A Model for Broad Choice Data

David Brownstone†  Phillip Li ‡

March 21, 2017

Abstract

This paper analyzes a discrete choice model where the observed outcome is not the exact alternative chosen by a decision maker but rather the broad group of alternatives in which the chosen alternative belongs to. The model is designed for situations where the choice behavior at a particular level is of interest but only broader level data are available. For example, consider analyzing a household’s choice for a vehicle at the make-model-trim level but only choice data at the make-model level are observed. The proposed model is a generalization of the multinomial logit model and collapses to it when there is full observability of the exact choices. We show that the parameters in the model are at least locally identified, but for certain configurations of the data, they are only weakly identified. Methods to address weak identification are proposed when there are data available on the overall market shares of all alternatives, and both maximum likelihood and Bayesian estimation methods are explored.

1 Introduction

Discrete choice models are usually estimated with data on the exact choices made by the decision makers from a well-specified choice set, as well as with observable attributes that are related to the

---

*We acknowledge funding provided by the University of California, Irvine through the Department of Economics, Institute of Transportation Studies, School of Social Sciences and Center for Economics and Public Policy, the University of California Center for Energy and Environmental Economics, the University of California Multi-Campus Research Program in Sustainable Transport: Technology, Mobility and Infrastructure, and the University of California Center on Economic Competitiveness in Transportation. This paper was started before Phillip Li joined the Office of Financial Research. The views and opinions expressed are those of the authors and do not necessarily represent official positions or policies of the Office of Financial Research or of the US Department of the Treasury. The authors are solely responsible for any errors or omissions.

†Department of Economics, University of California, Irvine, 3151 Social Science Plaza, Irvine CA 92697-5100. E-mail: dbrownst@uci.edu

‡Office of Financial Research, US Department of the Treasury, 717 14th St NW, Washington, DC 20005. E-mail: phillip.li@ofr.treasury.gov
choices, decision makers, or both. With these data standard discrete choice models like multinomial logit, probit, and generalized extreme value models can easily be estimated.

In contrast to this standard setting, our paper focuses on the situation where the econometrician does not observe the choices made by the decision makers at the level of interest, but rather only observes the broad groups of alternatives in which the chosen alternatives belong to. We refer to the choices at the original level of interest as exact choice data and the broader level group data as broad choice data. As a running example, suppose that it is of interest to model a household’s vehicle choice at the make-model-trim level where the choice set contains a Honda Civic LX, Honda Civic Hybrid, Toyota Camry LE, and Toyota Camry XLE Hybrid. Instead of observing the household’s exact choices from this four-vehicle choice set, the econometrician only observes the broad make-model group choices from the choice set, either from the Honda Civic or Toyota Camry group. The main objective of our paper is to only use the broad choice data to make inferences for the parameters belonging to the original exact choice data (e.g., alternative-specific constants).

There have been a few directions in the literature to address this data observability issue. From the statistics side, this type of data is referred to as either grouped data [Heitjan 1989; Gjeddebaek 1956a, b, 1961, 1949, 1957, 1959], partially categorized data [Blumenthal 1968; Nordheim 1984], or coarse data [Heitjan and Rubin 1990, 1991]. These three concepts are closely related and generally address the problem of only observing power sets from the sample space for the original random variable of interest. They differ only in terms of the type of random variable being analyzed: grouped data generally refer to observing interval data from continuous random variables, partially categorized data refer to observing set data from discrete random variables, and coarse data refer to observing general power set data from any random variable. Our definition of broad choice data is similar to partially categorized data and is similar to coarse data for discrete random variables.

Another direction of research is to redefine both the choice and attribute data into a common level of observability so that standard methods can be applied. For instance, in the vehicle choice example, the observable attributes at the make-model-trim level are either aggregated or averaged into the make-model level prior to estimation (e.g., average Honda Civic miles per gallon is used instead of specific trim-level fuel consumption). The attributes and choice data at the matching make-model level are then analyzed using standard discrete models. This approach has two major drawbacks. One is that using aggregate or average attributes will result in loss of precision for the parameter estimates when the members within a make-model level group are not homogenous with respect to their attributes. This is obvious since the miles per gallon ratings are significantly different between Honda Civic hybrids and non-hybrids, so averaging over this attribute within the make-model set will create measurement error which will lead to inconsistent parameter estimates. McFadden (1978) shows that if the distribution of attribute values being aggregated can be approx-
imated by a multivariate normal distribution, then this inconsistency can be removed by including the variances of the attributes within the group as well as the log of the number of alternatives in the group as additional explanatory variables. The second drawback is that, by averaging over the make-model-trim level attributes, there may not be enough variation to identify the parameters specific to the make-model-trim level, which is the level that we wish to make inferences in.

Multiple imputation is another direction of research. Intuitively, this approach imputes the exact choices from the original choice set of interest for each decision maker, estimates the model using the imputed exact choice data and attributes, and averages the parameter estimates over the numerous sets of imputed data. This is an attractive method since, given each set of imputed exact choices, standard discrete choice models can be used. Unfortunately a key requirement for multiple imputation estimators to be consistent is that the estimator must be consistent for each completed data set based on a single set of imputations. Unless the imputed alternative is the one actually chosen by the household, then the estimates on each completed set of data are not consistent.

In this paper, we propose a formal regression-based model for broad choice data that addresses the drawbacks in the current literature. In particular, our model is different than the work from the statistics literature in that it is a discrete choice model (i.e., the probabilities are based on utility maximization) and is a regression-based (i.e., attribute or covariate-based) model, while most of the current literature in statistics is based on general analysis of the data observability mechanism without covariates. In our framework, the broad choice data can be used together with the attributes at the exact choice level, avoiding the need of the previous literature to redefine data into a common level prior to estimation. The estimators we propose are either maximum likelihood or Bayes estimators, so they are fully efficient. And finally, our paper is unique in that we closely analyze the issue of identification when broad choice data are used over exact choice data. We show that the parameters in our model are locally identified, but for certain cases the parameters are only weakly identified (i.e., the likelihood function is almost completely flat). To address this weak identification issue, we introduce a novel technique to incorporate external information into the model in the form of parameter constraints or informative priors (in the Bayesian sense), and we also show how this information can be easily incorporated into maximum likelihood and Bayesian estimation routines. We only consider the case where the underlying choice model is conditional logit, but extensions to other discrete choice models are straightforward. Wong et al. (2017) provide Monte Carlo results showing that the broad choice model described in this paper performs much better than McFadden’s procedure, averaging over aggregated alternatives, or using a “representative alternative” in a realistic vehicle choice situation.

The paper proceeds as follows. The model for broad choice data is formally stated in Section 2, and the likelihood-based quantities are derived in Section 3. Using the quantities from the preceding
section, Section 4 discusses the identification issues associated with using the broad choice data. The details for maximum likelihood and Bayesian estimation of the parameters are discussed in Section 5 and Section 6 illustrates the various estimators on simulated data. Concluding remarks are in Section 7.

2 Model for broad choice data

The model specification is similar to that of a multinomial logit model and is based on random utility theory. Formally, the model is expressed as

\begin{equation}
U_{ij}^* = \delta_j + x_{ij}' \beta + \epsilon_{ij}, \quad \epsilon_{ij} \overset{i.i.d.}{\sim} \text{Type 1 Extreme Value},
\end{equation}

\begin{equation}
Y_i^* = j \quad \text{if} \quad U_{ij}^* \geq U_{ik}^* \quad \forall k \in C = \{1, 2, \ldots, J\},
\end{equation}

\begin{equation}
Y_i = m \quad \text{if} \quad Y_i^* \in C_m,
\end{equation}

for decision makers \(i = 1, \ldots, N\), alternatives \(j = 1, \ldots, J\), and groups \(m = 1, 2, \ldots, M\).

The latent utility that decision maker \(i\) obtains from alternative \(j\) is given by \(U_{ij}^*\) in (1). It is a function of an “average” level of utility that is constant for alternative \(j\) across all decision makers, \(\delta_j\), a column vector of \(K\) exogenous and observable attributes, \(x_{ij}\), a column vector of unknown coefficients, \(\beta\), and an unobserved error term, \(\epsilon_{ij}\), that is distributed i.i.d. Type 1 Extreme Value. For identification purposes, the average utility for the first alternative is normalized to zero (i.e. \(\delta_1 = 0\)). Choosing the first alternative for the normalization is inessential, as long as there is at least one such normalization.

In (2), the random variable \(Y_i^*\) denotes the exact alternative chosen by decision maker \(i\) from the choice set \(C\). The decision maker chooses alternative \(j\) if it provides the most utility among all the alternatives from the choice set. In (3), the variable \(Y_i\) represents the broad group of alternatives that \(Y_i^*\) is from. To be more specific about the values that \(Y_i\) can take, decompose each decision maker’s choice set into \(M\) groups such that \(C = \bigcup_{m=1}^{M} C_m\) and \(\bigcap_{m=1}^{M} C_m = \emptyset\). Then, \(Y_i\) equals the value \(m\) if the exact alternative chosen belongs to \(C_m\).

An important aspect of this paper is that only the outcomes for \(Y_i\) (and not \(Y_i^*\)) are observed. We refer to the observed outcomes for \(Y_i\) as the broad choice data, because they only broadly represent the exact choices made by the decision makers. For the running example, the choice set is partitioned into two groups. The first group, \(C_1\), contains the two Honda Civics, and the remaining group, \(C_2\), contains the two Toyota Camrys. The exact vehicle chosen by the household from \(C\) is not observed. Instead, we only observe either a 1 or 2, or equivalently, whether the exact vehicle chosen is a type of Honda Civic or Toyota Camry.
For the remainder of the paper, we refer to (1) to (3) as the model for broad choice data. Also, we refer to (1) to (2) with $Y_i^*$ observed for all decision makers as the model for exact choice data; this model is usually referred to as the conditional logit model in the literature. It is important to emphasize that the two models are equivalent when each group contains only a single alternative. To see their equivalence, note that $Y_i$ is equal to $Y_i^*$ for all decision makers when $|C_m| = 1$ for all groups, thus the two models are identical in this case.

3 Likelihood function and associated quantities

This section discusses the likelihood function of the sample, score function, and Hessian matrix of the log-likelihood function for the model with broad choice data assuming that the underlying choice model is conditional logit. The Hessian matrix is simple and provides insight into the likelihood function. It is also useful for the discussions on identification, information loss, and estimation in the subsequent sections.

Before discussing the likelihood function, some additional notation is needed. Define $\delta = (\delta_2, \delta_3, \ldots, \delta_J)'$, and let $\theta = (\delta', \beta')'$ be the parameter vector with $G = J - 1 + K$ elements. Also, define $w_{ij} = (z_j', x_{ij}')'$, where $z_j$ is a column vector of zeros and ones such that $w_{ij}' \theta$ is equal to the right hand side of the latent utility in (1). Intuitively, these vectors select out the appropriate average utility in $\theta$ for $U_{ij}^*$. As an illustration, with the $\delta_1 = 0$ normalization,

$$z_1 = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \quad z_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad z_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \quad \text{and} \quad z_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

when $J = 4$. In general, there are $J$ such vectors with $J - 1$ elements in each. They do not depend on $i$, because the composition of average utilities in (1) does not vary across decision makers.

The specifications from (1) to (3) imply that $Y_i$ takes a value of $m$ when one of the alternatives in group $m$ provides the highest level of utility among the alternatives in $C$. Since we do not observe the exact choice, the probability of observing $Y_i = m$ is equal to the probability that any alternative in $C_m$ may be the utility-maximizing alternative. Because these events are disjoint, this probability is equal to

$$\hat{P}_{im} = \Pr(Y_i = m),$$

$$= \Pr(Y_i^* \in C_m),$$

$$= \sum_{c \in C_m} P_{ic},$$

when the $\delta_1 = 0$ normalization is used.
If each probability within the summation is of the logit probability form, then

\[ P_{ic} = \Pr(Y_i^* = c) = \frac{\exp (w_i'c\theta)}{\sum_{j=1}^{J} \exp (w_{ij}\theta)}. \]  

(8)

The log-likelihood function of the sample is expressed as

\[ L_B(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \log(\hat{P}_{im}), \]  

(9)

where \( Y_{im} \) equals one if the outcome for decision maker \( i \) is equal to \( m \) and zero otherwise. The subscript \( B \) denotes a quantity corresponding to the model with broad choice data.

Differentiation of (9) with respect to \( \theta \) yields the score function

\[ S_B(\theta) = \frac{\partial L_B(\theta)}{\partial \theta} = \sum_{i=1}^{N} \left( \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} P_{ic|C_m} - \sum_{j=1}^{J} w_{ij} \hat{P}_{ij} \right), \]  

(10)

where

\[ P_{ic|C_m} = \frac{\exp (w_i'c\theta)}{\sum_{s \in C_m} \exp (w_s'\theta)}, \]  

(11)

and the Hessian matrix of the log-likelihood function

\[ H_B(\theta) = \frac{\partial^2 L_B(\theta)}{\partial \theta \partial \theta'} = L - F, \]  

(12)

where

\[ L = \sum_{i=1}^{N} \left( \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} (w_{ic} - \sum_{s \in C_m} P_{is|C_m} w_{is}) P_{ic|C_m} (w_{ic} - \sum_{s \in C_m} P_{is|C_m} w_{is})' \right), \]  

(13)

and

\[ F = \sum_{i=1}^{N} \left( \sum_{j=1}^{J} (w_{ij} - \sum_{r=1}^{J} P_{ir} w_{ir}) P_{ij} (w_{ij} - \sum_{r=1}^{J} P_{ir} w_{ir})' \right). \]  

(14)

The quantity in (11) is interpreted as the probability of decision maker \( i \) choosing alternative \( c \in C_m \) when the choice set is restricted \( C_m \). The derivations for the preceding quantities are given in the Appendix.

The analogous quantities for the model with exact choice data, which are denoted with the \( E \)
subscript, are given in McFadden (1973). The log-likelihood function of the sample is \( L_E(\theta) = \sum_{i=1}^{N} \sum_{j=1}^{J} Y_{ij}^* \log(P_{ij}) \), where \( Y_{ij}^* \) equals one if \( j \) is the exact alternative chosen by decision maker \( i \) and zero otherwise. The Hessian matrix of the log-likelihood, \( H_E(\theta) \), is equal to \(-F\).

There are two important characteristics of \( H_B(\theta) \) to emphasize. The first is that \( H_B(\theta) \) is not generally negative semidefinite. To see this, note that \( L \) and \( F \) are both positive semidefinite since they are equal to weighted moment matrices of the observed attributes, but the Hessian matrix, which is equal to the difference between these two matrices, does not need to be negative semidefinite. Hence, \( L_B(\theta) \) is generally not concave over the entire range of \( \theta \). On the other hand, McFadden (1973) shows that \( H_E(\theta) \) is negative semidefinite, so \( L_E(\theta) \) is concave in \( \theta \). Figure 1 confirms the shapes of the two log-likelihood functions for observed data with respect to a scalar parameter. Second, due to observing the broad choices instead of the exact choices, \( L_B(\theta) \) generally has less curvature than \( L_E(\theta) \). Both Figures 1 and 2 illustrate this fact. The diminished curvature has consequences on the identification of the parameters, which is discussed in the subsequent section.

4 Identification

Identification is assessed by analyzing whether the information matrix is nonsingular. The results in this section utilize Theorem 1 from Rothenberg (1971). Using the notation from the previous section, the theorem states that \( \theta \) is locally identified if and only if the information matrix \( \Pi_B(\theta) = -E(H_B(\theta)) \) is nonsingular, or equivalently, has rank \( G \) (the number of elements in \( \theta \)).

The information matrix corresponding to (9) is equal to

\[
\Pi_B(\theta) = -E(H_B(\theta)),
\]

\[
= F - IL,
\]

\[
= \Pi_E(\theta) - IL,
\]

where \( F = \Pi_E(\theta) \) from McFadden (1973), and

\[
IL = \sum_{i=1}^{N} (\sum_{m=1}^{M} \hat{P}_{im} \sum_{c \in C_m} (w_{ic} - \sum_{s \in C_m} P_{is|C_m} w_{is})P_{ic\mid C_m} (w_{ic} - \sum_{s \in C_m} P_{is\mid C_m} w_{is})'),
\]

This derivation is discussed in the Appendix. The expression for the information matrix in (17) has an intuitive form: it is equal to the difference between the information matrix with exact choice data and \( IL \). Loosely speaking, when viewed asymptotically, \( IL \) quantifies the econometrician's
Figure 1: Log-likelihood functions with respect to a scalar parameter. The function corresponding to exact choice data is concave everywhere, but the one for broad choice data is not globally concave.

"information loss" from using broad choice data instead of exact choice data, in the sense that $IL = \mathbb{I}_E(\theta) - \mathbb{I}_B(\theta)$ is positive definite unless $\mathbb{I}_E(\theta) = \mathbb{I}_B(\theta)$. By analyzing (18), the main factors that determine $IL$ are the observed attributes, choice probabilities, and most importantly, the sizes of the groups which determine the number of outer products being summed in (18).

We will discuss identification of the parameters using the information matrix for three distinct cases. The first case corresponds to each group having only one alternative (i.e. $|C_m| = 1$ for all groups) which results in the model for exact choice data. It is well known that $\theta$ is identified when $w_{ij}$ varies across alternative sets and is not collinear with the other attribute vectors (McFadden, 1973). Also, as expected, $\mathbb{I}_B(\theta)$ reduces to $\mathbb{I}_E(\theta)$ in this case. To see this, let $t_m$ be the only element in $C_m$, which implies that $P_{is|C_m} = 1$ for all $s \in C_m$. Then,

$$IL = \sum_{i=1}^N \left( \sum_{m=1}^M \hat{P}_{itm}(w_{itm} - \hat{w}_{itm})(w_{itm} - \hat{w}_{itm})' \right),$$

(19)

where $0_{(G \times G)}$ denotes a $G \times G$ matrix of zeros, and thus $\mathbb{I}_B(\theta) = \mathbb{I}_E(\theta)$.

The second case, presumably the most typical with broad choice data, occurs when the sizes
of the groups are strictly less than $J$. For this case, $\theta$ is locally identified, because given the aforementioned assumptions on $w_{ij}$, the information matrix is full rank. It is interesting to note that $IL$ is rank deficient since it has at most rank $G - M$, but subtracting it from $I_E(\theta)$ does not decrease the rank of $I_B(\theta)$. Local identification is achieved from the nonlinearities in the distributional form. Global and non-parametric identification are not analyzed in this paper and are topics for future research, but it is clear that the alternative specific constants, $\delta$, can only possibly be non-parametrically identified for alternatives that are fully observed (part of a group with only one member).

Although $\theta$ is locally identified in the second case, there are scenarios in which the broad choice data do not contain much information about the parameters, leading to either imprecise estimates during estimation or a nearly singular information matrix. As alluded to earlier, the amount of information in (15) about $\theta$ decreases as the sizes of the groups increase. One extreme scenario in the second case occurs when there are a large number of alternatives and the size of a particular group is close to the size of $C$. For these group configurations, $IL$ is “close” to $I_E(\theta)$, and the information matrix is nearly singular, leading to weak identification of $\theta$. By weak identification, we mean that although $\theta$ is locally identified, the information matrix is close to being singular.
This is consistent with Figure 2 which depicts that, when a particular group is close in size to $C$, $L_B(\theta)$ is almost flat relative to $L_E(\theta)$ with respect to a scalar parameter.

The third case occurs when there is a single group that is equal to $C$ in size, or equivalently, when every observed outcome for the broad choice data corresponds to the entire choice set. When this case occurs, $\theta$ is not identified, because the information matrix is rank deficient. To see this, assume that $m = M = 1$, which implies that $C_m = C$, $\bar{P}_m = 1$ for all decision makers and groups, and $P_{is|C_m} = P_{is}$ for all $s \in C$ and decision makers. Then

$$IL = \sum_{i=1}^{N} \left( \sum_{c \in C} (w_{ic} - \sum_{s \in C} P_{is}w_{is})P_{ic}(w_{ic} - \sum_{s \in C} P_{is}w_{is})' \right)^{1},$$

and therefore $I_B(\theta) = 0_{(G \times G)}$ is rank deficient. This result establishes that only knowing a decision maker’s choice belongs to the full choice set is not sufficient to discern the possible values of $\theta$.

5 Estimation

This section describes maximum likelihood (ML) and Bayesian estimation of $\theta$. Throughout the entire discussion, we assume the second case mentioned in Section 4 which implies that $\theta$ is at least locally identified. But for scenarios in which the broad choice data are not informative about $\theta$ (see Section 4 for an example), we propose incorporating additional information in the form of population market shares into the problem.

For ML estimation, the market share information is implemented as constraints on $\theta$. To begin the discussion, assume that the population market shares for alternatives 2 through $J$ are known. These market shares, denoted by $s_j$ for $j = 2, 3, \ldots, J$, are defined as the percentage shares of the population choosing each alternative. They are collectively denoted by $s$ and are related to the parameters by the nonlinear market share constraints $N^{-1} \sum_{i=1}^{N} P_{ij} = s_j$ for $j = 2, 3, \ldots, J$. The population shares are informative for the parameters, because for a well-specified model, the predicted market shares from a large representative sample should equal the population shares on average. In other words, the predicted market shares are unbiased and consistent estimates for the population market shares.

There are two important assumptions concerning these constraints. The first is that the predicted in-sample market shares, $N^{-1} \sum_{i=1}^{N} P_{ij}$, must equal the population market shares. As mentioned earlier, this assumption should be met for a well-specified model (with a full set of alternative-specific constants) estimated using a large representative sample on average, but equal-
ity in small samples is uncertain. The second assumption is that the population market shares are known with certainty and hence fixed. This assumption may not always hold as the population market shares may be measured with error or only known with a certain degree of certainty to the researcher. For the vehicle choice problem, overall market shares are known, but it is very difficult to separate out purchases by fleets (estimated to be about 20% of all light duty vehicle sales) from household purchases. But for the purposes of ML estimation, we assume that these assumptions are not violated.

To account for possible violations in the two preceding assumptions, we propose using Bayesian methods. In contrast to the ML methods, the proposed Bayesian methods formally account for the uncertainty in the market shares, and they do not strictly enforce the constraints onto the parameters. Instead, the constraints, which contain information for $\theta$ through the market shares, are only used to construct an informative prior for $\theta$. As a result, the prior can be used to indirectly reflect uncertainty for these constraints or equivalently for the two preceding assumptions. This is discussed in subsequent sections.

In summary, for each estimation method, we discuss how it can be implemented with or without incorporating information from the market share. Because the former case is relatively non-standard, a majority of the discussion is dedicated to it.

5.1 Maximum likelihood estimation

To enforce the constraints in an ML estimation routine, we use a result from Berry et al. (1995). They proved that, conditional on $\beta$ and $s$, the constraints are one-to-one mappings that relate $\delta$ to $s$ and $\beta$. Thus, assuming that the constraints can be inverted to solve for $\delta$ as a function of $\beta$, which we denote as $\delta(\beta)$, the maximum likelihood estimate (MLE) is obtained by maximizing the log-likelihood function for the observed sample

$$L_B(\beta, \delta(\beta)) = \sum_{i=1}^{N} \sum_{m=1}^{M} y_{im} \log(\tilde{P}_{im})$$ (22)

with respect to $\beta$, where $y_{im}$ is the observed value for $Y_{im}$, and the $\delta$ typically in $\tilde{P}_{im}$ is replaced with $\delta(\beta)$. It is tempting to interpret this approach as concentrating $\delta$ out of the likelihood, but the constraints (conditional on $\beta$ and $s$) are not the first order conditions of (9) with respect to $\delta$. As such, this approach is only used to enforce the constraints.

The main difficulty of obtaining the MLE is in inversion of the constraints. An analytic inverse is not obvious, so numerical methods must be used. We use the iterative techniques from Li (2012) which are more computationally efficient than the ones presented in Berry et al. (1995) by an order
of magnitude. The techniques rely on solving the market share constraints for a fixed point using an accelerated iterative system. When the system is iterated enough times, the iterates will converge to a unique vector for $\delta$ that solves the market share constraints.

The iterative system is given by

$$\delta^{(k+1)} = \delta^{(k)} + h(\delta^{(k)}) f(\delta^{(k)}),$$

where the $(k+1)$ and $(k)$ superscripts respectively indicate the $k+1$ and $k$-th iterations of the system. The step-size matrix $h$ is chosen from one of five forms in Table 1 depending on the practitioner’s desired stability and numerical performance of the system (see Li (2012) for a thorough discussion). The vector-valued function $f$ has elements of the form

$$\log \left( \frac{s_j}{\frac{1}{N} \sum_{i=1}^{N} P_{ij}} \right), \quad j = 2, \ldots, J,$$

(24)

where $\beta$ is fixed in each $P_{ij}$. To gain some intuition for (23) and (24), $f$ can be interpreted as an adjustment term when the step-size matrix is equal to an identity matrix. When $\delta^{(k)}$ results in predicted in-sample market shares that are too small relative to every element in $s$, then (24) is positive and the next value $\delta^{(k+1)}$ is equal to $\delta^{(k)}$ adjusted positively by (24). In turn, the positively-adjusted vector $\delta^{(k+1)}$ increases the predicted in-sample market shares relative to the last iteration. On the other hand, negative adjustments are produced when the predicted in-sample market shares are larger than the population shares. Theoretically, these adjustments continue until (24) equals zero, or equivalently, until the value of $\delta$ that sets the predicted in-sample market shares equal to the population market shares is found. However, because equality in (23) may not be possible due to machine precision, the convergence criterion is $\| \delta^{(k+1)} - \delta^{(k)} \| < c$, where $c$ is set to $10^{-14}$ or smaller following Dube et al. (2011). With the preceding intuition in mind, the different step-size matrices in Table 1 are used to accelerate this adjustment process.

The preceding iterative system is nested in gradient-based methods to obtain the MLE for $\theta$. The maximization algorithm searches over the space for $\beta$, and for each trial value, (23) is used to recover $\delta(\beta)$. The values of $\beta$ and $\delta(\beta)$ that jointly maximize (22) are the maximum likelihood estimates, denoted by $\widehat{\beta}_{MLE}$ and $\widehat{\delta}_{MLE}$. It is important to note that standard maximization algorithms in statistical software (e.g. Matlab, Gauss, etc.) will only output $\widehat{\beta}_{MLE}$ and $\widehat{H}(\widehat{\beta}_{MLE})$, the estimated Hessian matrix evaluated at the maximizing value, because $\beta$ is the only input to the algorithm. So, to obtain $\widehat{\delta}_{MLE}$, apply the iterative system in (23) to $\widehat{\beta}_{MLE}$ after the maximization algorithm has finished. And using a similar approach, the approximate covariance matrix for $\widehat{\theta}_{MLE} = (\widehat{\beta}_{MLE}, \widehat{\delta}_{MLE})$ is obtained numerically by sampling $\beta(g)$ from $N(\widehat{\beta}_{MLE}, -\widehat{H}(\widehat{\beta}_{MLE})^{-1})$
for \( g = 1, 2, \ldots, G \) draws, recovering \( \delta(\beta(g)) \) from (23) for each draw, and computing the sample covariance matrix using the collection of \((\beta(g), \delta(\beta(g)))\) vectors. This estimated covariance matrix, denoted by \( \hat{T} \), converges to the desired quantity when \( G \) is large. This parametric bootstrap procedure for estimating the covariance matrix is an alternative to the Generalized Method of Moments approach advocated by Nevo (2000) for the BLP model with full observability.

When the market share constraints do not need to be enforced, the MLE for \( \theta \) is the vector that maximizes (9) for observed data with respect to \( \theta \). The log-likelihood function is easily maximized with gradient-based algorithms since the analytic score function and Hessian matrix are given in (10) and (12), respectively. However, caution must be taken during implementation, because the log-likelihood function is not necessarily concave in \( \theta \) and may have almost-flat spots.

5.2 Bayesian estimation

5.2.1 Priors and posterior distribution

For Bayesian analysis, the model from (1) to (3) is completed with specifications for the prior distributions of the parameters. We will first discuss priors that incorporate the market share information and then describe priors without this information at the end of the section. The prior for \( \theta \), denoted by \( \pi(\theta | \mu) \), is multivariate normal with mean vector \( b \) and covariance matrix \( B \), and it depends on a hyperparameter \( \mu = (\mu_1, \mu_2, \ldots, \mu_J)' \) that contains the unknown market shares for all the alternatives in \( C \). Uncertainty for \( \mu \) is expressed in a hyperprior \( \pi(\mu) \). The posterior density of interest with market share information is given by

\[
\pi(\theta, \mu | Data) \propto L_B(\theta)\pi(\theta | \mu)\pi(\mu),
\]

where \( Data \) contains the observed broad choice data.

Uncertainty for the market shares is summarized in a hyperprior \( \pi(\mu) = D(\mu | \alpha) \), where

\[
D(\mu | \alpha) = \left( \frac{\Gamma(a_0)}{\prod_{j=1}^J \Gamma(\alpha_j)} \right) \prod_{j=1}^J (\mu_j)^{\alpha_j-1}
\]

is a standardized multivariate Dirichlet density that depends on a vector of parameters \( \alpha = (\alpha_1, \alpha_2, \ldots, \alpha_J)' \). In (26), \( a_0 = \sum_{j=1}^J \alpha_j \), and each \( \alpha_j \) is strictly positive. The Dirichlet prior is convenient for two reasons. First, this prior implies that the market shares are bounded in \((0, 1)\) and must sum to unity. Second, prior information for the market shares can be easily incorporated through \( \alpha \). As an example, suppose that we have prior information in terms of market shares from a previous time period. Then, one approach to incorporate this knowledge into the problem is to
set the elements in $\alpha$ such that the prior means are centered around the previous shares and the prior variances reflect our uncertainty about this knowledge. To find these parameters, recall the properties of the Dirichlet distribution: $E(\mu_j) = \alpha_j/a_0$, and $V(\mu_j) = \alpha_j(a_0 - \alpha_j)/a_0^2(a_0 + 1)$ for $j = 1, 2, \ldots, J$. Once a value for $a_0$ is chosen, then the elements of $\alpha$ that set the desired means and variances can be solved for. In general, larger values for $a_0$ result in smaller variances.

Information about $\theta$ from the market shares is incorporated into $\pi(\theta|\mu)$. We propose specifying the normal prior around the approximate distribution for $\hat{\theta}_{MLE}$ from a training sample. Specifically, set aside a quarter of the $N$ observations as the training sample, and using only the training sample, estimate $\theta$ with the ML methods from the previous section subject to the market share constraints where $s$ is replaced with $\mu$. We want to note that there is a disagreement in the literature on the appropriate size of the training sample. We therefore suggest choosing a sample size such that the training sample is representative of the population and that the sample is large enough to assume that the market share constraints are reasonable. Then, set the prior mean $b$ equal to $\hat{\theta}_{MLE}$ and the covariance matrix $B$ equal to $\hat{T} + \tau \times \text{diag}(\hat{T})$, where $\text{diag}(\hat{T})$ is a diagonal matrix with the elements of $\hat{T}$ on its main diagonal. The scalar variance inflation factor $\tau$ is used to control the degree of uncertainty in this prior and can also be used for prior sensitivity analysis. The advantages of using this prior specification are discussed next.

Important differences in the Bayesian analysis of the model must be stressed. First, the market shares are treated as unknown quantities, and their uncertainty is reflected in the hyperprior. Being able to express uncertainty in the population shares is important since they are extremely difficult to obtain in practice, and even if they are available, they may be measured with error. For example, in the case of vehicle sales, even “official” market share data are often convoluted with sales to car rental companies, leasing companies, government agencies, and businesses, so the observed shares may not accurately reflect the actual market shares among individual decision makers.

Second, the market share constraints, which relate $\theta$ to $\mu$, are not strictly enforced onto the parameters as they are in ML estimation. Instead, the information from $\mu$ is incorporated through the prior hyperparameters for $\theta$. Loosely interpreted, with a training sample that is representative of the remaining sample used for inference, the prior for $\theta$ is centered around the parameter vector that sets the predicted in-sample market shares equal to $\mu$, and the variance inflation factor $\tau$ quantifies our certainty around this vector. As an example, if $\tau$ and $\pi(\mu)$ are tightly specified, then our prior knowledge for $\theta$ is highly concentrated around the vector that satisfies the constraints implied by $\mu$, which is similar to enforcing the constraints into the model. This uncertainty is important to quantify since the constraints do not need to hold with equality in small sample, and any parameter values that are recovered by strictly enforcing the constraints may be misleading.

For the case without incorporating market share information, let $\pi(\theta)$ be a normal prior for $\theta$
with mean vector \( d \) and covariance matrix \( D \). The resulting posterior distribution is \( \pi(\theta | Data) \propto L_B(\theta)\pi(\theta) \). Note that although the market share information is not incorporated in the prior, we may learn about the parameters as long as the priors for \( \beta \) and \( \delta \) are dependent. Note that such dependencies arise naturally in models where there are interactions between the exogenous variables \( x_{ij} \) and the alternative specific constants. In the vehicle example it would make sense to include interactions between alternative specific constants for hybrid vehicles and income and education.

### 5.2.2 Markov chain Monte Carlo algorithm

Bayesian estimation is performed by Markov chain Monte Carlo (MCMC) methods with Metropolis-Hastings (MH) steps. Broadly speaking, this method generates samples from the posterior distribution by first proposing candidate values from a known proposal distribution and then accepting them with a specific MH probability. If a proposed candidate value is rejected, then the previous value is used instead. Because the same values may be repeatedly used, this method constructs a Markov chain. After a sufficient burn-in period, the draws from the constructed Markov chain are from the posterior distribution of interest by MH convergence results (Chib and Greenberg, 1995; Tierney, 1994). The final set of posterior draws is then used to construct quantities of interest (e.g. posterior means, standard deviations, etc.).

The MCMC algorithm is described first for the case with market share information. At iteration \( t \), candidate values for \( \beta, \delta, \) and \( \mu \) are proposed as follows

1. Draw \( \mu^{(t)} \) from \( q_1(\mu) = D(\mu | \alpha) \).

2. Given \( \mu^{(t)} \), maximize \( L_B(\theta)\pi(\theta | \mu^{(t)}) \) with respect to \( \theta \). Denote \( \hat{\theta} \) as the maximizing value and \( -\hat{H}(\hat{\theta})^{-1} \) as the negative inverse of the estimated Hessian evaluated at \( \hat{\theta} \).

3. Draw \( \theta^{(t)} \) from \( q_2(\theta | \mu) = f_t(\hat{\theta}, -\hat{H}(\hat{\theta})^{-1}, \nu) \), which is a \( t \) density with location parameter \( \hat{\theta} \), scale matrix \( -\hat{H}(\hat{\theta})^{-1} \), and degrees of freedom parameter \( \nu \), which is set to a small number to ensure heavy tails for this distribution.

The vector \( \eta^{(t)} = (\beta^{(t)}, \delta^{(t)}, \mu^{(t)}) \) constitutes a proposed candidate draw from the posterior distribution in (25), where the proposal density is given by

\[
q_3(\eta) = q_1(\mu)q_2(\theta | \mu).
\]  

(27)

Upon denoting the right hand side of (25) as \( p_1(\eta) \), the candidate vector is accepted for iteration
t with the MH transition probability
\[
\alpha_{\text{MH}}(\eta^{(t-1)}, \eta^{(t)}) = \min \left\{ 1, \frac{p_1(\eta^{(t)}) q_3(\eta^{(t-1)})}{p_1(\eta^{(t-1)}) q_3(\eta^{(t)})} \right\}.
\] (28)

If the candidate vector is not accepted, then treat \(\eta^{(t-1)}\) as the draw for iteration \(t\). This iterative process is repeated many times and the final collection of vectors represents the draws from the posterior distribution with market share information.

A few subtle points about the preceding algorithm are now noted. First, the iterative system from (23) is nested into each MCMC iteration. It is needed in Step 2 of the previous algorithm to compute the parameters in \(\pi(\theta|\mu^{(t)})\). Despite being nested in the algorithm, the iterative system is inexpensive to evaluate, because the previous draws of \(\eta\) can be used as the starting values to the system. This method significantly reduces the number of iterations needed for the iterative system to converge (Li, 2012). Second, if the maximization in Step 2 is difficult to perform, then a random walk proposal for \(q_2(\theta|\mu)\) is a viable alternative. And third, the candidate vectors for \(\mu\) are generated from the prior density in (26) instead of a density that is proportional to the posterior distribution. This method of generating candidate vectors may be inefficient in the sense that a large portion of the candidates for \(\mu\) may not be accepted, but its performance is quite good when (26) is tightly specified. Otherwise, valid draws of \(\mu\) from the posterior are incredibly difficult to obtain due to the support restrictions on the market shares.

The MCMC algorithm for the case without market share information is similar. Before starting the algorithm, maximize \(p_2(\theta) = L_B(\theta)\pi(\theta)\) with respect to \(\theta\), and denote \(\hat{\theta}\) and \(-\hat{H}(\hat{\theta})^{-1}\) as the maximizing value and negative inverse of the estimated Hessian evaluated at \(\hat{\theta}\), respectively. Then, at step \(t\) of the algorithm, candidate vectors for \(\theta^{(t)}\) are proposed from \(q_4(\theta) = f_t(\theta|\hat{\theta}, -\hat{H}(\hat{\theta})^{-1}, \nu)\) and accepted with the transition probability
\[
\alpha_{\text{MH}}(\theta^{(t-1)}, \theta^{(t)}) = \min \left\{ 1, \frac{p_2(\theta^{(t)}) q_4(\theta^{(t-1)})}{p_2(\theta^{(t-1)}) q_4(\theta^{(t)})} \right\}.
\] (29)

If the candidate vector is not accepted, then set \(\theta^{(t)} = \theta^{(t-1)}\).

6 Simulation results

This section applies the estimation methods developed in Section 5 to simulated data. The results are used to compare the different estimators and to highlight some key points regarding the inclusion of market share information. For this simulation study, the maximum likelihood and
Bayesian estimators are analyzed in both repeated sample and single sample settings.

We use 250 repeated samples. For each repeated sample, we simulate a population of 20,000 decision makers based on (1) to (3). The first exogenous attribute, $x_{i1}$, is discrete and takes values of either 1, 2, or 3. The second attribute, $x_{i2}$, is normally distributed, and the third attribute is an interaction term between $x_{i2}$ and the fourth alternative. The broad choice data and attributes are resimulated for each repeated sample, while the data-generating values for $\theta$, presented in Table 2 and in other tables, are fixed. There are ten alternatives in the choice set, and the broad groups are defined as $C_1 = \{1, 2, \ldots, 9\}$ and $C_2 = \{10\}$. Note that these group configurations are constructed so that the observed broad choice data are fairly uninformative for $\theta$. There are 12 parameters to estimate: $\theta = (\delta_2, \delta_3, \ldots, \delta_{10}, \beta_1, \beta_2, \beta_3)'$.

For each repetition, we calculate the population market shares based on the 20,000 decision makers as the average probability shares evaluated at the true parameter values for each alternative. The population shares for alternatives one through ten are roughly 5%, 9%, 6%, 41%, 12%, 3%, 4%, 6%, 8%, and 6%, respectively, and the in-sample market shares from each repetition closely mimic the ones from the population. Note that this data generation process does not enforce the market share constraints that are imposed by the BCDC estimator. We set aside the first 5,000 observations to use as a training sample for the Bayesian informative priors and use the remaining 15,000 observations for estimation.

The ML estimates corresponding to $\theta$ are presented in Table 2. Both ECD and BCD estimates are close to their true values, but there is substantially more variability when BCD is used instead of ECD. In particular, the standard errors corresponding to $\beta$ and $\delta$ roughly differ by a factor of four. This difference in variability is expected since there is generally less information in the broad choice data than in the exact choice data. In addition, this difference is amplified by the specified group configurations. Figure 3 illustrates the sampling distributions for the estimators of $\delta_2$ and $\beta$ (the distributions for the other estimators are omitted since they are qualitatively similar to the one for $\delta_2$) and confirms that the distributions based on BCD are wider than the ones based on ECD.

With the population market share constraints enforced in BCDC, the resulting ML estimates are generally close to their true parameter values. In terms of variability, the standard errors that correspond to $\delta$ are substantially smaller than the errors obtained with ECD and BCD. The exception is $\delta_4$ which is interacted with the continuous variable $x_{i2}$ in the design. This suggests that the constraints are helpful in pinning down the estimates of $\delta$. On the other hand, the standard error corresponding to $\beta$ does not differ from one obtained using BCD, which is around four times larger than the one from ECD. This suggests that the constraints do not contain much information with regards to $\beta$. Figure 3 confirms the first observation as the sampling distribution corresponding
Figure 3: Sampling distributions for the maximum likelihood estimators of $\delta_2$ and $\beta_3$. These are based on exact choice data (ECD), broad choice data (BCD), and broad choice data with population market share constraints (BCDC).
Figure 4: Posterior distributions for $\delta_2$ and $\beta_3$. These are based on exact choice data (ECD), broad choice data (BCD), and broad choice data with an informative prior (BCDIP). The informative prior has $\tau = 0.005$.

To $\delta_2$ is highly concentrated around the true value of $-0.68$ when using BCDC, even more so than the ones for ECD and BCD. The same figure also confirms that the distributions corresponding to $\beta$ are almost identical between ECD and BCDC.

Table 3 gives coverage probabilities for 80% nominal confidence intervals computed for each estimator in Table 2 across the 250 Monte Carlo repetitions. These coverage probabilities are close to their nominal values for all of the estimators except for $\delta$s for the BCDC estimator. Recall that the BCDC estimator imposes the constraint that the predictions from the model match the population market shares, but this constraint is not imposed in the data generation process. These constraints only affect the $\delta$s in the conditional logit model, so the $\beta$ confidence intervals are still valid as shown in Table 2 for the BCDC estimator.

Bayesian estimation is also based 250 repeated samples. Similar to the ML discussion, we present the posterior means and standard deviations for $\theta$ across three cases. The first two cases respectively correspond to ECD and BCD with priors that are fairly non-informative. For these priors, we set $b = 0_{(G \times 1)}$ and $B = 1000 \times I_{(G \times G)}$. The remaining case is based on broad choice data with an informative prior for $\theta$ (BCDIP). As discussed in Section 5.2.1, this prior is developed
Figure 5: Posterior distributions for $\delta_2$ and $\beta_3$. These are based on exact choice data (ECD), broad choice data (BCD), and broad choice data with an informative prior (BCDIP). The informative prior has $\tau = 0.1$. 
using the remaining 5,000 observations as a training sample and the population market shares. The variance inflation factor $\tau$ is set to 0.1 which slightly inflates the prior variances for $\theta$. Note that the known population market shares are used in this simulation exercise, so $\mu$ is known with certainty. And as a result, the hyperprior $\pi(\mu)$ does not need to be specified. The prior values for $b$ and the standard deviations implied by $B$ are in Tables 4 and 6. The off-diagonals of $B$ are close to zero and are not reported.

Table 4 contains the Bayesian estimates based on 10,000 runs of the MCMC algorithm with 2000 runs discarded for burn-in. The numerical results corresponding to ECD are similar to those from ML estimation. That is, when ECD is used, the posterior means are quite close to the true values of the parameters used to generate the data, and the standard deviations are fairly tight and similar to those derived from maximum likelihood. This suggests that the exact choice data are very informative about the parameters. For the case with BCD, the posterior means are in the neighborhood of the true values, but the posterior standard deviations are relatively large compared to the ones from ECD. Similar to the ML study, the posterior standard deviations for $\beta$ and $\delta$ differ by factors of four and eight, respectively. The posterior distributions depicted in Figure 4 confirm these observations.

In the case of BCDIP, the posterior means for $\delta$ are generally closer to the true values than the ones obtained using BCD, and the standard deviations are relatively smaller. For $\beta$, the posterior estimates do not change much over the ones obtained using BCD, which is a similar conclusion to the ML study. These observations are confirmed by Figure 4. From the figure, we see that the posterior distribution for $\delta_2$ has less variability in this case relative to BCD. Also, note that this distribution for $\delta_2$ is not as highly concentrated as the one from ML estimation. This additional variability comes from the prior which expresses some uncertainty in the constraints. For $\beta$, the figure confirms that the posterior distributions are almost identical. These Bayesian estimates suggest two conclusions. The first one is that the posterior distribution for $\delta$ is sensitive to the prior distribution despite the large sample size. There is little difference between the prior and posterior standard deviations for $\delta$. This is expected since the broad choice data in this simulation exercise are constructed to be uninformative about $\theta$. The second conclusion is that the broad choice data are informative for $\beta$, because the posterior standard deviation are about one half of the prior standard deviations for these parameters.

Table 5 examines the repeated sampling properties of the Bayesian estimators. The coverage probabilities for the ECD and BCD estimators are close to their nominal 80% values similar to the performance of the maximum likelihood estimators in Table 3. The coverage probabilities for

---

1For these simulations, we set $B$ to $\hat{T} + \tau \times I_{G \times G}$ instead of $\hat{T} + \tau \times \text{diag}(\hat{T})$.  

21
δ are below the nominal 80% value, but better than the maximum likelihood estimator (BCDC in Table 3) that imposes the exact market share constraints. If the prior distribution is more diffuse, then Table 6 shows that the posterior standard deviations are all substantially lower than the prior standard deviations. Table 7 shows that with the less informative prior the posterior confidence intervals for δ now lead to coverage probabilities slightly above the nominal 80% level.

A few conclusions for the entire simulation study are in order. First, the market shares in the form of constraints or prior information contain more information about δ than β. Second, the addition of market share information improves both estimation methods over the case without incorporating this information. However, we must warn that the effectiveness of this technique critically depends on the quality of the market shares and the sample size. If the population shares do not closely mimic the predicted in-sample shares, then inference is questionable. This issue is especially problematic for ML estimation, because the parameters are recovered using strict constraints based on the shares. In contrast, for the Bayesian methods, a large value for τ can be used to express uncertainty in this technique. At worst, the resulting prior is relatively non-informative, and we obtain results similar to the Bayesian estimates with BCD. Lastly, for some extreme group configurations, the broad choice data will not be informative for θ (especially δ). For these cases, the Bayesian estimates are highly sensitive to the prior. But despite this strong dependence, the numerical results demonstrate that, as long as the practitioner is thoughtful in forming the prior, the results are well behaved.

7 Conclusion

This paper introduces a new discrete choice model to analyze choice outcomes that only broadly represent the actual choices made by the decision makers. It is useful in analyzing situations where the choice behavior at a lower level is desired but only higher level choice data are observed. The parameters from the proposed model are locally identified, but in some perverse yet interesting cases, they may only be weakly identified. To efficiently recover the parameter estimates in these troublesome cases, we show how population-level market shares can be introduced as additional information into the problem. A simulation study shows that both maximum likelihood and Bayesian estimation techniques benefit from the inclusion of the market share information. Although the effectiveness of this approach depends critically on the quality of the population market shares, the results demonstrate that meaningful relationships can be uncovered using this new class of models.
Appendices

A Likelihood quantities and the information matrix

This section derives the score function and Hessian matrix of the log-likelihood function for the model with broad choice data. To obtain the score function, expand (5) in terms of (8), resulting in the log-likelihood function

\[ L_B(\theta) = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \left( \log( \sum_{c \in C_m} \exp(w_{ic}^t \theta)) - \log(\sum_{j=1}^{J} \exp(w_{ij}^t \theta)) \right). \]  \hspace{1cm} (30)

The score function is then

\[ S_B(\theta) = \frac{\partial L_B(\theta)}{\partial \theta}, \]   \hspace{1cm} (31)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \left( \sum_{c \in C_m} \frac{\partial}{\partial \theta} \log(\sum_{s \in C_m} \exp(w_{ic}^t \theta)) - \sum_{j=1}^{J} \frac{\partial}{\partial \theta} \log(\sum_{r=1}^{J} \exp(w_{ij}^t \theta)) \right), \]  \hspace{1cm} (32)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \left( \sum_{c \in C_m} \frac{w_{ic}}{\sum_{s \in C_m} \exp(w_{is}^t \theta)} \frac{\partial}{\partial \theta} \log(\sum_{s \in C_m} \exp(w_{is}^t \theta)) - \sum_{j=1}^{J} w_{ij} \frac{\partial}{\partial \theta} \log(\sum_{r=1}^{J} \exp(w_{ir}^t \theta)) \right), \]  \hspace{1cm} (33)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \left( \sum_{c \in C_m} w_{ic} P_{ic|C_m} - \sum_{j=1}^{J} w_{ij} P_{ij} \right), \]  \hspace{1cm} (34)

\[ = \sum_{i=1}^{N} \left( \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} P_{ic|C_m} - \sum_{m=1}^{M} Y_{im} \sum_{j=1}^{J} w_{ij} P_{ij} \right), \]  \hspace{1cm} (35)

\[ = \sum_{i=1}^{N} \left( \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} P_{ic|C_m} - \sum_{j=1}^{J} w_{ij} \left( \sum_{m=1}^{M} Y_{im} \right) P_{ij} \right), \]  \hspace{1cm} (36)

\[ = \sum_{i=1}^{N} \left( \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} P_{ic|C_m} - \sum_{j=1}^{J} w_{ij} \left( \sum_{m=1}^{M} Y_{im} \right) P_{ij} \right), \]  \hspace{1cm} (37)

since \( \sum_{m=1}^{M} Y_{im} = 1 \) for all decision makers.

Before deriving the Hessian matrix, a few miscellaneous quantities are needed. Note that

\[ \frac{\partial P_{ij}}{\partial \theta'} = P_{ij}(w_{ij'} - \sum_{r=1}^{J} w_{ir'} P_{rr}), \]  \hspace{1cm} (38)
and
\[
\frac{\partial P_{ic|C_m}}{\partial \theta'} = P_{ic|C_m} (w'_ic - \sum_{c \in C_m} w'_ic P_{ic|C_m}), \quad (39)
\]

Also,
\[
\sum_{j=1}^{J} w_{ij} P_{ij} w'_i - (\sum_{j=1}^{J} w_{ij} P_{ij})(\sum_{j=1}^{J} w_{ij} P_{ij})' = \sum_{j=1}^{J} (w_{ij} - \sum_{r=1}^{J} w_{ir} P_{ir}) P_{ij} (w_{ij} - \sum_{r=1}^{J} w_{ir} P_{ir})', \quad (40)
\]

which can be shown easily by adding and subtracting \((\sum_{r=1}^{J} w_{ir} P_{ir})(\sum_{r=1}^{J} w_{ir} P_{ir})'\) to the left hand side of (40) and manipulating the summation indices. And similarly,
\[
\sum_{c \in C_m} w_{ic} P_{ic|C_m} w'_ic - (\sum_{c \in C_m} w_{ic} P_{ic|C_m})(\sum_{c \in C_m} w_{ic} P_{ic|C_m})' = \sum_{c \in C_m} (w_{ic} - \sum_{s \in C_m} w_{is} P_{is|C_m}) P_{ic|C_m} (w_{ic} - \sum_{s \in C_m} w_{is} P_{is|C_m})', \quad (41)
\]

The Hessian matrix is
\[
H_B(\theta) = \frac{\partial S_B(\theta)}{\partial \theta'}, \quad (42)
\]
\[
= \sum_{i=1}^{N} \left( \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} \frac{\partial P_{ic|C_m}}{\partial \theta'} - \sum_{j=1}^{J} w_{ij} \frac{\partial P_{ij}}{\partial \theta'} \right), \quad (43)
\]
\[
= \left( \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} \frac{\partial P_{ic|C_m}}{\partial \theta'} \right) - \left( \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij} \frac{\partial P_{ij}}{\partial \theta'} \right), \quad (44)
\]
\[
= L - F, \quad (45)
\]
because

\[ \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} \frac{\partial P_{ic|C_m}}{\partial \theta} \]  \hspace{1cm} (46)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} P_{ic|C_m} (w_{ic}' - \sum_{c \in C_m} w_{ic} P_{ic|C_m}), \]  \hspace{1cm} (47)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} w_{ic} P_{ic|C_m} w_{ic}' - w_{ic} P_{ic|C_m} \sum_{c \in C_m} w_{ic} P_{ic|C_m}, \]  \hspace{1cm} (48)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \left( \sum_{c \in C_m} w_{ic} P_{ic|C_m} w_{ic}' - \left( \sum_{c \in C_m} w_{ic} P_{ic|C_m} \right) \left( \sum_{c \in C_m} w_{ic} P_{ic|C_m} \right)' \right), \]  \hspace{1cm} (49)

\[ = \sum_{i=1}^{N} \sum_{m=1}^{M} Y_{im} \sum_{c \in C_m} \left( w_{ic} - \sum_{s \in C_m} w_{is} P_{is|C_m} \right) P_{ic|C_m} \left( w_{ic} - \sum_{s \in C_m} w_{is} P_{is|C_m} \right)' \]  \hspace{1cm} (50)

\[ = L. \]  \hspace{1cm} (51)

where moving from (46) to (47) uses the expression from (39), and moving from (49) to (50) uses (41). Also, based on a similar argument, \( \sum_{i=1}^{N} \sum_{j=1}^{J} w_{ij} \frac{\partial P_{ij}}{\partial \theta} = F. \)

The information matrix is easy to derive using the preceding quantities. Note that the only random terms in \( H_B(\theta) \) are \( Y_{im} \) and that \( \mathbb{E}(Y_{im}) = \bar{P}_{im} \) for all decision makers and groups. Plugging the preceding expectations into \(-\mathbb{E}(H_B(\theta))\) results in the desired quantity in (15).

**B Forms for \( h(\delta^{(k)}) \)**

The different forms for \( h(\delta^{(k)}) \) in the iterative system \( \delta^{(k+1)} = \delta^{(k)} + h(\delta^{(k)}) f(\delta^{(k)}) \) are

(I) \( h_{BLP}(\delta^{(k)}) = I_{(J-1 \times J-1)}. \) This step size results in the contraction mapping algorithm from Berry et al. (1995).

(II) \( h_{ANJ}(\delta^{(k)}) = -[f'(\delta^{(k)})]^{-1}, \) where \( f' \) is the Jacobian matrix. This is the standard analytic Newton Jacobian (ANJ) step size and is generally not a diagonal matrix.

(III) \( h_{APNJ}(\delta^{(k)}) = -[a(s)]^{-1}, \) where \( a \) is an approximation to the Jacobian matrix that depends on the known market shares. A specific form for \( a \) is given at the end. This is the approximated Newton Jacobian (APNJ) step size and is also not a diagonal matrix.

(IV) \( h_{ADJ}(\delta^{(k)}) = -\text{Diag}(f'(\delta^{(k)}))^{-1}. \) This step size is referred to as the analytic diagonal Jacobian (ADJ). This specification results in a diagonal matrix with entries equal to the
negative reciprocals of the diagonal elements from the Jacobian. As an illustration, when there are two fixed points, \( \delta = (\delta_2, \delta_3) \), then

\[
f'(\delta) = \begin{pmatrix}
\frac{\partial f_1}{\partial \delta_2} & \frac{\partial f_1}{\partial \delta_3} \\
\frac{\partial f_2}{\partial \delta_2} & \frac{\partial f_2}{\partial \delta_3}
\end{pmatrix}, \quad \text{and} \quad h_{ADJ}(\delta) = \begin{pmatrix}
-(\frac{\partial f_1}{\partial \delta_2})^{-1} & 0 \\
0 & -(\frac{\partial f_2}{\partial \delta_3})^{-1}
\end{pmatrix}.
\]

(V) \( h_{APDJ}(\delta^{(k)}) = -Diag(a(s))^{-1} \). This specification is based on ADJ but uses the approximation to the Jacobian matrix (referred to as approximated diagonal Jacobian (APDJ)). Similar to ADJ, the matrix resulting from APDJ is diagonal with entries that equal the negative reciprocals of the diagonal elements of \( a(s) \).

The Jacobian matrix for \( f \) can be shown to equal

\[
f'(\delta) = \begin{pmatrix}
1 - \frac{\sum_{i=1}^{N} P_{i2}^2}{\sum_{i=1}^{N} P_{i2}} & -\frac{\sum_{i=1}^{N} P_{i2}P_{i3}}{\sum_{i=1}^{N} P_{i3}} & \cdots & -\frac{\sum_{i=1}^{N} P_{i2}P_{iJ}}{\sum_{i=1}^{N} P_{iJ}} \\
-\frac{\sum_{i=1}^{N} P_{i3}P_{i2}}{\sum_{i=1}^{N} P_{i3}} & 1 - \frac{\sum_{i=1}^{N} P_{i2}P_{i3}}{\sum_{i=1}^{N} P_{i3}} & \cdots & -\frac{\sum_{i=1}^{N} P_{i3}P_{iJ}}{\sum_{i=1}^{N} P_{iJ}} \\
\vdots & \vdots & \ddots & \vdots \\
-\frac{\sum_{i=1}^{N} P_{iJ}P_{i2}}{\sum_{i=1}^{N} P_{iJ}} & -\frac{\sum_{i=1}^{N} P_{iJ}P_{i3}}{\sum_{i=1}^{N} P_{iJ}} & \cdots & 1 - \frac{\sum_{i=1}^{N} P_{iJ}^2}{\sum_{i=1}^{N} P_{iJ}}
\end{pmatrix}.
\]

(52)

The step sizes \( h_{ADJ}(\delta^{(k)}) \) and \( h_{ANJ}(\delta^{(k)}) \) are constructed using (52).

The approximated step sizes \( h_{APDJ}(\delta^{(k)}) \) and \( h_{APNJ}(\delta^{(k)}) \) are based on known market shares around the fixed point and the assumption of homogenous decision makers. Specifically, from the market share constraints, we know that \( s_j = N^{-1} \sum_{i=1}^{N} P_{ij} \) must hold for \( j = 2, 3, \ldots, J \) around the fixed point, therefore all the denominators in (52) of the form \( \sum_{i=1}^{N} P_{ij} \) are approximated with \( s_j \times N \). Next, by assuming that each decision maker behaves like the market shares, we can approximate \( \sum_{i=1}^{N} P_{ij} P_{ik} \) from (52) with \( s_j \times s_k \times N \). With these assumptions, \( f'(\delta) \) is approximated as

\[
a(s) = \begin{pmatrix}
1 - s_2 & -s_3 & \cdots & -s_J \\
-s_2 & 1 - s_3 & \cdots & -s_J \\
\vdots & \vdots & \ddots & \vdots \\
-s_2 & -s_3 & \cdots & 1 - s_J
\end{pmatrix}.
\]

(53)

This approximated matrix is used to construct \( h_{APDJ}(\delta^{(k)}) \) and \( h_{APNJ}(\delta^{(k)}) \). The approximations are better when the iterative system is close to the fixed point and when the decision makers behave like the market shares. Even if these assumptions do not hold exactly, the resulting iterative systems should still be faster than the BLP system. Another feature of this approximation is that it does
not depend on \( \delta \). Unlike the analytic versions, it is not necessary to recalculate these step sizes at each \( \delta \) or in each iteration of the system. This saves a lot of time because multiple evaluations of the partial derivatives and inversions matrices are avoided. The biggest improvements occur when \( J \) is large.

C Tables
Table 1: Different forms for \( h(\delta^{(k)}) \) from Li (2012). In terms of convergence speed for the iterative system in (23), the ranking from fastest to slowest is as follows: ANJ, APNJ, ADJ, APDJ, and BLP. The specific expressions for \( h(\delta^{(k)}) \) in the context of (1) to (3) are given in the Appendix.

<table>
<thead>
<tr>
<th>Method Name</th>
<th>Form for ( h(\delta^{(k)}) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>BLP</td>
<td>( I_{(J \times J)} )</td>
</tr>
<tr>
<td>ANJ</td>
<td>(-[f'(\delta^{(k)})]^{-1})</td>
</tr>
<tr>
<td>APNJ</td>
<td>(-[a(s)]^{-1})</td>
</tr>
<tr>
<td>ADJ</td>
<td>(-Diag(f'(\delta^{(k)}))^{-1})</td>
</tr>
<tr>
<td>APDJ</td>
<td>(-Diag(a(s))^{-1})</td>
</tr>
</tbody>
</table>

Table 2: Repeated sample maximum likelihood estimates of \( \theta \) using exact choice data (ECD), broad choice data (BCD), and broad choice data with constraints (BCDC). “Est” and “SE” are sample averages of the estimators and their estimated standard errors over the 250 Monte Carlo samples.

<table>
<thead>
<tr>
<th></th>
<th>ECD</th>
<th>BCD</th>
<th>BCDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>Est. SE</td>
<td>Est. SE</td>
<td>Est. SE</td>
</tr>
<tr>
<td>( \delta_2 )</td>
<td>1.68 0.08</td>
<td>1.72 0.32</td>
<td>1.70 0.06</td>
</tr>
<tr>
<td>( \delta_3 )</td>
<td>0.44 0.08</td>
<td>0.44 0.35</td>
<td>0.49 0.04</td>
</tr>
<tr>
<td>( \delta_4 )</td>
<td>4.96 0.10</td>
<td>4.97 0.35</td>
<td>4.93 0.18</td>
</tr>
<tr>
<td>( \delta_5 )</td>
<td>2.51 0.08</td>
<td>2.53 0.31</td>
<td>2.54 0.07</td>
</tr>
<tr>
<td>( \delta_6 )</td>
<td>-0.78 0.08</td>
<td>-0.76 0.41</td>
<td>-0.75 0.04</td>
</tr>
<tr>
<td>( \delta_7 )</td>
<td>-0.22 0.09</td>
<td>-0.23 0.39</td>
<td>-0.22 0.04</td>
</tr>
<tr>
<td>( \delta_8 )</td>
<td>0.74 0.08</td>
<td>0.73 0.35</td>
<td>0.72 0.04</td>
</tr>
<tr>
<td>( \delta_9 )</td>
<td>1.49 0.08</td>
<td>1.48 0.35</td>
<td>1.50 0.05</td>
</tr>
<tr>
<td>( \delta_{10} )</td>
<td>0.61 0.08</td>
<td>0.62 0.25</td>
<td>0.64 0.04</td>
</tr>
<tr>
<td>( \beta_1 )</td>
<td>3.18 0.04</td>
<td>3.19 0.10</td>
<td>3.17 0.10</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>2.73 0.04</td>
<td>2.74 0.09</td>
<td>2.72 0.09</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>6.00 0.12</td>
<td>6.05 0.43</td>
<td>6.02 0.43</td>
</tr>
</tbody>
</table>
Table 3: Coverage probabilities for the frequentist 80% confidence intervals based on exact choice data (ECD), broad choice data (BCD), and broad choice data with constraints (BCDC).

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>ECD</th>
<th>BCD</th>
<th>BCDC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2$</td>
<td>0.77</td>
<td>0.84</td>
<td>0.67</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.81</td>
<td>0.85</td>
<td>0.14</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>0.79</td>
<td>0.83</td>
<td>0.77</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>0.79</td>
<td>0.83</td>
<td>0.74</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>0.83</td>
<td>0.82</td>
<td>0.36</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>0.74</td>
<td>0.80</td>
<td>0.18</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>0.74</td>
<td>0.84</td>
<td>0.44</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>0.78</td>
<td>0.83</td>
<td>0.71</td>
</tr>
<tr>
<td>$\delta_{10}$</td>
<td>0.80</td>
<td>0.85</td>
<td>0.31</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.80</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.80</td>
<td>0.77</td>
<td>0.79</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.77</td>
<td>0.81</td>
<td>0.82</td>
</tr>
</tbody>
</table>

Table 4: Bayesian estimates for $\theta$ based on exact choice data (ECD), broad choice data (BCD), and broad choice data with an informative prior (BCDIP). The informative prior uses $\tau = 0.005$. “Mean” and “SD” are sample averages over the 250 Monte Carlo samples of the posterior means and standard deviations for columns 2-8 and of the prior means and standard deviations for columns 9-10.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>ECD</th>
<th>BCD</th>
<th>BCDIP</th>
<th>BCDIP Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2$</td>
<td>1.68</td>
<td>1.69</td>
<td>0.08</td>
<td>1.74</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.44</td>
<td>0.45</td>
<td>0.08</td>
<td>0.44</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>4.96</td>
<td>4.97</td>
<td>0.09</td>
<td>4.99</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>2.51</td>
<td>2.52</td>
<td>0.08</td>
<td>2.56</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>-0.78</td>
<td>-0.77</td>
<td>0.09</td>
<td>-0.78</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>-0.22</td>
<td>-0.21</td>
<td>0.08</td>
<td>-0.23</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>0.74</td>
<td>0.75</td>
<td>0.08</td>
<td>0.74</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>1.49</td>
<td>1.50</td>
<td>0.08</td>
<td>1.50</td>
</tr>
<tr>
<td>$\delta_{10}$</td>
<td>0.61</td>
<td>0.62</td>
<td>0.08</td>
<td>0.64</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>3.18</td>
<td>3.18</td>
<td>0.04</td>
<td>3.20</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>2.73</td>
<td>2.73</td>
<td>0.03</td>
<td>2.75</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>6.00</td>
<td>6.00</td>
<td>0.12</td>
<td>6.10</td>
</tr>
</tbody>
</table>
Table 5: Coverage probabilities for the Bayesian 80% probability intervals based on exact choice data (ECD), broad choice data (BCD), and broad choice data with constraints (BCDC). The informative prior uses $\tau = 0.005$.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>ECD</th>
<th>BCD</th>
<th>BCDIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2$</td>
<td>0.76</td>
<td>0.84</td>
<td>0.52</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.81</td>
<td>0.83</td>
<td>0.26</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>0.78</td>
<td>0.82</td>
<td>0.70</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>0.79</td>
<td>0.83</td>
<td>0.55</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>0.82</td>
<td>0.83</td>
<td>0.47</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>0.75</td>
<td>0.81</td>
<td>0.58</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>0.74</td>
<td>0.85</td>
<td>0.63</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>0.78</td>
<td>0.81</td>
<td>0.68</td>
</tr>
<tr>
<td>$\delta_{10}$</td>
<td>0.81</td>
<td>0.84</td>
<td>0.47</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.80</td>
<td>0.81</td>
<td>0.78</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.78</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.78</td>
<td>0.82</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Table 6: Bayesian estimates for $\theta$ based on exact choice data (ECD), broad choice data (BCD), and broad choice data with an informative prior (BCDIP). The informative prior uses $\tau = 0.1$. “Mean” and “SD” are sample averages over the 250 Monte Carlo samples of the posterior means and standard deviations for columns 2-8 and of the prior means and standard deviations for columns 9-10.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>ECD</th>
<th>BCD</th>
<th>BCDIP</th>
<th>BCDIP Prior</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2$</td>
<td>1.68</td>
<td>1.69</td>
<td>0.08</td>
<td>1.74</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.44</td>
<td>0.45</td>
<td>0.08</td>
<td>0.44</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>4.96</td>
<td>4.97</td>
<td>0.09</td>
<td>4.99</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>2.51</td>
<td>2.52</td>
<td>0.08</td>
<td>2.56</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>-0.78</td>
<td>-0.77</td>
<td>0.09</td>
<td>-0.78</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>-0.22</td>
<td>-0.21</td>
<td>0.08</td>
<td>-0.23</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>0.74</td>
<td>0.75</td>
<td>0.08</td>
<td>0.74</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>1.49</td>
<td>1.50</td>
<td>0.08</td>
<td>1.50</td>
</tr>
<tr>
<td>$\delta_{10}$</td>
<td>0.61</td>
<td>0.62</td>
<td>0.08</td>
<td>0.64</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>3.18</td>
<td>3.18</td>
<td>0.04</td>
<td>3.20</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>2.73</td>
<td>2.73</td>
<td>0.03</td>
<td>2.75</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>6.00</td>
<td>6.00</td>
<td>0.12</td>
<td>6.10</td>
</tr>
</tbody>
</table>
Table 7: Coverage probabilities for the Bayesian 80% probability intervals based on exact choice data (ECD), broad choice data (BCD), and broad choice data with constraints (BCDC). The informative prior uses $\tau = 0.1$.

<table>
<thead>
<tr>
<th>$\theta$</th>
<th>ECD</th>
<th>BCD</th>
<th>BCDIP</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta_2$</td>
<td>0.76</td>
<td>0.84</td>
<td>0.90</td>
</tr>
<tr>
<td>$\delta_3$</td>
<td>0.81</td>
<td>0.83</td>
<td>0.87</td>
</tr>
<tr>
<td>$\delta_4$</td>
<td>0.78</td>
<td>0.82</td>
<td>0.82</td>
</tr>
<tr>
<td>$\delta_5$</td>
<td>0.79</td>
<td>0.83</td>
<td>0.86</td>
</tr>
<tr>
<td>$\delta_6$</td>
<td>0.82</td>
<td>0.83</td>
<td>0.88</td>
</tr>
<tr>
<td>$\delta_7$</td>
<td>0.75</td>
<td>0.81</td>
<td>0.90</td>
</tr>
<tr>
<td>$\delta_8$</td>
<td>0.74</td>
<td>0.85</td>
<td>0.91</td>
</tr>
<tr>
<td>$\delta_9$</td>
<td>0.78</td>
<td>0.81</td>
<td>0.86</td>
</tr>
<tr>
<td>$\delta_{10}$</td>
<td>0.81</td>
<td>0.84</td>
<td>0.84</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>0.80</td>
<td>0.81</td>
<td>0.79</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>0.78</td>
<td>0.77</td>
<td>0.76</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>0.78</td>
<td>0.82</td>
<td>0.78</td>
</tr>
</tbody>
</table>

References


URL http://www.jstor.org/stable/2171802


URL http://www.socsci.uci.edu/~phillil/research/cm.pdf


