care. When quality of care is omitted, the model

[8] Observe that a likelihood ratio test rejects the restrictions. The (simulated) χ^2 (2) statistic is 8.40 compared to a 5% critical value of 5.99. This result appears somewhat surprising if we take the asymptotic standard errors seriously, because the elements of A which we have restricted to be equal only differed by small amounts (as compared to their standard errors) when we estimated them freely.

must downplay the importance of provider choice in order to explain why such a large fraction of respondents (36.4%) choose the network HMO.

Next, we consider the contribution of the stated attribute importance weight information to the fit of the choice model. In Table 8, we report estimates for the simple heterogeneous logit or random coefficients model of Eq. (5) that does not

Table 8
Random parameter logit models of the health plan choices of Twin Cities Medicare beneficiaries

Parameter label	θ	2 Latent attributes	1 Latent attribute	No latent attribute	Conditional logit
Mean preference weights					
Premium	β_{01}	0.002	– 0.001	– 0.006	0.000
		(0.476)	(0.294)	(0.118)	(0.003)
Drug coverage	β_{02}	0.900	0.156	– 2.504	0.318
		(54.576)	(37.991)	(4.532)	(0.186)
Preventive	β_{03}	3.390	3.202	2.16	1.125
		(82.883)	(31.120)	(1.792)	(0.084)
Provider choice	β_{04}	– 0.016	0.072	– 2.724	– 0.205
		(99.958)	(67.178)	(3.693)	(0.227)
Standard deviation of preference weights					
Premium	σ_1	0.033	0.036	0.098	
		(0.048)	(0.0443)	(0.175)	
Drug coverage	σ_2	0.112	0.134	0.072	
		(745.352)	(324.693)	(18.731)	
Preventive	σ_3	0.062	0.281	0.751	
		(402.982)	(71.517)	(29.479)	
Provider choice	σ_4	0.009	0.418	0.832	
		(573.104)	(370.026)	(111.121)	
Latent attribute 1					
Medigap w/o drugs	A_{21}	– 0.183	– 0.212		
		(18.911)	(11.287)		
Medigap w/ drugs	A_{31}	0.118	0.102		
		(335.849)	(12.120)		
IPA	A_{41}	2.822	3.401		
		(143.026)	(104.041)		
Network	A_{51}	4.176	4.272		
		(99.471)	(264.305)		
Latent attribute 2					
Medigap w/ drugs	A_{32}	0.18235			
		(751.968)			
IPA	A_{42}	1.724			
		(703.541)			
Network	A_{52}	2.537			
		(425.189)			
Likelihood		– 1956.155	– 1957.909	– 1961.451	– 1964.270

Note: Standard errors are in parentheses.

incorporate the stated importance weight data. We report estimates for random coefficients models that contain two, one, and no latent attributes, respectively, and also for a simple (homogeneous) conditional logit model.

The most striking aspect of Table 8 is that the standard errors for all the parameters in all the heterogeneous logit models are extremely large, even in the model with no latent attributes. This indicates that the Hessians are close to singular and that severe identification problems are present. Only in the simple homogeneous conditional logit model are reasonably precise estimates obtained. Since we did not have such problems in the estimation of our extended heterogeneous logit model, it is clear that the attribute importance weight variables provide crucial information for identification of the heterogeneity distribution.[9]

We also find that the inclusion of the attribute importance information leads to substantial improvement in model fit. The (simulated) log-likelihood value for the ‘Full Model’ in Table 6 exceeds that of the two latent attribute model in Table 8 by 122.16 points. The former model has 18 parameters while the later has 15.[10] Thus, the Akaike information criteria are 3704.0 and 3942.3, respectively, and the ‘Full Model’ is superior by 238.3 points. The Bayes information criterion values (which impose a greater penalty for additional parameters) are 3796.7 and 4019.6, respectively. Thus, by both criteria the inclusion of the stated importance weight information substantially improves model fit. Also, note that the ‘ignorant’ model which assigns a probability of 0.20 to each alternative for each respondent would produce a log-likelihood value of -2050.4. Thus, the pseudo-R^2 for the Full Model is 0.106, while that for the random coefficients model that ignores the stated importance weight information is only 0.046.

4.3. A policy experiment

In this section, we illustrate the usefulness of our model of health plan choices by using it to forecast how elderly Medicare beneficiaries in the Twin Cities

[9] Interestingly, the χ^2 (2) statistic for elimination of latent attribute 1 is 7.08 compared to a 5% critical value of 5.99. Thus, we have a more severe version of the same problem that appeared in the extended heterogeneous logit model: the elements of A are jointly significant but the individual elements have huge standard errors.

[10] The model with two latent attributes in Table 8 is not nested within the ‘Full Model’ in Table 6. When importance weight information is excluded by setting the slope coefficients β_{11} through β_{15} to zero, it is no longer possible to restrict the standard deviation of the individual component of the preference weight for each attribute to be proportional to the slope coefficient, as in Eq. (9). Rather, in Table 8, we estimate separate standard deviations for the random coefficient corresponding to each observed attribute, σ_1 through σ_4. Note also that the requirement to submit claims attribute must be dropped completely from the random coefficients model, since it is collinear with preventive care. Finally, since in the random coefficients model the identity of the latent attributes is not pinned down by the importance weight questions, the rotational invariance restriction $A_{22} = 0$ is imposed.

Table 9
Predicted choice probabilities before and after simulated addition of new HMO alternative to the Twin Cities Medicare health plan market

	Basic Medicare	Medigap w/o drugs	Medigap w/ drugs	IPA	Network	New HMO
Before ($J = 5$)	0.091	0.094	0.124	0.256	0.436	N.A.
After ($J = 6$)	0.068	0.074	0.099	0.196	0.306	0.258

would respond to the introduction of a new health plan option.[11] Specifically, we use the Full Model from Table 6 and simulate the introduction of a new HMO plan. The new plan costs $45 per month, does not cover prescription drugs, covers preventive care, restricts choice of provider, and does not require enrollees to submit claims for reimbursement. It has a perceived quality level similar to that in the network HMO (i.e., we set $A_{62} = 0.161$) and a perceived cost sharing level that is less than that perceived in any of the four existing supplemental plans (i.e., we set $A_{61} = -0.150$).

Table 9 presents the predicted aggregate market shares (i.e., predicted average choice probabilities) before and after the introduction of the new HMO.[12] Observe that the new HMO is predicted to attract a disproportionate share of enrollment away from the existing HMOs. This illustrates how the heterogeneous logit model can generate deviations from the IIA property. With IIA, when a new alternative is added to the choice set, existing alternatives must lose market share proportionately, regardless of how similar they are to the new alternative. But in our model, since the new alternative is similar in structure to the two existing HMOs, it attracts a disproportionate number of enrollees away from the HMO options.

5. Conclusion

We have shown how revealed preference and attitudinal data can be fruitfully combined in the estimation of an extended version of the heterogeneous logit model. This model allows the distribution of the individual specific preference weight parameters to differ across groups of respondents who report that they

[11] A limitation of using our model for forecasting is that one must assume the attitudinal data used in estimation is representative of the preferences of the population whose behavior is being forecast.

[12] Note that baseline market shares do not equal sample market shares because the model does not include alternative specific intercepts.

assign different levels of importance to various attributes of alternatives. We successfully applied the framework to modeling the health plan choices of elderly Medicare beneficiaries in the Twin Cities. Our results indicate that the inclusion of stated preference weight information in the heterogeneous logit model results in a substantial improvement in fit. It also allows the parameters capturing the distribution of preferences for observed attributes of alternatives to be estimated much more precisely, and allows us to identify those unobserved attributes of alternatives that play an important role in driving decisions.

Substantively, our results indicate that failure to account for important unobserved attributes of health plans, such as perceived plan quality, generates an important omitted variable bias. For instance, estimation of logit models that ignore such unobserved attributes leads to the perverse finding that consumers dislike provider choice (because the perceived high quality plans tend to restrict choice). Our results also suggest that elderly Medicare beneficiaries in the Twin Cities do not accurately perceive the cost sharing requirements of the available health care options. Our estimates imply that respondents perceive the basic Medicare plan to have relatively low cost sharing requirements, when in fact it has high cost sharing requirements relative to available supplemental plans. Thus, informational interventions to enhance the elderly's understanding of Medicare and supplemental health plans may be desirable.

It is interesting that the stated attribute importance information in the Twin Cities Medicare data is so useful in predicting the health plan choices of consumers. Since there is no economic incentive for consumers to reveal their preferences accurately built into the survey design, there is no obvious reason to expect the stated importance data to be useful. Nor does our framework assume this a priori, as the model is free to determine that the distribution of preferences does not differ systematically across respondents with different stated preferences. An interesting area for future research is to examine the behavioral and cognitive processes that drive responses to attitudinal questions, and stated preference questions more generally, so we can better understand the contexts in which such data will be predictive of actual market behavior. Recent work by Ben-Akiva and Morikawa (1990) and Hensher and Bradley (1993) is a step in this direction.

Finally, we find that the parameters of a heterogeneous logit model for health plan choice based solely on cross-sectional revealed preference data are very poorly identified. Earlier work indicates that such models may be precisely estimated using panel data (e.g., see Elrod, 1988; Erdem, 1996; Revelt and Train, 1996). But collection of panel data is often very expensive and time consuming, and such data are not available for the study of many questions of interest. For example, to our knowledge there are no available panel data on health care plan choices. Our results indicate that data on consumers' stated attribute importance can in some sense serve as a substitute for panel data in contexts where only cross-sectional data are available, thereby enabling the researcher to obtain precise estimates of heterogeneous logit model parameters.

Acknowledgements

The collection of the data used in this study was funded by the Health Care Financing Administration under cooperative agreement 17-C-99040/5-01. Harris gratefully acknowledges financial support from Health Care Financing Administration Grant 30-P-90529/4-01. Keane's work on this project was supported in part by National Science Foundation Grant SBR-9511186.

References

Adamowicz, W., Louviere, J., Williams, M., 1994. Combining revealed and stated preference methods for valuing environmental amenities. Journal of Environmental Economics and Management 26, 271–292.

Ben-Akiva, M., Lerman, S., 1985. Discrete Choice Analysis: Theory and Application to Travel Demand. MIT Press, Cambridge.

Ben-Akiva, M., Morikawa, T., 1990. Estimation of switching models from revealed preferences and stated intentions. Transportation Research A 24A (6), 485–495.

Cafferata, G., 1984. Knowledge of their health insurance coverage by the elderly. Medical Care 22 (9), 838–847.

Davidson, B., Sofaer, S., Gertler, P., 1992. Consumer demand and biased selection in the demand for coverage supplementing Medicare. Social Science and Medicine 34 (9), 1023–1034.

Dowd, B., Moscovice, I., Feldman, R., Finch, M., Wisner, C., 1994. Health plan choice in the Twin Cities Medicare market. Medical Care 32 (10), 1–18.

Elrod, T., 1988. Inferring a product-market-map from panel data. Marketing Science 7, 21–40.

Elrod, T., 1991. Internal analysis of market structure: recent developments and future prospects. Marketing Letters 2, 253–266.

Erdem, T., 1996. A dynamic analysis of market structure based on panel data. Marketing Science 15 (4), 359–378.

Feldman, R., Finch, M., Dowd, B., Cassou, S., 1989. The demand for employer-based health insurance. Journal of Human Resources 23 (1), 115–142.

Geweke, J., Keane, M., Runkle, D., 1997. Statistical inference in the multinomial multiperiod probit model. Journal of Econometrics 80, 125–165.

Geweke, J., Keane, M., Runkle, D., 1994. Alternative computational approaches to statistical inference in the multinomial probit model. Review of Economics and Statistics 76, 609–632.

Hensher, D., Johnson, L., 1980. Applied Discrete Choice Modeling. Croom-Helm, London.

Hensher, D., Bradley, M., 1993. Using stated response choice data to enrich revealed preference discrete choice models. Marketing Letters 4 (2), 139–151.

Keane, M., 1992. A note on identification in the multinomial probit model. Journal of Business and Economic Statistics 10 (2), 193–200.

Keane, M., 1994. A computationally practical simulation estimator for panel data. Econometrica 62 (1), 95–116.

Lee, L.-F., 1995. Asymptotic bias in maximum simulated likelihood estimation of discrete choice models. Econometric Theory, 11, 437–483.

Lee, L.-F., 1997. Simulated maximum likelihood estimation of dynamic discrete choice statistical models: some Monte-Carlo results. Journal of Econometrics 82, 1–35.

Marquis, S., Kanouse, D., Brodsley, L., 1985. Informing consumers about health care costs: a review and research agenda. RAND Corporation, Santa Monica, CA.

McCall, N., Rice, T., Sangl, J., 1986. Consumer knowledge of Medicare and supplemental health insurance benefits. Health Services Research 20 (6), 633–657.

McFadden, D., 1974. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), Frontiers in Econometrics. Academic Press, New York, pp. 105–142.

McFadden, D., 1989. A method of simulated moments for estimation of discrete response models without numerical integration. Econometrica 57, 995–1026.

Revelt, D., Train, K., 1996. Incentives for appliance efficiency: random-parameters logit models of households' choices. Deptartment of Economics working paper, University of California-Berkeley.

Walker, A., 1990. Results of the Medicare beneficiary and physician focus groups. Lohr, K. (Ed.), In: Medicare: A strategy for Quality Assurance, vol. II, National Academy Press, Washington, DC, pp. 35–38.