Foundations of Stated Preference Elicitation: Consumer Behavior and Choice-based Conjoint Analysis

Moshe Ben-Akiva[1], Daniel McFadden[2,3] and Kenneth Train[2]

[1] *Department of Civil and Environmental Engineering, MIT Cambridge, MA 02139, USA*
[2] *Department of Economics, University of California, Berkeley, CA 94720-3880, USA*
[3] *Department of Economics, University of Southern California, CA 90089-921, USA*

ABSTRACT

Stated preference elicitation methods collect data on consumers by "just asking" about tastes, perceptions, valuations, attitudes, motivations, life satisfactions, and/or intended choices. Choice-Based Conjoint (CBC) analysis asks subjects to make choices from hypothetical menus in experiments that are designed to mimic market experiences. Stated preference methods are controversial in economics, particularly for valuation of non-market goods, but CBC analysis is accepted and used widely in marketing and policy analysis. The promise of stated preference experiments is that they can provide deeper and broader data on the structure of consumer preferences than is obtainable from revealed market observations, with experimental control of the choice environment that circumvents the feedback found in real market equilibria. The risk is that they give pictures of consumers that do not predict real market behavior. It

is important for both economists and non-economists to learn about the performance of stated preference elicitations and the conditions under which they can contribute to understanding consumer behavior and forecasting market demand. This monograph re-examines the discrete choice methods and stated preference elicitation procedures that are commonly used in CBC, and provides a guide to techniques for CBC data collection, model specification, estimation, and policy analysis. The aim is to clarify the domain of applicability and delineate the circumstances under which stated preference elicitations can provide reliable information on preferences.

Preface

Information on consumer preferences and choice behavior is needed to forecast market demand for new or modified products, estimate the effects of product changes on market equilibrium and consumer welfare, develop and test models of consumer behavior, and reveal determinants and correlates of tastes. Direct elicitation of stated preferences, perceptions, expectations, attitudes, motivations, choice intentions, and well-being, supplementing or substituting for information on revealed choices in markets is potentially a valuable source of data on consumer behavior, but can mislead if the information environments and decision-making processes invoked by direct elicitations differ from the settings for revealed choices in real markets.

The purpose of this monograph is to provide the reader with stated preference data collection methods, discrete choice models, and statistical analysis tools that can be used to forecast demand and assess welfare impacts for new or modified products or services in real markets, and summarize the conditions under which the reliability of these methods has been demonstrated or can be tested. One focus is the collection of preference and related data from consumer responses in hypothetical choice experiments, particularly choice-based conjoint analysis (CBC) methods that have proven useful in market research. Another is the economic theory and statistical analysis of choice behavior, revealed or stated, and an economic framework for forecasting market demand

and measuring consumer welfare. Stated choice data can be collected and combined with a broad spectrum of models of consumer behavior. This monograph will focus on the standard economic model of utility-maximizing consumers. Our treatment is informed by and benefits from experiments on perceptions and decision-making behavior in cognitive science and behavioral economics, and includes methods that can accommodate features of consumer choice that impact forecast reliability such as anchoring, adaptation to the status quo, and sensitivity to context. However, we will only touch on the implications of behavioral consumer theory for elicitation and analysis of stated preference data.

There are a number of good introductions to discrete choice analysis (Ben-Akiva and Lerman, 1985; Train, 1986; Train, 2009; McFadden, 1999; McFadden, 2001; McFadden, 2014b; Brownstone, 2001; Boyce and Williams, 2015), and to stated preference and conjoint analysis methods and market research applications (Louviere, 1988; Fischhoff and Manski, 1999; Louviere *et al.*, 2000; Wittink and Bergestuen, 2001; Hauser and Rao, 2002; Rossi *et al.*, 2005; Chandukala *et al.*, 2007; Raghavarao *et al.*, 2010; Rao, 1977; Rao, 2014; Green and Srinivasan, 1978; Green and Srinivasan, 1990). This monograph complements these introductions by filling in technical and behavioral backgrounds for these topics.

1

Some History of Stated Preference Elicitation

Stated preference methods date from the 1930s. The originator of empirical demand analysis, Henry Schultz, persuaded his University of Chicago colleague, the iconic psychologist Leon Thurstone (1931), to present a paper at the second meeting of the Econometric Society with this proposal for direct elicitation of indifference curves:

> "Perhaps the simplest experimental method that comes to mind is to ask a subject to fill in the blank space [to achieve indifference] in a series of choices of the following type:
>
> 'eight hats and eight pairs of shoes' versus 'six hats and____pairs of shoes'
>
> One of the combinations such as eight hats and eight pairs of shoes is chosen as a standard and each of the other combinations is compared directly with it."

Thurstone postulated that responses would obey Fechner's law, a common psychophysical regularity in the sensation produced by a stimulus. This turns out to be equivalent to postulating that respondents maximize a log-linear utility function. He collected experimental data on hats

5

vs. shoes, hats vs. overcoats, and shoes vs. overcoats, fit the parameters of the log-linear utility function to data from each comparison, treating responses as bounds on the underlying indifference curves, and tested the consistency of his fits across the three comparisons.

At the time of Thurstone's presentation, empirical demand analysis was in its early days. Pioneering studies of market demand for a single commodity (sugar) had been published by Frisch (1926), Schultz (1925), and Schultz (1928), but there were no empirical studies of multi-product demands. Least-squares estimation was new to economics, and required tedious hand calculation. Consolidation of the neoclassical theory of demand for multiple commodities by Hicks (1939) and Samuelson (1947) was still in the future. Given this setting, Thurstone's approach was path-breaking. Nevertheless, his estimates were rudimentary, and he failed to make a connection between his fitted indifference curves and market demand forecasts.[1] Most critically, he did not examine whether the cognitive tasks of stating indifferent quantities in his experiment and of choosing best bundles subject to a budget constraint in real markets were sufficiently congruent so that responses with respect to the first would be predictive for the second.

According to Moscati (2007), Thurstone's presentation was criticized from the floor by Harold Hotelling and Ragnar Frisch. First, they objected that Thurstone's indifference curves as constructed were insufficient to forecast market demand response to price changes; this objection failed to recognize that extending Thurstone's comparisons to include residual expenditure could have solved the problem. Second, they pointed out that the knife-edge of indifference is not well determined

---

[1]In retrospect, these two flaws were correctable: Denote by H, S, C, respectively, the numbers of hats, pairs of shoes, and coats consumed, let M denote the money remaining for all other goods and services after paying for the haberdashery, and let Y denote total income. Suppose Thurstone had asked subjects for the amounts M that made a comparison bundle (H, S, C, M) indifferent to a standard bundle (H_0, S_0, C_0, M_0). Then, he could have estimated the parameters of the log-linear utility function $u = \log M + \theta_H \log H + \theta_S \log S + \theta_C \log C$ by least squares regression of $\log(M/M_0)$ on $\log(H_0/H)$, $\log(S_0/S)$, and $\log(C_0/C)$. From this, he could have forecast demands, e.g., hat demand at price p_H and income Y would have been given by the formula $H = \theta_H Y/p_H(1 + \theta_H + \theta_S + \theta_C)$ that comes from utility maximization subject to the budget constraint.

in comparisons of bundles of discrete commodities.[2] Beyond these objections, Frisch and Hotelling were generally skeptical that stated indifference points, or non-market responses more generally, could be used to predict market behavior. The orthodoxy of that era, formed partly as a reaction to the casual introspections of Bentham and the utilitarians, was that empirical economics should rely solely on revealed market data; in the words of Irving Fisher (1892), *"To fix the idea of utility, the economist should go no further than is serviceable in explaining economic facts. It is not his province to build a theory of psychology."* Wallis and Friedman (1942) summarized this attitude in an attack that forcefully dismissed Thurstone's method or any other attempt to use experimental data for market demand analysis, pointing out difficulties in designing experiments that mimic the environment of real market choices: *"[Thurstone's] fundamental shortcomings probably cannot be overcome in any experiment involving economic stimuli and human beings."*

For 40 years following Thurstone's paper, consumer experiments were mostly limited to testing axioms for choice under uncertainty, and there was no systematic attempt to incorporate stated preference (SP) methods in demand analysis. There was some reason for this lack of interest. The language of economic analysis, then and now, is prediction of market demand and assessment of market failures in terms of dollars of equivalent lost income. Any measurement method that uses experimental data on preferences has to produce convincing results in this language by showing that stated preferences collected outside the market have the same predictive power for market behavior as implied preferences reconstructed from market data. With the advent of behavioral economics, we have learned that people are often not relentless utility maximizers, either in markets or in experiments, undermining the tight links neoclassical consumer theory postulates between consumer utility and demand behavior. This has led to calls for less focus on

---

[2] Additional objections could have been raised about the applicability of Fechner's law and the restrictiveness and realism of the log linear utility specification, lack of accounting for heterogeneity across consumers, and lack of explicit treatment of consumer response errors. Decades later, when this demand system was fitted to revealed preference (RP) data, these issues did arise.

assessment of consumer welfare in dollars of lost income deduced from market demand behavior, and more research on benchmarking stated satisfaction to psychological or neurological measures of well-being. This approach may eventually succeed, but at present market prediction and valuation remain the yardsticks against which any method for eliciting consumer preferences and inferring consumer welfare has to be judged.

The first sustained use of stated preference methods came out of the theory of conjoint measurement pioneered by Luce and Tukey (1964) and Luce and Suppes (1965), and developed as *conjoint analysis* by market researchers like Green (1974), Johnson (1974, 1999), Huber (1975, 1987), Srinivasan (1988) and Louviere (1988) and applied to the study of consumer preferences among familiar market products (e.g., carbonated beverages, automobiles). Good introductions to conjoint experiments, data, and analysis methods are Louviere *et al.* (2000) and Rossi *et al.* (2005). A central feature of conjoint analysis is the use of experimental designs that allow at least a limited mapping of the preferences of each subject, and multiple measurements that allow estimates of preferences to be tested for consistency and refined when necessary.

Early conjoint analysis experiments described hypothetical products in terms of price and levels of attributes in various dimensions, and asked subjects to rank attributes in importance, and rate attribute levels. These measurements were used by market researchers to classify and segment buyers, and target advertising, but they were not reliable tools for predicting market demand. However, Louviere and Woodworth (1983) and Hensher and Louviere (1983) introduced choice-based con- joint (CBC) elicitations that directly mimicked market choice tasks, and McFadden *et al.* (1986) and McFadden (1986) showed how these elicitations could be analyzed using the tools of discrete choice analysis and the theory of random utility maximization (RUM).[3] Subjects would

---

[3]The term "CBC" is used in marketing to include stated choice elicitations without an underlying framework of utility-maximizing discrete choice. More explicit terminology for the approach discussed in this monograph would be "CBC/RUM", or as suggested by Carson and Louviere (2011), "discrete choice experiments" (DCE). However, we will continue to use the umbrella term "CBC", leaving it to the reader to distinguish our economic approach from other forms of conjoint analysis.

be presented with a series of menus of products. Each product offered in each menu would be described in terms of price and levels of attributes. Subjects would be asked to choose their most preferred product in each menu. For example, subjects would be offered a menu of paper towels, with each product described in terms of price, number of towel sheets, a measure of the absorption capacity, a measure of strength when wet, and brand name. Choice data from these menus, within and across subjects, could then be handled in the same way as the real market choice data. Choice-based conjoint (CBC) surveys analyzed using discrete choice methods have become widely used and accepted in market research to predict the demand for consumer products, with a sufficient track record so that it is possible to identify some of the necessary conditions for successful prediction; see McFadden *et al.* (1988) and Green *et al.* (2001), Cameron and DeShazo (2013), and McFadden (2014, 2017).

Environmental economists developed independently a simple stated preference method termed *contingent valuation* (CV), and applied it to valuing environmental damage. This method traces its beginnings to a proposal by Ciriacy-Wantrup (1947) and a PhD thesis by Davis (1963a, 1963b) on the use-value of Maine woods. Its first published applications for values of environmental public goods seem to have been Brookshire *et al.* (1976), Bishop and Heberlein (1979), and Brookshire *et al.* (1980). CV can be viewed as a truncated form of conjoint analysis with two important differences. First, it does not have the experimental design features of conjoint analysis that allow extensive tests for the structure and consistency of stated preferences. Second, because of its applications, it usually does not have predictive accuracy in markets as a direct yardstick for reliability. Instead, it relies indirectly on the links between preferences, market demands, and valuations that hold when neoclassical consumer theory is valid, on analogies with stated preference studies of consumer goods in markets, and on limited internal consistency checks. The particular challenges of using contingent valuation for natural resource valuation are discussed by Carson *et al.* (2001), Carson (2012), Hausman (2012), and Kling *et al.* (2012). Its reliability in relation to stated preference elicitation for market goods is discussed by McFadden (2017).

Other elicitation methods for stated preferences, termed *focus groups*, *vignette analysis*, and measurement of *subjective well-being*, are popular among some applied economists and political scientists, see Rossi (1979), King *et al.* (2004), Caro *et al.* (2012), and Kahneman and Krueger (2013). Focus group methods are directed open-end discussions of products and their features in small samples of consumers. Focus groups do not provide direct data for market demand forecasts, but they can be quite useful in designing CBC experiments because of the insights they provide on product perceptions, levels of consumer knowledge, experience, and understanding, and the attributes consumers consider in forming their preferences. Vignette analysis uses detailed story descriptions of alternatives, often visual. Vignette presentations of alternatives can be used within conjoint analysis experiments, and may improve subject attention and understanding of alternatives. Subjective well-being methods elicit overall self-assessments of welfare, often on Likert or rating scales similar to those used in the early days of conjoint analysis. In the instances where vignette and subjective well-being methods have been tested, they prove to be strongly influenced by context and anchoring effects, see Deaton (2012). These effects compromise forecast accuracy in market demand forecasting applications. Psychometrics has developed an array of additional methods for measuring perceptions, attitudes, and motivations. Their usefulness for economic demand forecasting has not been demonstrated, but at least for perceptions and intentions it is clear that further development is potentially quite valuable for economic applications.

The focus of this monograph is market demand forecasting for new or modified products; we do not attempt here any overall assessment of the reliability of contingent valuation, vignette analysis, or subjective well-being methods in their primary uses. We urge readers to not casually pre-judge the use of stated preference methods in economic applications, but rather to acquire the tools needed to conduct and critique CBC studies, investigate how well these methods work under various conditions, and make a reasoned scientific judgement on when their use advances our understanding of consumer behavior and well-being.

2

Choice-Based Conjoint (CBC) Analysis

A choice-based conjoint (CBC) study offers a consumer a series of menus of alternative products with profiles giving levels of their attributes, including price, and asks the subject to choose the most preferred product in each menu.[1] The menus of products and their descriptions are designed to realistically mimic a market experience where a consumer can choose among various competing products. By changing the attribute levels for the included products and presenting each consumer with several menus, a CBC experiment obtains information on the relative importance that consumers place on each of the attributes. A classic CBC setup in marketing might be a laboratory experiment where subjects are offered a sequence of binary menus of actual products with varying attributes and prices, and asked to indicate the preferred

---

[1]This setup assumes that menu alternatives are mutually exclusive. Applications where the consumer buys more than one unit of the same or multiple products can be handled in this framework by listing each possible combination of purchases as a separate menu item. Alternately, CBC subjects could be allowed to check the number of units of each product they would purchase, with zero indicating "no purchase", as they do when placing product orders on Amazon. This alternative requires that the discrete choice models used to analyze stated preference data be modified to encompass utility-maximizing counts, see Burda *et al.* (2012), Bhat *et al.* (2014), and Chen (2017).

alternative in each menu. For example, subjects might be trained by tasting cola drink "brands" with various degrees of sweetness, carbonation, and flavor, and then asked in each menu to choose one brand–price combination or the other. They may be promised that at the end of the experiment, they will receive a six-pack of their chosen brand from one of the menus they faced, plus a participation payment less the menu price of the six-pack. Often, CBC is used in marketing studies of familiar products whose features can be described in words and pictures, with subjects asked to choose from a menu of products based on these descriptions. These studies often employ telephone, internet, or mail survey instruments, and recruit subjects using sampling frames that range from stratified random samples of the target population to samples obtained from internet volunteers, intercepts at retail outlets, or registered product buyers. For example, experiments on automobile brand and model choice have described alternatives in terms of price and attributes such as horsepower, fuel consumption, number of seats, and cargo space, and collected data using interactive internet sessions with randomly sampled consumers, see Urban *et al.* (1990), Urban *et al.* (1997), and Brownstone and Train (1999). These studies have determined with considerable predictive accuracy in the distributions of preference weights that consumers give to various vehicle features.

2.1 Issues in CBC study design

Based on the authors' experience in conducting stated preference experiments, and our review of the literature, we conclude that there are a number of key issues that need to be addressed in any CBC study, see also Carson *et al.* (1994) and McFadden (2018). We summarize these in the checklist given in Table 2.1, and discuss them further in the paragraphs that follow.

2.1.1 Familiarity

The selection of subjects, the setup and framing of the elicitation, training of subjects where necessary, and testing of subjects' understanding and consistency are critical in a CBC study, particularly when some products are novel with attributes that are not easily experienced or

Table 2.1: CBC study design checklist.

	Issue	Discussion
1	Familiarity: Subjects should be familiar with the class of products studied and their attributes, after training if necessary, and experienced in making market choices among them	2.1.1
2	Sampling, Recruitment, Background: Subjects should be sampled randomly from the target population, and compensated sufficiently to assure participation and attention. Background on socioeconomic status and purchase history should be collected for sample correction and modeling of heterogeneity in choice behavior	2.1.2
3	Outside Option: Menus should include a meaningfully specified "no purchase" alternative, or the study should be designed to use equivalently identifying external background or real market data	2.1.3
4	Menu Design: The number of menus, products per menu, attributes per product, and span and realism of attribute levels must balance market realism and the independent variation needed for statistical accuracy	2.1.4
5	Attribute Formatting: The clarity and prominence of attributes should mimic, to the extent possible given the goals of the analysis, the information environment the consumer will face in the real market	2.1.5
6	Elicitation Frame: Elicitation formats other than hypothetical market choice must balance information value against the risk of invoking incompatible cognitive frames	2.1.6
7	Incentive Alignment: Elicitations with a positive incentive for a truthful response reduce the risk of carelessness or casual opinion	2.1.7
8	Subject Training: Training may be necessary to provide sufficient familiarity and experience with products, but for reliable forecasting it needs to mimic the "street" training provided by real markets and minimize manipulation that is not realistic	2.1.8
9	Calibration and Testing: Consistency and reality checks should be built into the study, and forecasting accuracy should, if possible, be tested against external data	2.1.9

understood, or subjects have not had market experience with similar products. CBC studies can forecast market demand relatively well when the subjects' task is choice among a small number of realistic, familiar,

and fully described products, and subjects have had experience buying products of this type and can realistically expect to make further purchases in the future. However, there is a sharp reliability gradient, and stated preferences become problematic when the products are unfamiliar or incompletely described, or involve public good or social aspects that induce respondents to consider social welfare judgments or peer opinions. For example, when stated preferences for automobile models are collected in an elicitation that emphasizes the energy footprint and environmental consequences of the alternatives, subjects may give these attributes more weight than they would in a real market, or state more socially responsible preferences than they would reveal in actual market choices.[2]

Familiarity may be a given when the CBC study is of ordinary household products such as cola drinks, but not when novel products or novel or highly technical product features are outside the subject's range of experience. Where possible, subjects should have an opportunity to "test-drive" novel products. For example, in a study of consumer choice among streaming music services, it should improve prediction to give subjects hands-on experience with different features, in a setting that mimics their opportunities to investigate and experience these features in real shopping. In some cases, this can be accomplished with actual or mock-up working models of the offered products, or with computer simulation of their operation. There is a tradeoff: Attempting to train consumers and providing mock-ups can inadvertently create anchoring effects. Consumers who are unfamiliar with a product may take the wording in the training exercises about the attributes, or the characteristics of the mockups, as clues to what they should feel about each attribute. Even the mention of an attribute can give it more prominence in a subject's mind than it would have in real shopping. The

---

[2]This is not an argument that energy footprint should be excluded from the attribute list in studying automobile choice, but rather a caution that when it is included, responses are likely to be quite sensitive to how statements about energy footprints and environmental consequences are framed. In general, valuation of non-use aspects of natural resources is challenging for conjoint methods because these applications seek to measure preference that lies outside the normal market experience of consumers and may invoke sentiments regarding social responsibility that are partially formed and unstable.

researcher needs to weigh the often conflicting goals of making subjects knowledgeable about products and avoiding distortion of the relative valuations of attributes these subjects would exhibit in real markets.

2.1.2 Sampling, recruitment, background

Target populations may differ depending on the objectives of the study — e.g., all current users of a class of products, only technologically savvy current users, the general population. For many policy applications, it is a major blunder to sample only current buyers of a class of products, as the preferences and behavior of non-buyers are normally also needed to determine the effect of policy changes on market equilibrium, the preferences of marginal buyers in equilibrium, and welfare consequences for both initial non-buyers and buyers. It is important that the sampling frame draws randomly from the target population, without excessive weighting to correct for stratification and non-response. It is generally a bad idea to use opportunistic pools of potential respondents, such as subjects recruited from volunteers at internet survey sites or intercepted at shopping malls or transportation nodes, as these pools may differ substantially from the target population in unobserved characteristics, even after weighting ("racking") to match the target population distribution on observed dimensions such as age and education.

It is sometimes desirable to sample from a target population that has experience or expertise with a class of products. Then it may be informative to study the preferences of experienced users, and separately study the differences in tastes of these users and the general population. An example is study of consumer demand for relatively esoteric technical attributes of products, say the levels of encryption built into telecommunication devices, where only technically savvy device users will appreciate the meaning of different encryption levels. In this case, a good study design may be to conduct an intensive conjoint analysis on technically knowledgeable users, and separately survey the target population to estimate the extent and depth of technical knowledge, and the impact of technical information on the purchase propensities of current general users and non-users.

Background information on socioeconomic status and product purchase history should be collected on subjects in a CBC study, to enable

weighting to correct for sampling imperfections such as uneven attrition rates, and to provide the foundation for estimation of choice models that are accurate for various demographic groups. Importantly, wage, asset, and transfer income are needed to model income effects in choice behavior. Relatively universal internet access has led to inexpensive and effective surveying using internet panels, rather than telephone, mail, or personal interviews. However, it is risky to recruit internet survey subjects from volunteers, as they can look quite different than a target population of possible product buyers. Better practice is to use a reliable sample frame such as random sampling of addresses, then recruit internet panel subjects from the sampled addresses. It is also important to compensate subjects for participation at sufficient levels to minimize selection due to attrition; see McFadden (2012). Experience with "professional" subjects who are carefully recruited and paid to participate in internet panels, with extensive background information collected prior to the CBC study, is positive: Subjects who view responding as a continuing "job" with rewards for effort seem to be more attentive and responsive than casual respondents.

2.1.3   Outside option

Obviously, a consumer's history and future opportunities influence product choice. If the consumer has in the past purchased substitutes or complements that are still on hand, this affects the desirability of items on a current menu. Further, if a current menu choice opens, closes, or changes the attractiveness of options in the future, the utility of a product purchase will include allowance for its future impacts. For example, faced with a menu of cars with various attributes and prices, a consumer will consider how a current choice will influence the use or disposition of vehicles already owned, and how this choice would fit not only with anticipated needs, but also with future purchases. The timing of consumer durables and perishable restocking purchases is an important aspect of consumer behavior in real markets. A reliable CBC study must give subjects a decision-making context that is realistic, taking account of their market history, and providing clear instructions on purchase timing and how current stated menu choices affect future

options. Critically, if subjects strategically "game" their experimental responses based on their perceived market opportunities in other times and places, and these perceptions do not match strategic opportunities in a future real market, then CBC forecasts for this real market will tend to be unreliable.

For a neoclassical consumer, the decision to buy at a particular decision point is made by comparing the maximum utility from a purchase now and the utility of "no purchase", where the latter is the option value of opportunities at future times and places for choices that complement or substitute for a purchase now. A CBC study that offers menus of products without a "no purchase" option can predict market shares, but not overall product demand. To go further and answer important questions such as the extent to which new products expand demand rather than simply capture market share, one can either turn to external market data to calibrate product demand, or include a "no purchase" option in the CBC menus. Both strategies present design issues.

Consider first external calibration. If real market demand data with existing products, attributes, and prices are available at the degree of granularity on demographics and purchase history attainable from CBC subjects, then it is straightforward to start from a model fitted to CBC data without a "no purchase" option, calibrate the model's "no purchase" utility as a function of history and socioeconomics so that it matches current demand under current market conditions, and then use this calibrated model to forecast future demand. A more general approach stacks CBC product choice data and baseline real market data, with the latter broken down by available demographics and history, and estimates a pooled model for forecasting. The accuracy of external calibration is limited by the granularity of real market data, as it cannot capture variations in purchase probability arising from unobserved differences in consumer histories. For example, it is unusual to have real market demand broken down by buyers' income, age, or purchase history, while these are commonly collected as background in CBC studies. Reliable external calibration obviously requires consistency in the framing of study questions on real market behavior and the definitions used in collecting real market demand data.

Consider next the strategy of including a "no purchase" option in CBC elicitations. One can simply make "no purchase" as an alternative on some or all menus. Alternately, one can ask first for the best product choice from a menu without a "no purchase" option, and follow up the response by asking if the subject would in fact buy this chosen product. An advantage of the alternative approach is that it keeps subjects from using a "no purchase" choice as a way to avoid the time and effort required to evaluate the offered products; it may also mimic the way consumers make choices in real markets for complex products such as cars or houses. However, a drawback of the alternative approach is that unless strong separability conditions on preferences are met, choice probabilities among products without a "no purchase" option will not in general coincide with product choice probabilities conditioned on a decision to purchase. Another drawback is that conditional product choice may unrealistically color subsequent stated purchase choice.

An issue with any method of including a "no purchase" option in CBC elicitations is that its meaning needs to be carefully specified for subjects. If a subject states "no purchase" now, what are they being asked to assume about their possibilities for hypothetical and real market purchases in the future? If "no purchase" in the CBC setting is interpreted differently by different subjects, and inconsistently with real market data definitions such as the numbers of units of products sold in a calendar quarter, a CBC study will fail the mechanical accounting requirement for reliable forecasting that the CBC menu and the real market forecast have consistent purchase periods. For example, in a car choice study, a subject might interpret a "no purchase" option offered without explanation in a CBC study as meaning that they would continue to use indefinitely a vehicle that the household currently owns; that they would do without a new car in the current model year; or that they could at any time go to the real market as an alternative to having a product choice from a CBC menu fulfilled. These interpretations are all inconsistent with a study goal of forecasting the number of cars sold in the next model year. Better practice in CBC design is to make the meaning of "no purchase" specific and explicit, and establish that subjects understand and accept it.

2.1.4 Menu design

The design of a conjoint experiment establishes the number of menus offered to each subject, the nature and number of products on each menu, the number of attributes and attribute levels introduced for each product, and the design of the profiles of the products placed on each menu. The levels of attributes of the products offered on different menus can be set by experimental design, so that it is possible to separate statistically the weights that consumers give to the different attributes. In its early days, menu designs were often of a "complete profile" form that mimicked classical experimental design and allowed simple computation of "part-worths" from rating responses, but currently the emphasis is simply on ensuring that menus are realistic and incorporate sufficient independent variation in the attributes, so that the impact of each attribute on choice can be isolated statistically. The classical statistical literature on experimental design focused on analysis of variance and emphasized orthogonality properties that permitted simple computation of effects, and treatments that provided minimum variance estimates. Designs that reduce some measure of the sampling variance under specified model parameters (such as the determinant of the covariance matrix for "D-efficiency") have been implemented in market research by Kuhfield *et al.* (1994), Bleimer and Rose (2009), and Rose and Bleimer (2009), and others. It is important that conjoint studies be designed to yield good statistical estimates, but there is relatively little to be gained from adherence to designs with classical completeness and orthogonality properties. With contemporary computers, the computational simplifications from orthogonal designs are usually unimportant. Further, for the nonlinear models used with CBC, orthogonality of attributes does not in general minimize sampling variance. Unlike classical analysis of variance problems, it is not usually possible in nonlinear choice models to specify efficient designs in advance of knowing the parameters that are the target of the analysis; see also Walker *et al.* (2018).

In real markets, as a result of competition, surviving products are often similar in their attributes and prices. Much of the power of a CBC study is the opportunity to identify preferences in menus

with independent variation in attributes that are highly correlated in real markets. However, relatively mechanistic statistical approaches to setting attribute levels in a CBC study may lead to profiles that are unrealistic or impractical. Menus and their framing that are unlike familiar real market menus can jar the subject and invite cognitive framing that differs from the frames that express preferences and drive choices in market settings. Considerable care is needed to balance statistical objectives with realism of the experiment, see Huber and Zwerina (1996).

2.1.5 Attribute formatting

The formatting, clarity, and prominence of attributes and prices of products are critical aspects of real market environments, and are correspondingly important in the hypothetical market menus in CBC studies. Advertising and point-of-sale product presentations in real markets often feature "hooks" that attract the consumer's attention and make products appealing, and understate or shroud attributes that may discourage buyers. Thus, prices may not be prominently displayed, or may be presented in a format that shrouds the final cost; e.g., promotions of "sales" or "percent-off" discounts without stating prices, statements of prices without add-ons such as sales taxes and baggage fees, subscriptions at initial "teaser" rates. Products like mobile phones, automobiles, and hospital treatments are often sold with total cost obscured or shrouded through ambiguous contract terms, through a two-part tariff that combines an access price and a usage fee, or through framing (e.g., "pennies a day"). A CBC study that is reliable for forecasting evidently needs to mimic the market in its presentation of product costs, incorporating the same attention-getting, persuasion, ambiguities and shrouding that consumers see in the real market.[3] However, a CBC study whose goal is to estimate consumers' "genuine" willingness to pay for attributes, rather than forecast market demand, needs the subjects to understand the tradeoffs clearly without masking information in the way that often occurs in real markets.

---

[3]A CBC study may embed alternative presentations of product features and costs, so that the influence of presentation format on choice can be quantified and predicted.

Prominence and ease of comparison are known to be factors that influence the attention subjects give to different aspects of decision problems; for example, market researchers sometimes report that subjects in their stated choices systematically place more weight on price relative to other product attributes than they do in real markets,[4] perhaps because this dimension is clearly visible and comparisons are easy in a conjoint analysis menu, whereas prices in real markets often come with qualifications and may not be displayed side-by-side. Widespread folklore is that subjects have trouble processing more than six attributes and more than four or five products, and begin to exhibit fatigue when making choices from more than 20 menus, see Johnson and Orme (1996). However, other researchers argue that many more attributes and products can be included, as long as the subjects understand how the attributes affect them, see Louviere *et al.* (2010) and Hensher *et al.* (2011).[5] To achieve verisimilitude with complex markets such as those for houses, cars, and jobs, and induce similar decision protocols, it can be desirable to allow CBC menus to be comparably complex. Often conjoint analysts will limit the dimensionality of the attribute profile, presenting *incomplete profiles* in which subjects are explicitly or implicitly asked to assume that in excluded attribute dimensions, the products are comparable to brands currently in the market and/or the omitted attributes are held equal across products. This *fill-in problem* leaves subjects free to make possibly heterogeneous and unrealistic assumptions about these omitted attributes, or asks them to digest and remember lengthy specifications for omitted attributes and their assumed levels. These design restrictions may make responses easier to analyze, but they will degrade prediction if correlations of included attributes with the fill-ins actually assumed

---

[4]Alternatively, subjects may place less weight on price if payment is hypothetical, or real but made from discounted house money, in the CBC exercises.

[5]In a personal communication, David Hensher says that what matters in both an SP and an RP setting is attribute relevance. Complexity is often in the eyes of the respondent, defined by what matters to them. To exclude relevant attributes in itself creates an element of complexity, since there is a mismatch between what people take into account and what the analyst imposes on them. Hensher says that he often uses a lot of attributes and 13–14 is not uncommon, and that "What makes it relevant is the way that we design the survey instrument so that respondents can activate attribute processing rules to accommodate what matters to them."

by respondents are not reproduced in the real market; see Islam *et al.* (2007).

In real markets with many complex products, such as the market for houses, consumers often use *filtering heuristics*, screening out alternatives that fail to pass thresholds on selected aspects before choosing among the survivors, see Swait and Adamowicz (2001). If the primary focus of the conjoint study is prediction, then the best design may be to make the experiment as realistic as possible, with approximately the same numbers and complexity of products as in a real market, and possibly sequential search, so that consumers face similar cognitive challenges and respond similarly even if decision-making is less single-minded than neoclassical preference maximization. However, if the primary focus is measurement of consumer welfare, there are deeper problems in linking well-being to demand behavior influenced by filtering. While it may be possible to design simple choice experiments that eliminate filtering and give internally consistent estimates of willingness-to-pay for product attributes, there is currently no good theoretical or empirical foundation for applying these tradeoffs to assess well-being in real markets for complex products where filtering influences consumer behavior.

2.1.6 Elicitation frame

The canonical elicitation frame for a CBC study asks the subject which alternative they would choose in a menu if it were offered in a real market, sometimes with the incentive that with some probability their selection will be fulfilled. This resembles real market choice environments, and experience is that it elicits responses that for familiar products are predictive for future market behavior. Studies of elicitation frames other than first choice, such as rating products on some scale, adjusting some attribute (e.g., price) to make alternatives indifferent, or ranking products from best to worst seem to induce cognitive "task-solving" responses different from the task of maximizing preferences.[6] Subjects seem to approach less familiar choice tasks as if they were school exams — they cast about for "correct" answers by making inferences on what

---

[6]See Wright and Kriewall (1980), Chapman and Staelin (1982), and Elrod *et al.* (1992).

the experimenter is looking for. While some may use their responses to air opinions, most give honest answers, but not necessarily to the question posed by the experimenter. They may "solve" problems other than recovering and stating their true preferences, indicating instead the alternative that seems the most familiar, the most feasible, the least cost, the best bargain, or the most socially responsible, see Schkade and Payne (1993).

For some products, it would seem that unfolding menu designs with elicitation of stated choices within groups of similar products, followed by elicitation of stated choices across these groups, or alternately designs in which subjects state conditional choices across groups, followed by stated choices that "customize" a preferred product in a chosen group, mimic market decision-making experience. For example, the last format mimics the real market choice process for products like automobiles or computers where the buyer picks a product provisionally, then chooses options, with recourse if customization is unsatisfactory. However, even elicitation formats that are apparently reasonably close to a market experience, such as asking for the next choice from a menu if the first is unavailable, may generate inconsistent responses as a result of anchoring to the initial response, self-justification, or a "bargaining" mind-set that invites strategic responses, see McFadden (1981), Green *et al.* (1998), Hurd and McFadden (1998), List and Gallet (2001), Louviere (1988), and Hanemann *et al.* (1991).

2.1.7 Incentive alignment

The idea behind incentives is that when subjects have a realistic chance of really getting what they say they prefer, and they understand this, they have a positive disincentive to misrepresent their preferences and risk getting an inferior outcome. Economic theorists have developed a number of mechanisms for incentive-compatible elicitation of preferences. The simplest offers the subject a positive probability that every stated choice will result in a real transaction. If subjects understand the offer, the probabilities are sufficiently high so that the subject does not dismiss them as negligible, and subjects view the transactions as being paid for from their own budgets rather than in terms of "house money" that

they feel is not really theirs, then it is a dominant strategy for the subject to honestly state whether or not a product with a given profile of attributes is worth more to them than its price; this is shown by Green *et al.* (1998) for a leading variant of this mechanism due to Becker *et al.* (1964). However, experimental evidence is that people have difficulty understanding, accepting, and acting on the incentives in these mechanisms.[7] Thus, it can be quite difficult in practice to ensure incentive alignment in a CBC, or to determine in the absence of strong incentives whether subjects are responding truthfully. Fortunately, there is also considerable evidence that while it is important to get subjects to pay attention and answer carefully, they are mostly honest in their responses irrespective of the incentives offered or how well they are understood, see Bohm (1972), Bohm *et al.* (1997), Camerer and Hogarth (1999), Yadav (2007), and Dong *et al.* (2010).

Where possible, conjoint studies should be designed to be *incentive-compatible*, so that subjects have a positive incentive to be truthful in their responses. For example, suppose subjects are promised a Visa® cash card, and then offered menus and asked to state whether or not they would purchase a product with a profile of attributes and price, with the instruction that at the end of the experiment, one of their choices from one of their menus will be delivered, and the price of the product in that menu deducted from their cash card balance. If they never choose the product, then they get no product and the full Visa balance. If subjects

---

[7]Several mechanisms have been shown to be incentive-compatible in circumstances where choices involve social judgments as well as individual preferences; for example, McFadden (2012) shows how the Clark–Groves mechanism can be used in an economic jury drawn at random from the population to decide on public projects. Small transaction probabilities are an issue. Suppose a CBC experiment on choice among automobiles offers a one in ten-thousand chance of receiving either a car with a selling price of \$40,000 or \$40,000 in cash, depending on stated choice. If the true value of the car to this subject is \$V, then a truthful response is to choose the car if and only if \$V > \$40,000. If a consumer declines the car when \$V > \$40,000, then his expected loss is \$V/10,000–\$4, a small number. This incentive is still enough in principle to induce the rational consumer to state truthfully whether or not he prefers the car. However, misperceptions of low-probability events and attitudes toward risk may in practice lead the consumer to ignore this incentive or view it as insufficient to overcome other motivations for misleading statements.

understand, perhaps through training or experience, that it is in their interest to say they would choose a product if and only if its value to them is higher than its price, then they have a positive incentive to be truthful, and the experiment is said to be *incentive-aligned* or *incentive-compatible*.

In many situations it will not be practical to provide an incentive-compatible format while maintaining the objectives of the analysis. For example, the researcher might want to consider combinations of attributes that have not yet been manufactured. Or, the cost of providing a product at any meaningful probability is prohibitive. For example, automobile manufacturers might want to test consumers' reactions to new features during the design phase of the manufacturers' product development. For existing but expensive products such as cars, it is impractical to offer subjects enough money to buy a new car and then provide them one of their chosen cars at the listed price. A lottery may be incentive-compatible in principle, but the probabilities required to make it practical (e.g., a one in a thousand chance of being offered the vehicle you chose in one menu at its stated price) may be so low that subjects do not take the offer seriously. While incentive compatibility is desirable, CBC experiments without incentive compatibility have been found to be predictive in many settings, see Dong *et al.* (2010).

2.1.8   Subject training

Extensive experiments from cognitive psychology show that context, framing, and subject preparation can have large, even outsize, effects on subject response. It is particularly important that subjects have familiarity with the products and features they are being asked to evaluate that is comparable to their real market experiences, as attention, context, and framing effects are particularly strong when subjects are asked to respond in unusual or unfamiliar circumstances. Familiarity may be automatic if the target population is experienced users of a particular line of products. For inexperienced users, tutorials on the products and hands-on experience can reduce careless or distorted responses, but may also influence stated preferences in ways that reduce forecasting accuracy.

It is useful to recognize that training of subjects can occur at several levels, and that training can manipulate as well as educate, leading to unreliable demand predictions. First, subjects have to get used to answering questions that may be difficult or intrusive, and learn that it is easier or more rewarding to be truthful than to fabricate. Some of this is mechanical: practice with using a computer for an internet-based conjoint survey, and moving through screens, buttons, and branches in a survey instrument. Second, subjects need to be educated as to what the task of stating preferences is. Subjects can be trained in "Decision-Making 101" how to optimize outcomes with assigned preferences, and how to avoid mistakes such as confusing the intrinsic desirability of a product with process issues such as availability or dominance by alternatives. Such training can be highly manipulative, leading to behavior that is very different from and not predictive for real market choices. But real markets are also manipulative, providing the "street" version of "Decision Making 101" that teaches by experience the consequences of poor choices. The goal of a conjoint study designed for prediction should be to anticipate and mimic the training that real markets provide. Third, the study designer needs to determine what information will be conveyed to the subject, in what format, and assess what information the subject retains and understands. Typically a conjoint survey begins by describing the types of products the subject will be asked to evaluate, their major attributes, and the structure of the elicitation, asking for most preferred alternatives from a series of menus. Details may be given on the nature and relevance of particular attributes. Particularly important are sufficient explanations of unfamiliar attributes, and where possible a cross-walk between described attribute levels and historical or trial experience, to allow informed choice. Instructions may be given on the time the subject has to respond, and what rules they should follow in answering. For example, the survey may either encourage or discourage the subject from consulting with other family members, finding and operating past products in the same line, or consulting outside sources of information such as internet searches. Finally, subjects need to be instructed on the incentives they face, and the consequences of their stated choices. At various stages

in subject training, they may be monitored or tested to determine if they have acquired information and understood it. For example, a protocol in market research called *information acceleration* gives subjects the opportunity to interactively access product descriptions, consumer reviews, and media reports, and through click-stream recording and inquiries during the choice process collects data on time spent viewing information sources and impact on interim propensities. This protocol seems to improve subject attention and understanding of product features, and also identify the sources and content of information that has high impact on stated choices, see Urban *et al.* (1990) and Urban *et al.* (1997).

In summary, while training may educate subjects so they are familiar with the products being compared, it is difficult to design training that is neutral and non-manipulative. Real markets are in fact often manipulative, via advertising and peer advice, and one goal for CBC is to achieve accurate prediction by mimicking the advertising and other elements of persuasion the consumer will encounter in the real market. One caution is that particularly in cases where preferences are not well formed in advance, subjects will be particularly vulnerable to manipulation, and training that embodies manipulation that is not realistic risks inducing stated responses that are not predictive for real market behavior.

2.1.9 Calibration and testing

Where possible, CBC results should be tested against and calibrated to consumer behavior in real markets. In some cases, CBC menus will coincide with product offerings in existing markets, or can be designed to include such products. In this case, it is useful to compare models estimated from the CBC study and the market data to assess whether people are weighing attributes similarly. Calibration is not limited to specific targeted products or policies; it may in fact be helpful to test or impose consistency in stated preference data across disparate choice situations when consistency is expected from rational

consumers.[8] Improved forecasts may be obtained by imposing real market constraints such as product shares on the estimation of choice models from CBC data, by calibrating CBC model parameters to satisfy market constraints, or by combining CBC and market choice data and estimating a combined model with scaling and shift parameters for CBC data as needed. CBC studies should when possible embed tests for response distortions that are commonly observed in cognitive experiments, such as anchoring to cues in the elicitation format, reference point or status quo bias, extension neglect, hypersensitivity to context, and shadowing from earlier questions and elicitations. While some of these cognitive effects also appear to influence market choices, many are specific to the CBC experience and have the potential to reduce forecasting accuracy. Ideally, a CBC study with adequate calibration will not show much sensitivity in its bottom-line forecasts to these sources of possible response distortion.

An advantage of CBC experimental designs is that through the presentation of a slate of menus, they give an opportunity to test the consistency of individual stated choices with neoclassical preference theory, to confront respondents and ask them to explain and reconcile stated choices, and to incorporate menus that allow direct cross-validation between stated and revealed market choices. For example, menus can be offered that allow for the possibility of testing whether stated choices are consistent with the axioms of revealed preference, and specifically whether they violate the transitivity property of preferences. For example, even under the relaxed standard that consumers have stochastic preferences with new preference draws for each choice, their responses can be tested for the regularity property that adding alternatives cannot increase the probability that a previous alternative is chosen. If menus contain current market alternatives, and past purchase behavior of the subjects is known, then one can test whether revealed and stated

---

[8]In a personal communication, Danny Kahneman suggests that stated preferences for environmental policies might be improved by eliciting rankings of values for disparate policy interventions, possibly including past social decisions that establish calibrating precedents. The task of constructing valuation scales from such data is similar to that of developing psychological scales from test inventories where item responses are noisy.

preferences for the same alternatives are consistent in their weighting of attributes. For example, Morikawa *et al.* (2002) find that there are systematic differences in weights given to attributes between stated and revealed choices, and that predictions from stated choices can be sharpened by calibrating them to revealed preferences; see also Ben-Akiva and Morikawa (1990) and Brownstone *et al.* (2000), and Hensher and Bradley (1993). This step of testing and validating CBC is important particularly in studies where verisimilitude of the conjoint menus and congruity of the cognitive tools respondents use in experimental and real situations are in question, for example, when the products being studied are complex and unfamiliar, such as choices of college, house to purchase, cancer treatment to pursue, or remedies for environmental damages.[9] For example, the battery of consistency tests conducted on the contingent valuation method of eliciting stated preferences given in McFadden and Train (2017) could in many cases be generalized and embedded in any CBC study.

In marketing applications, it is possible to validate CBC forecasts against actual market performance of new or modified products, judged by market shares in the population or in population segments. We have not found a comprehensive survey of the performance of forecasts from CBC studies. Natter and Feurstein (2002) compare consumer-level CBC forecasts with scanner data on actual purchases, and conclude that accounting for individual heterogeneity leads to market-level forecasts no better than aggregate models. However, they do not use a statistical method that accounts for the unreliable estimation of individual preference weights. Moore (2004) compares CBC with other elicitation and forecasting methods, and concludes that CBC data analyzed using the methods described in this paper out-performed the alternatives. Wittink and Bergestuen (2001) cite studies in which CBC forecasts of market shares of data terminals, commuter modes, state lottery

---

[9]A large literature compares and tests stated preference elicitation methods, and is relevant to questions of CBC reliability, see Acito and Jain (1980), Akaah and Korgaonkar (1983), Andrews *et al.* (2002), Carmone *et al.* (1978), Hauser and Urban (1979), Hauser and Rao (2002), Huber *et al.* (1993), Louviere (1988), Neslin (1981), Jain *et al.* (1979), McFadden (1994), Orme (1999), and Segal (1982), Srinivasan (1988, 1998), Bateson *et al.* (1987), and Reibstein *et al.* (1988).

products, personal computers, and fork-lift trucks are close to actual results, and conclude that "These results provide strong support for the validity of self-explicated conjoint models in predicting marketplace choice behavior". At a disaggregate level, Wittink and Montgomery (2009) use CBC data to estimate individual weights on eight attributes of jobs open to MBA graduates, and four months later observe actual job choices and the actual attributes of jobs offered. They report 63% accuracy (percent hits) in predicting the jobs students chose out of those offered, compared to a 26% expected hit rate if the students had chosen randomly. They attribute the failure to achieve higher accuracy to noise in estimates of weights for individuals, and to the influence of job attributes not included in the CBC study. They conclude: "On balance, the published results of forecast accuracy are very supportive of the value of conjoint results." They do caution that "One should keep in mind that positive results (conjoint analysis providing accurate forecasts) are favored over negative results for publication. Nevertheless, the evidence suggests that marketplace forecasts have validity."

2.2 A conjoint study example

To make CBC more concrete, consider the problem of estimating consumer preferences for table grapes, to guide growing and supermarket pricing decisions. In preliminary focus groups, it is found that in addition to price per pound, the attributes that consumers mention as important are sweetness, crispness, grape size, and whether or not the grapes are organically grown. Subjects interviewed at a mall are given eight menus, with each menu containing a no-buy alternative and three alternative types of grapes pictured and described in words giving their price per pound and attribute levels. The CBC experimental design varies price and product attributes over various levels, and assigns these randomly to bunches in each menu, making sure that the bunches offered all differ from each other in two or more dimensions. The levels in the study are $1.00 to $4.00 for the Price of a one-pound bunch, {Tart, Sweet} for Sweetness, {Soft, Crisp} for Crispness, {Small, Large} for Size, and {No, Yes} for Organic. In a parametric model of

the utility of alternatives, these "raw" attributes may enter linearly, and also as transformations such as an interaction of the sweetness and crispness attributes, (Sweetness = Tart)*(Crispness = Soft) or an interaction of a product attribute such as sweetness and subject gender, (Gender = Female)*(Sweetness = Tart). Subjects are told that at the end of the interview, one of the menus they have responded to will be drawn, and they will receive their choice from this menu plus a Visa cash card for $10 less the price of their choice. This CBC design has strong incentive alignment, provided subjects believe that the promised transaction will be carried through. The study assumes that consumers buy at most one bunch of grapes per day, and does not ask for purchase intensity. Subjects are told that checking the "no purchase" option on a menu means that they would not buy any grapes today if this were the only menu available to them in a food market or in this study, but that after the experiment they are free to purchase table grapes or not as they wish. If subjects understand and accept this frame, and do not respond strategically in the experiment even if they anticipate better deals for table grapes outside, then the experiment should predict daily market demand. Table 2.2 gives an example of a menu.

Table 2.3 illustrates data obtained from this CBC, with the attributes and levels coded as indicated in Table 2.2. "Bunch intercept" is a dummy variable that is one except for the no-purchase alternative, and "Choice" is a choice indicator for each subject and menu. The data continues with the remaining menus for the first subject, next the data for the second subject, and so on.

Table 2.2: Table grape CBC menu.

Menu 1	Bunch #1	Bunch #2	Bunch #3	No Purchase
Price/lb.	$2.50	$2.75	$3.00	$0.00
Sweetness	Tart (0)	Sweet (1)	Sweet (1)	
Crispness	Crisp (1)	Soft (0)	Crisp (1)	
Size (count per lb.)	Small (0)	Large (1)	Small (0)	
Organic?	No (0)	No (0)	Yes (1)	
Check One:	☐	☐	☐	☐

Table 2.3: Extract of data from the table grape CBC.

Subject	Menu	Alternative	Price	Sweetness	Crispness	Size	Organic?	Sweet*Crisp	Sweet*Gender	Bunch intercept	Gender	Choice
1	1	1	2.98	0	0	0	1	0	0	1	0	0
1	1	2	1.51	0	1	0	1	0	0	1	0	0
1	1	3	1.42	1	0	0	0	0	0	1	0	1
1	1	4	0	0	0	0	0	0	0	0	0	0
1	2	1	1.12	1	0	0	1	1	0	1	0	1
1	2	2	1.63	1	1	1	1	1	0	1	0	0
1	2	3	2.77	1	0	0	0	0	0	1	0	0
1	2	4	0	0	0	0	0	0	0	0	0	0

This example and its incentive scheme show some of the problems in designing a conjoint survey to closely mimic the real market, while providing the information needed to guide supplier decisions. First, the technical attributes of table grapes relevant to a supplier are measurements such as sugar content (in percent by volume), pH (unbuffered acidity on a scale of 0 for highly acid to 7 for distilled water), and berry count per pound. Consumers rely on familiarity and experience to give terms like "tart" and "soft" meaning, and not all subjects will have the same interpretation. Hence, tying stated preferences among attribute levels such as "tart" and "soft" to specific technical attributes is a problem, and it is difficult to translate stated preferences phrased in terms of these words into demand for products with specific technical attributes. Untrained subjects are unlikely to recognize technical attribute levels and be able to reliably state preferences among them. However, it may be possible, and important, to train subjects by having them taste grapes with specific technical attribute levels and attach words to them; thus, "tart" may be associated with grapes of 9% sugar, and "sweet" with grapes of 12% sugar; "crisp" associated with 4 pH grapes and "soft" with 6 pH grapes, "small" with 40 count/pound and "large" with 20 count/pound. This training could be part of a "practice" introduction to the experiment in which menus of currently available grape products with associated verbal attribute descriptions are offered, and subjects are asked to taste the different grapes and associate their preferences for grapes with various verbal attribute profiles. This "vignette" approach allows a direct association of implicit technical attributes and tastes. The feasibility of this approach depends on the availability of existing products, the variability of attribute levels in these products, the extent to which potential future products will have technical attributes whose desirability can be reasonably extrapolated from existing products, and the ability of subjects to taste grapes through multiple menus without becoming satiated or bored.

Second, the incentive alignment in the grape study, offering a chosen alternative from one menu a subject faces, will be feasible only if one of the stated choices is an existing product. Otherwise, it may not be possible to fulfill a promised transaction. Incentives can be aligned through more general promises to deliver an available product

(or cash) consistent with the subject's stated preferences, but this makes it more difficult for the subject to recognize that this makes it in their interest to be truthful. Third, whether subjects view the offered prices as reasonable, or grapes as tempting, will depend on their shopping histories and grape inventories, their expectations regarding availability and prices of grapes elsewhere, particularly awareness of and response to real market promotions and sales, and their anticipations of how they will feel if they receive grapes. Overall, the predictive power of the "no purchase" option for grapes will depend on its being given a sufficiently specific description and context, so that it corresponds realistically to the product and purchase timing options the consumer will face in the forecast market.

2.3   Stated perceptions, expectations, and well-being

The focus of CBC analysis in the literature is on preferences among products described in terms of their objective attributes, including objective probabilities of events when the nature or utility of a product is uncertain. In principle, recovery of preferences can incorporate subjective probabilities, and perceptions through studies that include existing products with objective attributes, and products whose delivery is contingent on verifiable future events. As in Ramsey (1931), Savage (1954), and Arrow (1971), if markets or experiments offer a rich family of lotteries on different events, consumers have complete, transitive, continuous preferences over contingent goods and services, and satisfy some plausible axioms on the separation of subjective probabilities of events and preferences over consequences of events, then these consumers have consistent preferences and subjective probabilities that should be recoverable using stated preference methods. However, it has been recognized since the work of Allais (1953), and emphasized by the experiments of Kahneman and Tversky (1979) and Kahneman and Tversky (2000), that consumers are hypersensitive to context when asked to perform cognitive tasks requiring personal probabilities or perceptions. Consequently, the task of recovering consistent, stable personal probabilities, and perceptions is at least as challenging as the task of recovering preferences. At this point, there is no systematic

approach to measuring perceptions that is comparable to CBC analysis
of preferences, although one can in principle conduct CBC studies of
choice among contingent events. Manski has found that stated prob-
abilities on a 100-point rating scale can be relatively consistent and
fairly predictive in some circumstances, although coding of responses
to focal points is an issue, see Manski (1999), Manski (2004), Blass
et al. (2010), and Delavande and Manski (2015). Manski's elicitation
format allows subject uncertainty in decision-making by asking for the
chances of picking particular products rather than identifying a single
choice. His argument is that if subjects consider subjective probabilities
over completing partially profiled products, then their most realistic
responses are their personal probabilities that each alternative is chosen.
He points out that subjective uncertainty here may include uncertainty
about the subject's state of mind at the time of actual future choice,
and points out that these uncertainties are *resolvable*; i.e., they will
be known in an actual choice setting. Hudomiet *et al.* (2018) intro-
duce a paired-menu elicitation format, asking for the probabilities of
a choice at two levels of a single attribute, with other attributes left
(implicitly or explicitly) unchanged, that shows considerable promise for
controlling the fill-in problem and identifying predictive causal effects. A
question for future research is whether elicitation of conditional choice
probabilities, which no longer force the subject to resolve uncertain-
ties in some manner and state a single choice, introduces cognitive
elements that are different from those operating for choice in actual
markets. The answer will depend in part on the degree to which subjects
allow the elicitation format to influence their definition of the cogni-
tive task they have been given. The Kahneman–Tversky experiments
indicate that errors in the arithmetic of personal probabilities, and
systematic biases, are likely to be intrinsic to subjective probability
judgments, and a problem for the future is to combine Manski-type
elicitations, a framework for incentive alignment, and experimental
designs like those of conjoint analysis, to map out structural models
of beliefs, personal probabilities, and risk response that are predic-
tive for market demand for new products with shrouded contingent
attributes.

3

Choice Behavior

The neoclassical economic theory of consumer choice behavior is based on the "consumer sovereignty" postulate that utility is predetermined and stable, and consumer choice maximizes utility subject to a budget constraint determined by income and prices. Utility functions are defined over bundles of personally consumed continuous and discrete market goods. The utility of a particular bundle of goods is independent of the existence or availability of alternatives; thus, independent of prices and income, and of attributes of unchosen bundles. Intertemporal dependence is allowed, with current preferences influenced by consumption history and expectations, but interpersonal influences (e.g., altruism, social network effects) are usually excluded.[1] When consumers face uncertainty regarding events that affect product attributes, additional postulates for strict neoclassical consumer theory are that consumers have realistic, statistically consistent perceptions regarding

---

[1]Most economic studies of consumer behavior treat households as unitary decision-making units with a single household utility function. While there is an extensive literature on household and social group shared and individual tastes, production and consumption, communication, bargaining, reciprocity, and decision-making, there is no widely accepted model for analyzing demand behavior and well-being of these groups.

future events, and that they handle uncertainty by maximizing expected utility. We use the term "direct utility" for an index of desirability of bundles of goods and services with specified attributes, and "indirect utility" for the level of maximum direct utility achievable with a budget determined by income and the prices of goods and services. We also use the term "decision utility" for the index of desirability that neo-classical consumers anticipate they will achieve through their market purchases, and "experience utility" for the level of satisfaction they achieve in retrospect. In general, experience utility cannot be recovered from revealed or stated demand behavior, but there are important exceptions where the gap between decision and experience utility is due solely to removal of deception or resolution of uncertainty. The target of most stated or revealed preference studies is "indirect decision utility" and its implications for choice and well-being. Preference heterogeneity leads to a distribution of market good demands and discrete choices among consumers facing the same incomes and prices. However, a single consumer's choices from different menus offered at different times or in different stages of a stated choice experiment will in the neoclassical theory be explained by maximization of a single decision-utility function. It is the structural stability of neoclassical decision utility that makes it possible to forecast the effects of policy changes on the choices of each individual, and the distribution of choices in the population.

Consumer choices from repeated menus in laboratory and market experiments often deviate from strict neoclassical theory, e.g., McFadden (1999).[2] Consumers may have trouble discriminating among similar alternatives, particularly if they have ambiguous or shrouded attributes. They may be poor statisticians, with inconsistent perceptions

---

[2]McFadden (2006) summarizes the testable implications on choice probabilities that come from a population of preference maximizers. For example, the probability of a specified alternative being chosen cannot rise when a new alternative is added to the menu, and the sum of the probabilities of choices in an intransitive cycle of length K cannot exceed $K - 1$. When these testable conditions are satisfied, behavior is consistent with a stable distribution of preferences over the population. These tests are silent on the extent to which there is heterogeneity in individual consumer preferences across different choice occasions as opposed to heterogeneity in tastes across consumers.

and expectations about the attributes of alternatives, e.g., unrealistic expectations about the probability of a product defect. They may not be relentless and exact preference maximizers, instead making choices that are influenced by whimsy, fatigue, inattention, bias toward immediate gratification, and a willingness to settle for "good enough". Also, preferences may encompass unmeasured attributes of alternatives, and may be conditioned on previous choices, experience, and social motivations that are not fully accounted for. Finally, preferences may be situational, anchored or adapted to the status quo, and sensitive to context. There is a simple device for accommodating behavior alternatives to neoclassical theory. The population of consumers may represent a mixture of people who conform to the neoclassical model, others who adopt decision algorithms that provide economies in decision-making effort or protection from severe mistakes, such as heuristics that may be broadly but not universally consistent with utility maximization (e.g., Ariely *et al.*, 2006), reductive processes such as filtering (e.g., Lleras *et al.*, 2017) or elimination-by-aspects (Tversky, 1972), or regret-minimization (Manski, 2007). In principle, the array of choice models for these consumer classes can be mixed, with latent shares of consumers in each class, to form an omnibus model of consumer behavior. There are technical problems of identification of the parameters of such a model in the extreme that some consumer classes are empty. However, econometric methods for testing non-nested hypotheses, which also use the device of mixing the non-nested models, provide solutions, see Pesaran and Weeks (2001) and Hong and Preston (2012). Further, in a Bayesian approach to analyzing CBC data, it is rather natural to consider mixtures of models, and with suitable priors, identification is not an issue.

The presence of non-neoclassical behavioral elements in consumer behavior is not necessarily a major impediment to forecasting new product demand, which requires only that the *distribution* of decision-rules in the population be recoverable from observed behavior and be stationary between the observation and forecast environments; decision rules are not required to be realistic, or to map nicely into realized well-being. However, behavioral elements break the tight neoclassical link between the utility and observed choices, complicating welfare analysis.

The early models of individual choice behavior that came out of psychophysics, Thurstone (1927), Luce (1959), and Marschak (1960), focused on stochastic elements in individual preferences originating from imperfect psychophysical discrimination. McFadden's (1974) adaptation of these models for econometric analysis of choice behavior in markets recognized both inter-consumer and intra-consumer sources of preference heterogeneity, but lumped these together since they are indistinguishable when observing a single choice by each consumer. However, when considering stated choice data from subjects confronted with a panel of menus, intra-consumer and inter-consumer preference heterogeneities can be identified, and the question is whether it is theoretically and statistically useful to do so. For *one-shot* forecasting of market demand at a single point in time, one modeling strategy is to ignore the panel data structure generated by multiple menu choices of multiple consumers, and treat each observation as the result of an independent draw from a distribution of preferences. This loses information and statistical efficiency, but is nevertheless a consistent model that in many applications will have good predictive power for market behavior. Another strategy is to retain the neoclassical assumption of stable preferences within consumers, with minor allowance for the effects of imperfect psychophysical discrimination. This *nearly neoclassical* approach exploits the information contained in multiple choices from a neoclassical consumer, and is particularly valuable if a target of the analysis is recovery of information on "permanent" individual preferences, with drift and whimsy treated as nuisance factors. However, it will give a misleading picture of individual preferences if intra-consumer heterogeneity is in fact significant, and is more than is needed for one-shot forecasting. For dynamic demand forecasting, where the behavior of preferences over time is critical, a third strategy is to introduce a structural system with both intra-consumer and inter-consumer stochastic elements in preferences, see Hess and Rose (2009), Hess and Train (2011), Yanez *et al.* (2011), Daly *et al.* (2012), and Rose *et al.* (2015). This structural approach can provide valuable information about systematic drifts in tastes that will influence choice in temporal sequences of real markets, and offers a solution to the Heckman (1981) initial values problem of identifying observed and unobserved state dependence in panel data. All the issues

and opportunities that arise in time series analysis appear in the structural approach: Is the stochastic taste process Markovian, ergodic, or described by an ARMA process? In this monograph, we will set up the consumer model using the structural approach, and give an application. However, for the most part we will concentrate on the nearly neoclassical modeling strategy, assuming that aside from errors due to the limits of psychophysical discrimination, a consumer's choices over multiple menus maximize the same preferences. In our view, this is usually the most parsimonious modeling approach to forecasting, focusing on predictive features of consumer behavior and avoiding the complexity of modeling features that are descriptive but not predictive. This approach also covers the first strategy through the expedient of treating multiple choices from one consumer as if they were single choices from multiple consumers.

3.1 Utility

We set notation and review the elements of random utility models (RUM) and utility-maximizing choices that can be applied to either revealed or stated choice data, and can be used to forecast demand for new or modified or repriced products. We follow the RUM setup of McFadden (1974) that is a common starting point for discrete choice analysis, with the generalizations and regularities developed in McFadden (2018) and innovations to facilitate analysis of CBC data from multiple menus. Table 3.1 summarizes notation, which in general also follows McFadden (2018).

Suppose a consumer faces $m = 1, \ldots, M$ menus containing a set $\mathbf{J}_m$ of mutually exclusive alternative products or services, including in general a "no purchase" alternative, and makes one choice from each menu.[3] Let z_{jm} denote the "raw" attributes and p_{jm} denote the real unit price of product j in menu m. The attributes and price of a "no purchase" alternatives will be normalized to zero unless noted otherwise.

---

[3]Assume that one alternative is the designated default if the consumer refuses to make a unique choice. The number of alternatives $|\mathbf{J}_m|$ can vary with m, but to simplify notation we often suppress m and let J denote the number of alternatives.

Table 3.1: Summary of notation.

W	Wage income
A	Net Asset income
T	Net transfer income (subsidies minus taxes)
$I = W + A + T$	Disposable income
s	Other individual characteristics, past experience, background prices
$m = 1, \ldots, M$	Menus of alternatives offered to a consumer
$j \in \mathbf{J}_m$	Alternatives offered in menu m
$\mathbf{z}_m$	Vector of attributes of menu m products, components z_{jm} for $j \in \mathbf{J}_m$
$\mathbf{p}_m$	Vector of prices of menu m products, components p_{jm} for $j \in \mathbf{J}_m$
$\rho_m = \zeta + \eta_m$	Tastes (an abstract vector, finite in applications) at menu m, where ζ represents neoclassical tastes that are invariant across menus, and η_m represents transitory tastes or whims that are i.i.d. across menus
ε	Vector of i.i.d. standard Extreme Value Type 1 (EV1) psychometric discrimination noise, with components ε_j for $j \in \mathbf{J}_m$
$\sigma(\rho_m)$	Positive inattentiveness coefficient that scales discrimination noise
$u_{jm} = U(W, A, T$ $-p_{jm}, s, z_{jm} \| \rho_m) + \sigma(\rho_m)\varepsilon_j$	Conditional decision utility of alternative j in menu m
$x_{jm} = X(W, A, T - p_{jm}, s, z_{jm})$	Vector of predetermined transformations of observed choice environments
$\beta(\rho_m)$	Vector of taste coefficients, commensurate with X, depends on ρ_m
$v_{jm} = v_{jm}(\rho_m) \equiv V(W, A, T,$ $s, z_{jm}, p_{jm}, \rho_m) \equiv X(W, A,$ $T - p_{jm}, s, z_{jm})\beta(\rho_m) - p_{jm}$	Quality-adjusted net value, the negative of quality-adjusted price
$F(\zeta\|s)$ and $H(\eta_m\|\zeta, s)$	CDF of ζ and conditional CDF of i.i.d. η_m given ζ
$F(\rho_1, \ldots, \rho_M\|s) = \int_\zeta$ $F(d\zeta\|s) \prod_{j=1}^{M} H(\rho_m - \zeta\|\zeta, s)$	CDF of the taste vectors
$n \equiv (W, A, T, s)$	All observed socioeconomic characteristics, experience
$u_{jm} = I + v_{jm}(\rho_m) + \sigma(\rho_m)\varepsilon_{jm}$ $\equiv I + X(W, A, T -$ $p_{jm}, s, z_{jm})\beta(\rho_m) - p_{jm} + \sigma(\rho_m)\varepsilon_j$	Linear additive money-metric utility
$u_{jm} = I + v_{jm}(\zeta) + \sigma(\zeta)\varepsilon_j$	Linear additive money-metric utility with neoclassical tastes; i.e., $\eta_m \equiv 0$.

The consumer has a latent taste vector ρ_m (abstract in general, finite-dimensional in practice) when confronted with a menu m that is a function of neoclassical "permanent" tastes ζ that are fixed across choice situations and non-neoclassical "transitory" whims η_m that vary across menus. The tastes ρ_m will in general vary across consumers. The consumer is also characterized by a vector of socioeconomic characteristics s that include household demographics, location, and past consumption experience, and by real disposable income I, written as $I \equiv W + A + T$, where the components are wage and salary income W, net asset income A, and net lump-sum transfer income T, the last typically negative due to taxes. The reason for the income decomposition is that in a population, choice probabilities can vary with income for three reasons: First, a consumer with given preferences may shift choices and consumption patterns as disposable income rises, a neoclassical income effect. Second, changes in wage or asset income have links to unobserved effective prices of discrete alternatives or attractiveness of features, e.g., through (unobserved) opportunity cost of leisure or the cost of financing durable purchases. Consequently, observed response of choice to changes in these income components confounds neoclassical income effects and unobserved price effects on choice. Third, observed income may be ecologically correlated with tastes across the population due to common causes, e.g., consumers with high (unobserved) rates of impatience may have lower incomes because of unwillingness to delay consumption to invest in human capital, and also have higher probabilities of choosing less durable products. This can induce a correlation of income and demand across the population even if individual demands exhibit no neoclassical income effect. Then, to the extent that the measured impact of income changes on choice is different for T than it is for W, A, or I, the response to T is more likely to reflect a pure neoclassical income effect. The reasons this matters are that neoclassical welfare analysis is greatly simplified if neoclassical income effects are sufficiently small to be neglected, and that either hypothetical or fulfilled compensation are typically adjustments to transfer income, so that compensating values are appropriately calculated with required income adjustments in terms of the transfer income component.

We assume that consumer tastes ρ_m are represented for each discrete alternative j in menu m by a *conditional indirect utility* function $U(W, A, T - p_{jm}, s, z_{jm} | \rho_m) + \sigma(\rho_m)\varepsilon_j$, obtained by optimization over the remaining market goods subject to the disposable income $I - p_{jm}$ left after buying j. In this function, ε_j is *psychometric noise*, scaled by $\sigma(\rho_m) > 0$, that comes from imperfect attention, diligence, and acuity.[4,5] Given the previous discussion, we allow the W and A components of income to enter this indirect utility as separate arguments to accommodate their possible effects on utility other than through neoclassical income effects. We assume that consumers make discrete choices j from the alternatives available to maximize conditional indirect utility. In neoclassical consumer theory, utility functions are ordinal, and interpersonal comparisons of utility differences are meaningless. However, in a population of heterogeneous consumers, it is important to represent preferences so that utilities vary regularly with ρ_m and are denominated in a metric that permits interpersonal economic comparisons. We choose to scale utility in *money-metric* or *WTP-space* units, with the implication that the marginal utility of a dollar is the same for every consumer. This regularizes the modeling of choice behavior and facilitates applications where compensating transfers are calculated for benefit–cost or welfare analysis (with social weights applied when deemed appropriate). The term "money-metric" utility is due to Samuelson (1950), who observed that indirect neoclassical utility evaluated at given prices could with an increasing transformation be expressed in units of income. Hurwicz and Uzawa (1971) relate this form of utility to measurement of consumer surplus, see also McFadden

---

[4]The price p_{jm} will be treated in the following analysis as the purchase price of product j, but it is also possible to interpret p_{jm} as an "access fee", with the conditional indirect utility U also depending on "user fees" or other nonlinear pricing structures, as well as on the underlying prices of goods available in continuous quantities. It is also possible to interpret p_{jm} as a market or implicit rental cost of a consumer durable, or as the cost of auxiliary activities required to consume product j, e.g., the cost of travel to a site where the product is located. We omit these background and nonlinear prices to simplify notation, but they can be reintroduced when interactions of discrete and continuous choices or pricing structures are a focus of study.

[5]We assume the psychometric noise ε_j for the product indexed by j is the same in every menu.

(2014a, 2014b, 2018). We standardize money-metric utility using total disposable income, but in our suggested specification with wage and asset income included as separate factors to capture unobserved prices of leisure and capital, this is equivalent to standardization using transfer income. Train and Weeks (2005) use the term "WTP-space" utility to refer to money-metric utility, as the incremental utility associated with a change in an attribute equals the consumer's WTP for this change. They recommend this scaling for stabilizing empirical measures of consumer welfare in discrete choice problems, see also Miller *et al.* (2011), Daly *et al.* (2012), and Carson and Czajkowski (2013). Train and Weeks use the terminology "preference-space" utility when it is written with an additive stochastic with a coefficient of one, a normalization introduced by McFadden (1974) and employed in most previous applications. Train and Weeks point out that the two normalizations are formally equivalent in the sense that any distribution of coefficients in preference-space translates into a distribution of WTP coefficients in WTP-space, and vice versa. However, commonly assumed distributions are often inconsistent. For example, a model in preference-space that assumes a non-degenerate log-normal price coefficient and normal non-price coefficients is not equivalent to a model in WTP-space with an assumption of normally distributed WTP coefficients. Instead, the equivalent distribution in WTP-space is of the ratio of a normal to a log-normal, which relative to a conventional normal is skewed in the direction of the mean of the normal with a heavier tail, see Yang (2008). While a normal log-normal specification could be assumed for WTP-space analysis, we have not seen this in applications. If some consumers filter alternatives based on the presence of a discrete attribute before completing utility maximization, their behavior is consistent with a high WTP for this attribute. Then, specification of a normal log-normal distribution of WTP coefficients may capture a real feature of behavior. Alternatives to money-metric scaling are to express utility in units of leisure time ("minute-metric" utility) or a market basket of goods ("basket-metric" utility). The last alternative could be used for example to scale utility in multiples of a designated poverty level. Again, these scalings are all equivalent with appropriate (non-conventional) assumptions on the distributions of coefficients. Without equivalent

distributional specifications, the various scaling alternatives give non-nested models of utility that could be discriminated using non-nested tests, or plausibly combined by mixing in an omnibus model of the distribution of WTP coefficients.

Our *money-metric* or *Willingness-To-Pay-space* specification of conditional indirect utility forces $U(W, A, T, s, j_0|\rho_m) \equiv I$ for a designated "no purchase" alternative j_0, and adopts an *additive linear* approximation to this utility,

$$u_{jm} = U(W, A, T - p_{jm}, s, z_{jm}|\rho_m) + \sigma(\rho_m)\varepsilon_j$$
$$\equiv I - p_{jm} + X(W, A, T - p_{jm}, s, z_{jm})\beta(\rho_m) + \sigma(\rho_m)\varepsilon_j. \quad (3.1)$$

Then Equation (3.1) is denominated in real monetary units with the property that $u_{j_0m} = I + \sigma(\rho_m)\varepsilon_{j_0m}$. The function $x_{jm} = X(W, A, T - p_{jm}, s, z_{jm})$ is a vector of predetermined transformations (e.g., logs, polynomials), weighted by taste coefficients $\beta = \beta(\rho_m)$, that can include nonlinear transformations and interactions of raw product attributes, interactions of raw product attributes and consumer characteristics, and dummy variables that identify types or classes of products (e.g., indicators for "brand"), with $X(W, A, T, s, j_0) \equiv 0$ by normalization in most applications. Tastes ρ_m are in general heterogeneous across consumers, and may be heterogeneous across menus, and will be treated in our most general analysis as random effects with a CDF $F(\rho_1, \ldots, \rho_M|s)$.[6] By construction, the coefficients $\sigma(\rho_m)$ and $\beta(\rho_m)$ will be predetermined transformations (e.g., linear, exponential) of the underlying taste vector ρ_m. The consumer's *Willingness-To-Pay* (WTP) vector for unit increases in corresponding components of X is given by $\frac{dp_{jm}}{dx_{jm}}|_{u_{jm}=\text{const.}} = \beta(\rho_m)/\{1 + [\partial X(W, A, T - p_{jm}, s, z_{jm})/\partial T]\beta(\rho_m)\}$. In marketing, models are often specified without imposing the money-metric scaling in Equation (3.1), and the coefficients $\beta(\rho_m)$ are termed

---

[6]The socioeconomic effects s may enter as arguments in X, as conditioning vectors in the distribution of taste parameters, or both. The former is more natural if socioeconomic characteristics change the perceived quality of attributes and levels, while the latter is more natural if socioeconomic factors shift the location of the consumer in the field of preferences. If choices influence tastes through learning, then F is structurally stable, but if the conditioning comes from selection on serially correlated tastes, one has the challenging initial values problem of Heckman (1981).

part-worths. If the functions X do <u>not</u> depend on net transfer income $T - p_{jm}$, then Equation (3.1) is of *Gorman Polar Form*, Engle curves for continuous goods are linear and parallel for all consumers, discrete choices are independent of T, showing <u>no</u> neoclassical income effect, and WTP simply equals the dollar-denominated vector $\beta(\rho_m)$, see Gorman (1953), Gorman (1961), Chipman and Moore (1980), Chipman and Moore (1990), McFadden (1981), McFadden (2004), and McFadden (2014b). Welfare analysis of discrete choice simplifies considerably when a neoclassical income effect is absent or sufficiently minor to be ignored. The $\beta(\rho_m)$ coefficients will play a central role in business or public policy analysis of the demand and consumer welfare impacts of modifications in the offerings of products. However, simple measures of central tendency such as population mean or median β's do not account for choice selection effects or neoclassical income effects, and a more careful analysis is required in many cases to translate the distributions of β coefficients into demand and welfare impacts, see McFadden and Train (2018).

Examples of $X(W, A, T - p_{jm}, s, z_{jm})$ functions from a study of automobile choice are horsepower divided by vehicle weight and the number of seats in the vehicle less the number of persons in the consumer's household. Another example is the fuel efficiency of an automobile alternative, with a coefficient interpreted as the value of a unit increase in fuel efficiency, reflecting both tastes and subjective expectations on the future price of fuel. In travel mode choice, W and A may influence utility even with no neoclassical income effect — W interacts with tastes for time-saving attributes and A interacts with credit requirements for owning a car, but if there is no neoclassical income effect, X does not depend on T. X can also include dummy variables for "brand name" or for commodity classes (e.g., "buses" or "public transit" in a study of mode choice), with the associated components of $\beta(\rho_m)$ capturing the common impact of these factors on the utilities of various choices. One implication of X transformations that interact with consumer characteristics is that a product modification that changes a raw attribute z can have an impact on multiple components of X, and these impacts can vary across consumers.

The requirements of first-stage utility maximization are that the indirect utility Equation (3.1) be homogeneous of degree zero and quasi-

convex in nominal income and prices, including the market prices of the goods demanded in the first-stage optimization, increasing in income, and non-increasing in all prices, see McFadden (1974), Deaton and Muellbauer (1980), Fosgerau *et al.* (2013), McFadden (2014b), and McFadden (2018). The homogeneity requirement is met by writing all income components and prices in real rather than nominal terms, but the quasi-convexity requirement is a substantive restriction on the function forms of the X transformations. Quasi-convexity can be forced on Equation (3.1) by requiring that components of X for which the corresponding component of $\beta(\rho_m)$ is unrestricted in sign (resp., restricted positive, restricted negative) be linear (resp., convex, concave) in background prices (and in $T - p_{jm}$ if it is an argument in X). However, these restrictions are stronger than necessary, and it will be more practical to not impose them a priori, and instead test estimated utility functions for the required quasi-convexity on the empirically relevant domain.

An issue that is important for predicting demand for new products is whether X depends only on "generic" attributes of products, or also on "branded" attributes. Here, "generic" means that consumers respond to the attribute in a new product in the same way as in an existing products, e.g., "schedule delay time" (the time interval between when the consumer would like to leave and when the mode is scheduled to leave) in a study of transportation mode choice has the same definition for both existing and new modes. Class indicators in x_{jm} may be either generic (e.g., "public transit") or branded (e.g., "Red Bus Company"). A "branded" attribute interacts with a "brand" class dummy, so that the associated WTP can depend on the "brand" of the product, e.g., WTPs may differ for "Red Bus" and "Blue Bus" schedule delay time. While branded attributes may provide a parsimonious "explanation" of tastes, they are shorthand for deeper generic attributes, such as the comfort of the waiting environment in the mode choice example. When brand-specific effects are not easily quantified in measures of perceived reputation, these shorthand variables may be convenient, or expedient when it is clear how to extend branded attributes to new products, e.g., when new products are introduced as new lines in an existing brand. However, the analyst should recognize that recasting "brand" in terms of

generic attributes such as reputation for reliability or durability extends the ability of analysis to forecast demand for innovative products.

Consider the roles of the latent taste parameter vector ρ_m and the disturbances ε_j in Equation (3.1). The ε_j are interpreted as perturbations in utility that vary from one consumer and alternative to the next, and come from the psychometric difficulty consumers have in distinguishing between products and attribute levels, from whimsy, from lack of consumer attention and acuity, and from mental mechanisms consumers use to "break ties" when the utility levels of alternatives seem indistinguishable. The ρ_m locates the consumer in the field of possible preference orders. They reflect "permanent" tastes, and for a strictly neoclassical consumer are the same for each menu faced. However, more generally they may vary across menus due to non-neoclassical decision elements. If the preference field can be represented as a linear vector space, the combination of inter-consumer and intra-consumer heterogeneity can be modeled as $\rho_m = \zeta + \eta_m$, where ζ characterizes permanent tastes with CDF $F(\zeta|s)$, and the η_m, independently identically distributed with CDF $H(\eta_m|\zeta, s)$, capture perturbations in tastes from one menu to the next.[7] The joint CDF of the taste parameters in Equation (3.1) is then $F(\rho_1, \ldots, \rho_M|s) = \int_{-\infty}^{+\infty} H(\rho_1 - \zeta|\zeta, s) \ldots H(\rho_M - \zeta|\zeta, s) F(d\zeta|s)$. A strictly neoclassical consumer with sharp psychophysical discrimination corresponds to $\mathrm{H}(\eta_m|\zeta, s)$ having unit probability at $\eta_m = 0$ and a tiny positive scaling factor $\sigma(\zeta)$ in (3.1) so that the ε_{jm} just provide a tie-breaking mechanism. This setup will also describe a consumer with some non-neoclassical elements in decision-making, such as unrealistic perceptions or dependence of tastes on social motives that lie outside the neoclassical assumptions, so long as this consumer behaves consistently across menus "as if" she had fixed tastes. This is the case favorable for forecasting future market demand on which we will concentrate, with observations from each consumer for multiple menus consistent with a

---

[7]The assumption that the η_m are i.i.d. could be generalized to allow a stochastic process governed by the rate of decay of memory, and perhaps influenced by external events. However, identification and estimation of the features of a more general process is a challenging task in time series analysis. Later, both F and H will be made functions of a "deep" parameter vector θ, and ζ and the η_m will be termed "intermediate" parameters. In a Bayesian interpretation, these parameters then have a hierarchical structure.

single decision-utility function that has the stability property that it will also be the decision-utility function in the future market. However, it is also possible to use the utility setup (3.1) to model joint inter-consumer and intra-consumer heterogeneity in decision utility, with coefficients written as $\sigma(\zeta + \eta_m)$ and $\beta(\zeta + \eta_m)$. One limiting case that may be expeditious for forecasting applications is "no taste memory" with $F(\zeta|s)$ having unit probability at $\zeta = 0$, so that every consumer for every menu draws tastes from the same distribution without regard for previous taste history. It would have been possible to write the model (3.1) more compactly by absorbing the terms $\sigma(\rho_m)\varepsilon_j$ into the vector ρ_m; if these effects are needed to rationalize choice, they would then appear as components of $\beta(\rho_m)$ associated with components of X that are dummy variables for the individual alternatives $j = 0, \ldots, J_m$. The justification for introducing the "redundant" $\sigma(\rho_m)\varepsilon_j$ terms sepa-rately is that isolating these as independent additive terms is useful for obtaining relatively simple expressions for choice probabilities and for computation. While the perturbations ε_j may be needed to reconcile observed choices with utility maximization, they are not an explanation of choice behavior. If their presence looms large, this says that many of the determinants of choice are unobserved and unaccounted for, and this will limit the accuracy of predictions. Ideally the $\sigma(\rho_m)\varepsilon_j$ terms have a modest effect on utility differences relative to the "systematic" component $v_{jm} \equiv X(W, A, T - p_{jm}, s, z_{jm})\beta(\rho_m) - p_{jm}$ that contains the relevant information on the consumer's choice environment. When the distribution of the taste coefficients $\beta(\rho_m)$ across consumers and menus is sufficiently broad, and X contains a sufficiently rich array of dummy variables whose coefficients capture idiosyncratic tastes for various product classes, the disturbances ε_j are not really needed to "explain" correlations between utilities of different alternatives and id-iosyncratic choices. Then one can treat the ε_j as elements that simply smooth responses and facilitate computation. In this case, it should be harmless to assume that the ε_j are independently, identically distributed with a convenient Extreme Value type 1 (EV1) distribution, and to assume that the scaling coefficient $\sigma = \sigma(\rho_m)$ is constant. However, if choice is poorly predicted in practice by the systematic component of Equation (3.1), specification of the distribution of ε_j can have a

significant impact on predicted choice probabilities, and it will become important to avoid restricted specifications such as i.i.d. EV1 that result in unrealistic choice probabilities.

The additive linear specification of the systematic component in Equation (3.1) seems quite restrictive, but McFadden and Train (2000, Theorem 1) and McFadden (2018, Theorem 5.1) prove that Equation (3.1) with sufficiently flexible specifications of the X transformations and the distribution $F(\rho_1, \ldots, \rho_M | s)$ can mimic any preference field and choice situation that meets mild regularity conditions.[8] Of course, for Equation (3.1) to be a good model of decision utility, the analyst must be prepared to introduce complex, high-order terms in X and to allow flexible characterizations of the distribution of preferences. It will be an empirical question as to whether any specific truncated basis function series approximation to random utility is sufficient to achieve desired degrees of accuracy and avoid forecasting bias due to model misspecification. In McFadden (2018), the scaling factor σ on the EV1 noise can be taken to be any sufficiently small positive constant. A constant σ assumption is particularly convenient for consumer welfare calculations. However, in applications, it may be parsimonious, or useful for interpretation in terms of heterogeneity in consumer acuity, to allow $\sigma(\rho_m)$ to vary across consumers. For the sake of generality, we will allow heterogeneous $\sigma(\rho_m)$.

The specification (3.1) of utility is predicated on the structural assumption that consumers are decision-utility maximizers, with additional restrictions if consumer behavior is assumed to be almost or strictly neoclassical. Without further restrictions, the utility is a "partial reduced form" that convolves the effects of true tastes, subjective perceptions of product attributes, and subjective probabilities of con-

---

[8]The assumptions needed for the result are (1) the maximum numbers of discrete alternatives and menus are finite, $J \leq J_{max} < +\infty$ and $M < +\infty$; (2) the domains of $(W, A, T, p_1, \ldots, p_M, s, z_1, \ldots, z_M)$ and $(\rho_1, \ldots, \rho_M)$ are finite-dimensional, closed, and bounded; (3) the probability measure on the domain of $(\rho_1, \ldots, \rho_M)$ is absolutely continuous; (4) the preference field has Lipschitz-continuity properties; and (5) there is zero probability of ρ_m such that $U(W, A, T - p_{jm}, s, z_{jm} | \rho_m) = U(W, A, T - p_{km}, s, z_{km} | \rho_m)$ for $j \neq k$. Sufficient conditions for the absolute continuity and zero probability of ties conditions are that all alternatives in menus are distinct in some attribute and that tastes for this attribute are continuously distributed.

tingent events (e.g., "Will these Bluetooth headphones work reliably, and will I enjoy using them if I buy them?"). When it is important for policy analysis to identify true tastes, perceptions of attributes that may respond to advertising or information campaigns, and subjective probabilities that may be altered by product warranties and guarantees, a combination of CBC experiments focused on subjective perceptions and probabilities and structural assumptions on Equation (3.1) may allow these components of decision-utility to be isolated. Development of a consistent system of hypotheses on choice behavior, structural assumptions on Equation (3.1), and CBC experimental designs to accomplish these tasks is beyond the scope of the current monograph, but is an important topic for future research.

3.2 Choice probabilities

Consumers with decision utility functions of the form (3.1) choose alternatives $j = 0, 1, \ldots, J$ in menu $m = 1, \ldots, M$ that give maximum utility. Abbreviate notation by defining $v_{jm} = v_{jm}(\rho_m) \equiv V(W, A, T, s, z_{jm}, p_{jm}, \rho_m) \equiv X(W, A, T - p_{jm}, s, z_{jm})\beta(\rho_m) - p_{jm}$, then $u_{jm} = I + v_{jm} + \sigma(\rho_m)\varepsilon_j$. We will call v_{jm} "*quality-adjusted net value*" or the negative of quality-adjusted price. Define $\mathbf{z}_m = (z_{1m}, \ldots, z_{Jm})$, $\mathbf{p}_m = (p_{1m}, \ldots, p_{Jm})$, and $\mathbf{v}_m = (v_{0m}, \ldots, v_{Jm})$, with $v_{0m} = 0$ by construction. The next step in discrete choice analysis is to go from (3.1) to a choice probability model consistent with its maximization. A useful tool for this analysis is the *expected maximum money metric utility*,

$$G(I, \mathbf{v}_m) \equiv \mathbf{E}_{\rho_m|s}\mathbf{E}_\varepsilon \max_{j=0,\ldots,J}\{I + v_{jm} + \sigma(\rho_m)\varepsilon_j\}$$

$$= I + \mathbf{E}_{\rho_m|s}\sigma(\rho_m) \cdot \left\{ \log \sum_{j=0}^{J} \exp\left(\frac{v_{jm}(\rho_m)}{\sigma(\rho_m)}\right)) + \gamma_0 \right\}. \quad (3.2)$$

The first form of this expression holds for any distribution of the psychometric noise, while the second form holds, with γ_0 denoting Euler's constant, when the ε_{jm} are i.i.d. EV1 distributed, see McFadden (2018, Appendix B). The argument I in G denotes direct dependence through the linear term in Equation (3.2); there will also be indirect dependence on W or A if they influence the effective prices of alternatives, and on

T through the function $v_{jm} = V(W, A, T, s, z_{jm}, p_{jm}, \rho_m)$ when there is a neoclassical income effect. The derivative of G with respect to v_{jm}, obtained by differentiating inside the expectations and the maximization operator,[9] gives the *choice probability* $P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m)$ for alternative j:

$$\partial G(I, \mathbf{v}_m)/\partial v_{km} \equiv P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m)$$
$$\equiv \mathbf{E}_{\rho_m|s}\mathbf{E}_\varepsilon \mathbf{1}(v_{jm} + \sigma(\rho_m)\varepsilon_j > v_{km}$$
$$+ \sigma(\rho_m)\varepsilon_k \text{ for } k \neq j). \qquad (3.3)$$

Due to this property, G is called a *choice probability generating function* (CPGF). Fosgerau *et al.* (2013) show that every well-behaved random utility model has a CPGF, and give necessary and sufficient conditions for a function to be a CPGF. Their approach embeds a given random utility model in a random preference field with additive perturbations of the utilities of various alternatives, and shows that the gradient of expected maximum utility with respect to these perturbations gives the choice probabilities. When there is no neoclassical income effect, G coincides with the social surplus function utilized by McFadden (1981) to characterize choice probabilities when the effect of income on indirect utility is additive, the derivative $\partial G(I, \mathbf{v}_m)/\partial v_{km}$ is identical to $-\partial G(I, \mathbf{v}_m)/\partial p_{km}$, the CPGF is the utility function of a "representative" consumer, and the CPGF gradient property is equivalent to application of Roy's identity to obtain demand shares that equal the choice probabilities. The CPGF approach is useful because it provides a tight link from expected maximum utility to choice probabilities that holds whether or not there is a neoclassical income effect. This link can be used to generate systems of choice probabilities from functions that meet the conditions to be a CPGF, and to test systems of choice probabilities with random utility maximization.

In the case of i.i.d. EV1 psychometric noise, the choice probabilities given by the gradient of the last form in Equation (3.2) are of mixed

---

[9]We use v_{jm} to denote quality-adjusted value, a scalar variable, and also as shorthand notation for the function $v_{jm}(\rho_m) = V_{jm}(W, A, T, s, z_{jm}, p_{jm}, \rho_m)$. The derivative $\partial G(I, \mathbf{v}_m)/\partial v_{jm}$ is taken with respect to the scalar variable. A reader who finds the dual use of the v_{jm} notation confusing should add an explicit perturbation w_{jm} to v_{km}, differentiate G with respect to w, and evaluate the result at $w = 0$.

(or, random parameters) multinomial logit form,

$$P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m) \equiv \mathbf{E}_{\rho_m|s} P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m, \rho_m), \quad (3.4)$$

where

$$P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m, \rho_m)$$

$$= \frac{\exp\left(\frac{x_{jm}\beta(\rho_m) - p_{jm}}{\sigma(\rho_m)}\right)}{\sum_{k=0}^{J} \exp\left(\frac{x_{km}\beta(\rho_m) - p_{km}}{\sigma(\rho_m)}\right)} \quad \text{for } j = 0, 1, \ldots, J \quad (3.5)$$

is a multinomial logit (MNL) model. A cross-price elasticity from Equation (3.5), $\partial \log P_j / \partial \log p_{km} = -(\delta_{jk} - P_k)p_{km}/\sigma(\rho_m)$, where δ_{jk} is one if $j = k$, zero otherwise, will be useful in later analysis.

A strong and potentially seriously limiting property of MNL models is *Independence of Irrelevant Alternatives* (IIA): The odds that one alternative is chosen over a second are unchanged when a third alternative is added to the menu. When the third alternative shares many features with the second alternative, but not the first (e.g., the first two alternatives are "car" and "red bus", and the new alternative is "blue bus"), it may primarily split bus choices rather than drawing from the car choice, and the IIA property is unrealistic. To avoid poor predictions due to IIA, Equation (3.4) must be specified with flexible x_{jm}, and latent tastes ρ_m must be allowed to be heterogeneous across consumers, as in Ben-Akiva and Bolduc (1996). With this flexibility, latent taste heterogeneity overwhelms the IIA property and the probability (3.4) can approximate the choice probability from any random utility model. However, often in applications the taste parameters $(\sigma(\rho_m), \beta(\rho_m)) = (\sigma, \beta)$ are assumed to be the same for all consumers and menus, and Equation (3.4) reduces to a *flat MNL* model where IIA is an important issue,

$$P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m, \sigma, \beta)$$

$$= \frac{\exp\left(\frac{x_{jm}\beta - p_{jm}}{\sigma}\right)}{\sum_{k=0}^{J} \exp\left(\frac{x_{km}\beta - p_{km}}{\sigma}\right)} \quad \text{for } j = 0, 1, \ldots, J. \quad (3.6)$$

This model is often successful in forecasting despite its restrictive IIA feature, but if IIA is rejected empirically, better modeling options are to

alter the assumed distribution of the disturbances ε_j, say by adopting a Multivariate Extreme Value distribution for these disturbances, leading in applications to various nested logit models, or more generally, to allow random parameters preference heterogeneity as in Equation (3.4). Current estimation software makes the model (3.4) practical.

When a population of consumers is considered, inter-consumer heterogeneities in tastes will appear in Equations (3.1) and (3.4) as heterogeneities in $\sigma(\rho_m)$ and $\beta(\rho_m)$. Possible modeling strategies when confronted with choice data generated by Equation (3.4) are (A) to impose homogeneous (σ, β), obtaining the flat MNL model (3.6) in which any inter-consumer heterogeneities that are present are "subsumed" in the ε_j; (B) to assume $\rho_m = \zeta$ is a separate "fixed effects" parameter vector for each consumer, with no intra-consumer homogeneity, so that $\eta_m = 0$; or (C) to specify the $\rho_m = \zeta + \eta_m$ as "random effects" with CDF's $F(\zeta|s,\theta)$ and $H(\eta_m|\zeta,s,\theta)$ depending on deep parameters θ. In strategies (B) and (C), unobserved variations across consumers in tastes for product classes (e.g., "SUV's" in automobile choice, "buses" in commuter mode choice) can be introduced as additive fixed or random effects embodied in the coefficients of indicators in x_{jm}. In (C), the intermediate parameter vectors ζ and η_m that characterize preferences can be interpreted as a hierarchical mixture, given θ. In working with choice data where there is a single or limited number of observed choices for each consumer, strategy (A) is practical and often surprisingly good at predicting demand despite IIA rigidity and the failure to capture taste heterogeneity. The reason is that this model will often successfully capture predictive central tendencies in tastes even when it does a poor job of representing dispersion. The alternative (B) is unworkable, with no identification of individual consumer taste parameters η_m and poor identification of taste parameters ζ in choice data from a limited number of menus. In (C) the mixture distributions $F(\zeta|s,\theta)$ and $H(\eta_m|\zeta,s,\theta)$ will be identified under fairly mild conditions on the distributions of attributes and the structure of utility. Alternative (C) is practical and sufficient for many purposes, particularly in the nearly neoclassical case where intra-consumer heterogeneity is absent; i.e., the H have unit probability at $\eta_m = 0$ and ζ characterizes tastes. Train (2009, Chs. 11 and 12) discusses estimation of individual consumer parameters under

classical estimation and under Bayesian estimation. Huber and Train (2001) describe and compare both procedures, see also Hensher and Greene (2003).

Let d_{jm} denote an indicator that is one if the consumer chooses alternative j from menu m, zero otherwise, $\mathbf{d}_m = (d_{0m}, \ldots, d_{Jm})$, and $\mathbf{d} \equiv (\mathbf{d}_1, \ldots, \mathbf{d}_M)$ denote a portfolio of choices for this consumer. Define $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_M)$ and $\mathbf{p} = (\mathbf{p}_1, \ldots, \mathbf{p}_M)$. The probability of a portfolio $\mathbf{d}$ implied by the model (3.4) under the random-parameters modeling strategy (C) is

$$P(\mathbf{d}|W, A, T, s, \mathbf{x}, \mathbf{p})$$

$$\equiv \int_\zeta \prod_{m=1}^M \prod_{j=0}^J \left[\int_{\eta_m} P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m, \zeta + \eta_m) H(d\eta_m|\zeta, s) \right]^{d_{jm}}$$

$$\cdot F(d\zeta|s)$$

$$\equiv \mathbf{E}_{\zeta|s} \prod_{m=1}^M \prod_{j=0}^J [E_{\eta_m|\zeta,s} P_j(W, A, T, s, \mathbf{x}_m, \mathbf{p}_m, \zeta + \eta_m)]^{d_{jm}}$$

$$\equiv \mathbf{E}_{\zeta|s} \prod_{m=1}^M \prod_{j=0}^J \left[E_{\eta_m|\zeta,s} \frac{\exp\left(\frac{x_{jm}\beta(\zeta+\eta_m)-p_{jm}}{\sigma(\zeta+\eta_m)} \right)}{\sum_{k=0}^J \exp\left(\frac{x_{km}\beta(\zeta+\eta_m)-p_{km}}{\sigma(\zeta+\eta_m)} \right)} \right]^{d_{jm}} . \qquad (3.7)$$

In the nearly neoclassical case with the $\eta_m \equiv 0$, Equation (3.7) reduces to

$$P(\mathbf{d}|W, A, T, s, \mathbf{x}, \mathbf{p})$$

$$\equiv \mathbf{E}_{\zeta|s} \prod_{m=1}^M \prod_{j=0}^J \left[\frac{\exp\left(\frac{x_{jm}\beta(\zeta)-p_{jm}}{\sigma(\zeta)} \right)}{\sum_{k=0}^J \exp\left(\frac{x_{km}\beta(\zeta)-p_{km}}{\sigma(\zeta)} \right)} \right]^{d_{jm}} , \qquad (3.8)$$

and when there is a single menu m so that the expectation can be moved inside, reduces further to

$$P(\mathbf{d}_m|W, A, T, s, \mathbf{x}_m, \mathbf{p}_m)$$

$$\equiv \prod_{j=0}^J \left[\mathbf{E}_{\zeta|s} \frac{\exp\left(\frac{x_{jm}\beta(\zeta)-p_{jm}}{\sigma(\zeta)} \right)}{\sum_{k=0}^J \exp\left(\frac{x_{km}\beta(\zeta)-p_{km}}{\sigma(\zeta)} \right)} \right]^{d_{jm}} . \qquad (3.9)$$

Then Equations (3.7) and (3.8) are mixed MNL models for portfolios of choices. The nearly neoclassical model (3.8), with its generalization (3.7) to allow non-neoclassical intra-consumer heterogeneity, or its specialization (3.6) to flat MNL, will be the models we will estimate using CBC data, and then use in Equation (3.9) to forecast demand for new or altered products. A consequence of Equation (3.8) that is useful for some applications is that the probability that nearly neoclassical tastes ζ are contained in any measurable set $\mathbf{C}$, given a vector of choices $\mathbf{d}$, is

$$
F(\mathbf{C}|s,\mathbf{d}) = \frac{\int_{\zeta \in \mathbf{C}} \prod_{m=1}^{M} \prod_{j=0}^{J} \left[\frac{\exp\left(\frac{x_{jm}\beta(\zeta)-\mathrm{P}_{jm}}{\sigma(\zeta)}\right)}{1+\sum_{k=1}^{J} \exp(\frac{x_{km}\beta(\zeta)-p_{km}}{\sigma(\zeta)})} \right]^{d_{jm}} F(d\zeta|s)}{\int_{\zeta} \prod_{m=1}^{M} \prod_{j=0}^{J} \left[\frac{\exp\left(\frac{x_{jm}\beta(\zeta)-p_{jm}}{\sigma(\zeta)}\right)}{1+\sum_{k=1}^{J} \exp\left(\frac{x_{km}\beta(\zeta)-p_{km}}{\sigma(\zeta)}\right)} \right]^{d_{jm}} F(d\zeta|s)}.
$$

(3.10)

4

Choice Model Estimation and Forecasting with CBC Data

The CBC elicitation format produces data on choices from hypothetical market experiments which must then be analyzed to model preferences and forecast demand. A natural approach to these tasks is to treat the choice data from a CBC study like observed market choice data, with allowance for the portfolios of choice observations that come from responses to multiple menus. Such data may circumvent two limitations of real market observations — lack of independent variation in some attribute dimensions and levels and in prices, due to the fact that in equilibrium markets offer only products that are sufficiently attractive to consumers to be viable, and endogeneity of product features and prices as a result of market equilibrium.[1] The key questions are whether experiments with hypothetical markets and incentive structures can be designed to give context, incentives, and information similar enough to real markets to elicit the same perceptions, expectations, and cognitive processes, so that stated choice data from these hypothetical markets

---

[1]When the disturbances that determine an individual consumer's choice have components that do not "net out" at the market level, equilibrium market level prices will depend on these components, and cannot be treated as predetermined in estimating the choice model parameters, see Berry *et al.* (1995), Berry *et al.* (1998), and Berry *et al.* (2004).

are predictive for choice behavior in real markets, and whether tests of the predictive validity of forecasts based on stated choice data establish a domain in which they are reliably accurate.

When the data generation process for subject n is described by the nearly neoclassical model (3.8) written as a non-linear function of deep parameters θ, a traditional econometric approach to estimating θ is maximum likelihood estimation. Evaluation of Equation (3.8) requires a relatively high-dimensional integration that cannot in general be performed analytically, requiring either numerical integration or simulation methods such as maximum simulated likelihood (MSL) or method of simulated scores (MSS). An alternative (and asymptotically equivalent) approach is hierarchical Bayesian estimation, which avoids the iterations necessary to maximize likelihood, but requires computationally intensive (and usually iterative) sampling from the Bayesian posterior. We discuss and compare these approaches with estimation in Sections 5–7. For forecasting, it is necessary to consider a sample from the population of interest, which may be the sample of CBC subjects weighted to match known population characteristics (e.g., size, age distribution, income distribution), or another real or simulated population sample. For each person in this forecasting sample, the choice probabilities (3.9) are simulated for values of the explanatory variables that will prevail in the forecast marketplace, including the attribute levels and prices of the products offered in this marketplace. Generally, the required simulations are similar to those used in the estimation stage.

A first step in an analysis program is to make the model (3.8) more concrete by specifying the predetermined transformations X of raw attributes and consumer characteristics, the predetermined transformations $\sigma(\zeta)$ and $\beta(\zeta)$ of the intermediate parameter ζ, and the mixing distribution $F(\zeta|s,\theta)$ and the socioeconomic characteristics s and deep parameter vector θ on which it depends. Often, these specifications will be application-specific, selected to be parsimonious in parameters and focused on attributes of primary interest. They can be modified through specification testing during estimation. While exact model specification strategies may be application-specific, a reasonable general approach is to start with simple models including raw attributes and interactions

that are likely to matter, and test for the presence of significant omitted terms.[2]

A typical (but not the most general) random parameters specification assumes component-wise one-to-one mappings $\sigma(\zeta)$, $\beta(\zeta)$ that are restricted in sign for taste parameters that are signed *a priori*. An important and practical case for applications is $F(\zeta|s,\theta)$ multivariate normal, with the deep parameters θ specifying the mean and variance of this distribution; often with the mean a linear function of s. Alternately, a quite general specification is $F(\zeta|s,\theta) = C(F_1(\zeta_1|s,\theta),\ldots,F_T(\zeta_T|s,\theta);s,\theta)$, where C is any copula[3] and the F_i are the univariate marginal CDFs of the components of ζ, typically exponential, log normal, or normal with or without censoring. For example, let $\boldsymbol{\Phi}(\zeta;\Omega)$ denote a multivariate normal CDF with standard normal marginals and a correlation matrix Ω. Then $C(u_1,\ldots,u_T) \equiv \boldsymbol{\Phi}(\Phi^{-1}(u_1),\ldots,\Phi^{-1}(u_T);\Omega)$ is a multivariate normal copula that generates the multivariate distribution $F(\zeta|s,\theta) = \boldsymbol{\Phi}(\Phi^{-1}(F_1(\zeta_1|s,\theta)),\ldots,\Phi^{-1}(F_T(\zeta_T|s,\theta));\Omega(s,\theta))$. In applications, it is sometimes parsimonious to restrict $\Omega(s,\theta)$ to have a low-dimensional factor-analytic structure: Let Λ be a $T \times K$ matrix of rank K whose columns are factor loadings, scaled so that $\sum_{t=1}^{T} \Lambda_{ti}^2 < 1$, and let Ψ denote a diagonal matrix with diagonal elements $\Psi_{ii} = \sqrt{1 - \sum_{t=1}^{T} \Lambda_{ti}^2}$. Then $\Omega(s,\theta) = \Psi(s,\theta)^2 + \Lambda(s,\theta)\Lambda(s,\theta)'$ is the factor-analytic correlation matrix. Sampling from the multivariate normal copula with a correlation matrix $\Omega(s,\theta)$ can be carried out by setting $\varsigma = \Omega(s,\theta)^{1/2}\xi$, or in the case of a factor-analytic structure, $\varsigma = \Psi(s,\theta)\xi + \Lambda(s,\theta)\omega$, where ξ and ω are vectors of standard normal variates, and then setting $\zeta_i = F_i^{-1}(\varsigma_i|s,\theta)$ for each component.

It is also sometimes important to allow segmentation of the population into classes, with each class having its own distribution of ζ, see Desarbo *et al.* (1995). Segmentation may be unobserved, with the

---

[2]In the framework of maximum likelihood estimation, Lagrange Multiplier tests that can be calculated at fitted base model parameters are often practical for model specification tests.

[3]A copula is a multivariate CDF $C(u_1,\ldots,u_T)$ whose univariate marginals are all uniform $[0,1]$. Any multivariate distribution can be written in terms of its copula and marginals, and there are necessary and sufficient conditions for a function to be a copula, see Joe (1997), and Fosgerau *et al.* (2013).

probability of consumers falling in a "latent class" becoming a parameter of the problem, or may be observed, e.g., classification by age or gender. An example would be a population that is segmented by preferences for "brands" or "classes" of products. The vector x_{jmn} can contain indicators that are turned on for specific categories of products, and a class might be defined as the sub-population that has positive or negative preferences for a specific category (i.e., the component of $\beta(\zeta_n)$ is non-zero for sample members in this class, and the remaining components of $\beta(\zeta_n)$ for product categories are zero). The probability of a person falling in this class would be another parameter in θ, and within each class, all other taste parameters would have a (conventional) distribution conditioned on the class.

5

Maximum Simulated Likelihood (MSL) Analysis of CBC Data

Consider CBC data in the format described in Section 2. Suppose consumers are nearly neoclassical, with utility (3.1) and tastes $\rho_m = \zeta$ that are homogeneous across menus. Then their choices are generated by the discrete choice model (3.8). Assume this model is parameterized with deep parameters θ, and consider maximum likelihood estimation of θ. In the simplest case, the parameters $\sigma(\zeta)$ and $\beta(\zeta)$ in this model are assumed to be homogenous over the population, so that (3.8) reduces to the flat MNL model (3.6). If a flat MNL model is believed to be sufficient to explain and forecast real market choices, then, the log likelihood of the stated choice data is

$$\log L(\sigma, \beta) = \sum_{n=1}^{N} \sum_{m=1}^{M} \sum_{j=0}^{J} d_{jmn} \log \left(\frac{\exp\left(\frac{x_{jmn}\beta - p_{jmn}}{\sigma}\right)}{\sum_{i=0}^{J} \exp\left(\frac{x_{imn}\beta - p_{imn}}{\sigma}\right)} \right).$$

$$(5.1)$$

Next allow for preference heterogeneity across consumers. As discussed earlier, CBC studies normally cannot make the number of menus M large enough to make estimation of "fixed effects" taste coefficients for each consumer reliable. The most practical alternative is to make $\sigma(\zeta)$ and $\beta(\zeta)$ "random effects" drawn for each consumer from a distribution

61

$F(\zeta|s,\theta)$ that depends on deep parameters θ and may depend on some components of the socioeconomic vector s. For simulation, it is sometimes useful to assume that ζ can be written as a transformation $\zeta = W(\omega, s, \theta)$ of a random vector ω that has a "standard" CDF $\Phi(\omega)$. Assume the distribution F has a density $f(\zeta|s,\theta)$. Then the likelihood for this consumer is

$$L_n(d_n|x_n, p_n, \theta) = \int_\zeta P^n(\zeta) \cdot f(\zeta|s,\theta)d\zeta$$

$$\equiv \int_\omega P^n(W(\omega, s, \theta))\cdot\Phi(d\omega), \qquad (5.2)$$

where $P^n(\zeta) \equiv \prod_{m=1}^M \prod_{j=0}^J \left[\dfrac{\exp\left(\frac{v_{jmn}(\zeta)}{\sigma(\zeta)}\right)}{\sum_{i=0}^J \exp\left(\frac{v_{imn}(\zeta)}{\sigma(\zeta)}\right)} \right]^{d_{jmn}}$ and $v_{jmn}(\zeta) \equiv$ $X(W_n, A_n, T_n - p_{jmn}, s_n, z_{jmn})\beta(\zeta) - p_{jmn}$. The gradient of Equation (5.2) is

$$\partial L_n(d_n|x_n, p_n, \theta)/\partial\theta = \int_\zeta P^n(\zeta) \cdot \frac{\partial \log f(\zeta|s_n, \theta)}{\partial\theta} \cdot f(\zeta|s_n, \theta)d\zeta$$

$$\equiv \int_\omega [\partial P^n(W(\omega, s_n, \theta))/\partial\theta] \cdot \Phi(d\omega). \qquad (5.3)$$

Standard maximum likelihood theory states that under mild regularity conditions that will ordinarily be satisfied in this application, iteration of the BHHH condition[1]

$$\sqrt{N}(\theta^{i+1} - \theta^i) = (\Sigma_N)^{-1}\frac{1}{\sqrt{N}}\sum_{n=1}^N \left(\frac{\partial \log L_n\left(d_n|x_n, p_n, \theta^i\right)}{\partial\theta} \right),$$

$$(5.4)$$

where

$$\Sigma_N = \frac{1}{N}\sum_{n=1}^N \left(\frac{\partial \log L_n(d_n|x_n, p_n, \theta^i)}{\partial\theta} \right) \left(\frac{\partial \log L_n(d_n|x_n, p_n, \theta^i)}{\partial\theta} \right)',$$

$$(5.5)$$

---

[1]Berndt *et al.* (1974). Alternately, Gauss–Newton or Newton–Raphson iteration to the maximum can be used.

with an added step-size adjustment to avoid overshooting and accelerate convergence, is guaranteed to converge from any initially consistent estimator θ^0 to an estimator $\hat{\theta}$ that maximizes the likelihood of the sample $L(\theta) \equiv \prod_{n=1}^{N} L_n(d_n|x_n, p_n, \theta)$. Further, $\sqrt{N}(\hat{\theta} - \theta^*)$ is asymptotically normal with mean zero for the true θ^*, and a covariance matrix that is consistently estimated by $(\Sigma_N)^{-1}$. Note that this result does not guarantee convergence to the maximum likelihood estimator from all starting values for θ, nor can one in general be assured of the global strict concavity that would imply such global convergence. Consequently, it is useful to start the MLE iteration from a grid of starting values, and perhaps introduce annealing (discussed in the next section) during the iterations, to assure that search spans the parameter space and has a high probability of finding a global rather than a local maximum.

Maximum simulated likelihood (MSL) replaces $L_n(d_n|x_n, p_n, \theta) = \int_\omega P^n(W(\omega, s_n, \theta)) \cdot \Phi(d\omega)$ with the approximation $\frac{1}{R} \sum_{r=1}^{R} P^n (W(\omega_r, s_n, \theta))$, where the ω_r are R independent pseudo-random draws from the CDF $\Phi(\omega)$, and uses a commensurate approximation $\frac{1}{R} \sum_{r=1}^{R} \frac{\partial P^n(W(\omega_r, s_n, \theta))}{\partial \theta}$ to the gradient $\frac{\partial L_n(d_n|x_n, p_n, \theta)}{\partial \theta}$ to iterate to the maximum simulated likelihood using Equation (5.4), see McFadden and Ruud (1994), and Hajivassiliou and Ruud (1994). Alternately, a *method of simulated score* (MSS) uses Equation (5.4) to find a root of the score of the likelihood, with $L_n(d_n|x_n, p_n, \theta)$ and $\partial L_n(d_n|x_n, p_n, \theta)/\partial\theta$ approximated respectively by $\frac{1}{R} \sum_{r=1}^{R} P^n(\zeta_r)$ and $\frac{1}{R} \sum_{r=1}^{R} P^n(\zeta_r) \cdot \frac{\partial \log f(\zeta_r|s_n, \theta)}{\partial \theta}$, where the ζ_r are independent pseudo-random draws from $F(\zeta|s_n, \theta)$, see Hajivassiliou and McFadden (1998). If R rises at least as rapidly as $\sqrt{MN}$, then either of these simulation estimators will be asymptotically equivalent to the (intractable) exact maximum likelihood estimator. Train (2007) points out that the gradient approximations in MSL and MSS are not identical for fixed R, even when generated by the same random draws, and that for numerical stability, MSL should use its commensurate gradient approximation.

Suppose $F(\zeta|s, \theta)$ is multivariate normal with density $n(\zeta; \mu, \Omega)$, where $\Omega = \Lambda\Lambda'$ and $\theta = (\mu, \Lambda)$, and $\sigma(\zeta)$, $\beta(\zeta)$, $\delta(\zeta)$ are predetermined component-wise transformations of ζ; μ may be a linear function $\mu_0 +$

$\mu_1 \cdot s$. Differentiating, $\log n(\zeta; \mu, \Omega) \propto -\log|\Lambda| - (\zeta - \mu)'(\Lambda\Lambda')^{-1}(\zeta - \mu)/2$ implies $\frac{\partial \log n(\zeta; \mu, \Lambda\Lambda')}{\partial \mu} = \Omega^{-1}(\zeta - \mu)$ and $\frac{\partial \log n(\zeta; \mu, \Lambda\Lambda')}{\partial \Lambda} = \{\Omega^{-1}(\zeta - \mu)(\zeta - \mu)' \Omega^{-1} - \Omega^{-1}\}\Lambda$, and the gradient of Equation (5.2), given by Train (2007), is

$$\frac{\partial L_n(d_n|x_n, p_n, \theta)}{\partial \mu} = \int_\zeta P^n(\zeta) \cdot \Omega^{-1}(\zeta - \mu) \cdot F(d\zeta|s_n, \theta)$$

$$\frac{\partial L_n(d_n|x_n, p_n, \theta)}{\partial \Lambda} = \int_\zeta P^n(\zeta) \cdot \{\Omega^{-1}(\zeta - \mu)(\zeta - \mu)'\Omega^{-1} - \Omega^{-1}\}\Lambda$$

$$\cdot F(d\zeta|s_n, \theta). \tag{5.6}$$

A straightforward MSS approach to computation of Equations (5.2)–(5.6) in the multivariate normal case is to start from a large repository $r = 1, \ldots, R$ of draws of i.i.d. standard normal vectors ω_r, which are held fixed during estimation to avoid chatter and maintain stochastic equicontinuity, and iteratively consider trial values $\theta = (\mu, \Lambda)$, with $\Omega = \Lambda\Lambda'$, and values $\zeta_r \equiv \mu + \Lambda\omega_r$ for the intermediate parameter vectors ζ that enter $P^n(\zeta)$. Then approximate Equation (5.2) at θ by $\widehat{L}_n(d_n|x_n, p_n, \theta) = \frac{1}{R}\sum_{i=1}^R P^n(\zeta_r)$, Equation (5.3) by

$$\nabla_\mu \widehat{L}_n = \frac{1}{R}\sum_{i=1}^R P^n(\zeta_r)\Omega^{-1}(\zeta_r - \mu)$$

$$\nabla_\Lambda \widehat{L}_n = \frac{1}{R}\sum_{i=1}^R P^n(\zeta_r)\{\Omega^{-1}(\zeta_r - \mu)(\zeta_r - \mu)'\Omega^{-1} - \Omega^{-1}\}\Lambda, \tag{5.7}$$

and Equation (5.5) by an average over n of the outer product of the gradient (5.7), reshaped into a vector; i.e., define g_n to be a column vector that concatenates $\nabla_\mu \widehat{L}_n$ and a vector containing the lower triangular elements of $\nabla_\Lambda \widehat{L}_n$, column by column, and $\widehat{\Sigma}_N = \frac{1}{N}\sum_{n=1}^N g_n g_n'$.[2] Other treatments of the problem of estimating mixed MNL models, using MSL rather than MSS, are Ben-Akiva *et al.* (1993), Revelt and Train (1998), Train (1999), Train (2007), Train (2009), Brownstone and Train (1999),

---

[2]Technically, the independent elements in the lower triangular array Λ are vectorized and concatenated with μ to form the vector θ; the R command for doing this is $\theta < -c(\mu, \Gamma [\text{lower.tri}(\Gamma, \text{diag} = \text{TRUE})])$. The corresponding vector g_n is formed by the R command $\nabla_\Lambda \widehat{L}_n [\text{lower.tri}(\nabla_\Lambda \widehat{L}_n, \text{diag} = \text{TRUE})])$.

Brownstone *et al.* (2000), Chesher and Santos-Silva (2002), Hensher and Greene (2002), and Walker *et al.* (2007).[3]

When the target of the CBC is market prediction rather than estimates of individual preferences, there is only a modest loss of generality or efficiency in ignoring σ, β homogeneity across menus for the same consumer. Then model (5.2) reduces to the mixed MNL model (3.9) that ignores the panel structure of stated choices from multiple menus. The simulation method described above still applies, but it is necessary to use a "sandwich" formula to obtain an estimated covariance matrix of the estimates that is consistent when there are common tastes that induce correlation in choices across menus.

---

[3]Prof. Arne Rise Hole has developed STATA modules to estimate mixed logit models in preference-space and wtp-space. The modules, called "mixlogit" for preference space and "mixlogitwtp" for wtp space, are available at https://www.sheffield.ac.uk/economics/people/hole/stata or by typing "ssc install mixlogit" and "ssc install mixlogitwtp" in the STATA command window. The two modules share syntax and include options for estimation by various maximization techniques, including Newton–Raphson and BHHH; for correlated or uncorrelated random terms (coefficients in preference space, WTPs and alpha in wtp-space); and for normal and lognormal distributions.

Matlab, Gauss, and R codes to estimate mixed logits in preference-space by MSL are available on K. Train's website at: https://eml.berkeley.edu/~train/software.html.

6

Hierarchical Bayes Estimation

An alternative and/or complement to classical statistical estimation is the use of Bayes methods. These turn out to be particularly convenient for analysis of CBC data, and have been refined and applied widely in marketing, led by the contributions of Allenby and Rossi (1993, 2006), Allenby and Lenk (1994) and Rossi *et al.* (2005). Textbook treatments of Bayesian statistics can be found in Koop (2003), Gelman *et al.* (2004), and McFadden (2015) gives a primer on decision-theoretic and statistical properties. Under general conditions, the mean of the Bayesian posterior of a parameter is asymptotically equivalent to the classical maximum likelihood estimator (MLE), and the variance of the Bayesian posterior is asymptotically equivalent to the sampling variance of the MLE. This equivalence implies that a researcher can implement a Bayesian estimation procedure and treat the mean and variance of the posterior the same as classical estimates and their covariances, without necessarily adopting a Bayesian approach to statistics. This section first reviews the basics of Bayes estimation, then shows its use for CBC analysis.

6.1 Bayesian estimation

Observations $\mathbf{y}$ are realizations of a random vector that are governed by a probability law or data generation process (DGP) indexed by a parameter vector θ and described by a density or likelihood $L(\mathbf{y}, \theta)$. For example, in a study of automobile stated purchase choices, one might specify a flat MNL data generation process, so that $L(\mathbf{y}, \theta)$ would have the form (5.1). Bayesian statistical approach postulates that θ can itself be treated as a random quantity, see for example Zellner and Rossi (1984). Prior to observing $\mathbf{y}$, the researcher has beliefs on what values of θ are likely. These beliefs can be expressed in terms of a prior density $k(\theta)$. Then, the joint density of $\mathbf{y}$ and θ is $L(\mathbf{y}, \theta)k(\theta)$. By Bayes Law, the conditional density of θ given $\mathbf{y}$, which is called the researcher's *posterior* belief about θ given the data, is

$$K(\theta|\mathbf{y}) = \frac{L(y, \theta)k(\theta)}{\int_{\theta'} L(\mathbf{y}, \theta')k(\theta')d\theta'}. \tag{6.1}$$

The researcher can use the posterior to evaluate payoffs that depend on θ, or to calculate an estimate $t(\mathbf{y})$ of θ. Usually, the mean of the posterior is taken as the estimator of θ. The Bayes estimator balances the information on θ contained in the likelihood $L(\mathbf{y}, \theta)$ for the observed data and that contained in the prior $k(\theta)$. Priors can be either diffuse or tightly concentrated, and if they are tight, they can be "good", concentrated near the true value of θ, or "bad", concentrated far away from it. The Bayes estimator is asymptotically equivalent to the classical ML and MSL estimators under general conditions, see Train (2001). To see this, assume for simplicity that θ is a one-dimensional variation from a true value θ_0, and start from the log likelihood function

$$\log L(y, \theta) = \sum_{n=1}^{N} \log f(y_n, \theta), \tag{6.2}$$

where y_n is the data, $f(y_n, \theta)$ is the density for observation n, and N is sample size. Take the identity

$$1 \equiv \int_{-\infty}^{+\infty} f(y_n, \theta)dy_n \equiv \int_{-\infty}^{+\infty} \exp(\log f(y_n, \theta))dy_n \tag{6.3}$$

and differentiate it in θ to get

$$0 \equiv \int_{-\infty}^{+\infty} \frac{\partial \log f(y_n, \theta)}{\partial \theta} f(y_n, \theta) dy_n \quad \text{and}$$

$$0 \equiv \int_{-\infty}^{+\infty} \left[\left(\frac{\partial \log f(y_n, \theta)}{\partial \theta} \right)^2 + \frac{\partial^2 \log f(y_n, \theta)}{\partial \theta^2} \right] f(y_n, \theta) dy_n. \quad (6.4)$$

For regular problems, $I(\theta') = \int_{-\infty}^{+\infty} \left(\frac{\partial \log f(y_n, \theta')}{\partial \theta} \right)^2 f(y_n, \theta) dy_n$, termed the *Fisher Information*, is positive for θ' near θ. A Taylor's expansion around the true value θ_0 of θ gives

$$K(\theta|\mathbf{y})/K(\theta_0|\mathbf{y}) = \exp[S_N \cdot \sqrt{N}(\theta - \theta_0) - \frac{1}{2} I_N (\sqrt{N}(\theta - \theta_0))^2$$

$$+ \frac{1}{\sqrt{N}} k' \cdot \sqrt{N}(\theta - \theta_0)]. \quad (6.5)$$

By a central limit theorem, $S_N \equiv \frac{1}{\sqrt{N}} \sum_{n=1}^{N} \frac{\partial \log f(y_n, \theta_0)}{\partial \theta}$ converges to a normal random variable with mean zero and variance $I(\theta_0)$. By a law of large numbers and additional technical arguments, $I_N = \frac{1}{N} \sum_{n=1}^{N} \left(\frac{\partial \log f(y_n, \theta)}{\partial \theta} \right)^2$ converges to $I(\theta)$. Then, for N large, the argument in the exponent on the right-hand-side of Equation (6.5) behaves like $\sqrt{N}(\theta - \theta_0)$ times the expression $S_N + \frac{1}{\sqrt{N}} k' - \frac{1}{2} I_N \sqrt{N}(\theta - \theta_0)$. But this expression has the negative of the sign of $\sqrt{N}(\theta - \theta_0)$ when $\sqrt{N}(\theta - \theta_0)$ is large in magnitude, establishing from Equation (6.5) that the posterior mode has $\sqrt{N}(\theta - \theta_0)$ bounded. Further, the contribution of the prior, $\frac{1}{\sqrt{N}} k'$, is asymptotically negligible, so that the mode of the posterior and the MLE converge and have the same asymptotic distribution. Finally, the concentration of the density $K(\theta|\mathbf{y})$ in a neighborhood where $\sqrt{N}(\theta - \theta_0)$ is bounded, along with an assumption that the tails of $k(\theta)$ are sufficiently thin, so that the mean of the posterior is not dominated by its tails, implying that the mean of $K(\theta|\mathbf{y})$ and its γ-quantiles for γ in the interior of $(0, 1)$ all are in a $\sqrt{N}(\theta - \theta_0)$ bounded neighborhood of θ_0, and hence $\frac{1}{\sqrt{N}}$ close to θ_0 and the MLE for N large. More precise statements can be found in Ibragimov and Has'minskii (1973) and Strasser (1975). Huber and Train (2001) find that in a dataset of typical size, the MLE and Bayes procedures produce very similar estimates for a mixed logit model in preference space.

Scarpa *et al.* (2008) obtain similar results from Bayesian and MLE for mixed logit models in willingness-to-pay space. Train (2009) provides further discussion and examples.

6.2 Selecting priors and computing posteriors

For any given likelihood function, if the prior and posterior distributions have the same form but different parameters, then the prior is called "conjugate." Statistical models with conjugate priors are special, but convenient when applicable. Fink (1997) describes the conditions under which models admit conjugate priors, notably distributions in the exponential family, and discusses various transformations of the models that allow conjugate priors to be used. Gelman *et al.* (2004) give a compendium of known priors, see also the Wikipedia entry on conjugate priors. Of value in CBC applications are the Dirichlet prior that is conjugate for a multinomial likelihood,[1] and the Normal or Wishart prior that is conjugate for a multivariate normal likelihood with unknown mean or unknown covariance matrix.[2] Drawbacks of concentrating on conjugate priors are, first, that even if the posterior density is analytically determined, it is not necessarily simple to compute moments or draw simulated values from it, and, second, that natural parameterizations of models for CBC data often do not admit conjugate priors.

Usually a CBC analysis can start with a prior density and a choice model likelihood that can be expressed in terms of computationally

---

[1]If the likelihood of counts $(k_1, \ldots, k_J)$ from N trials is $\begin{pmatrix} N \\ k_1 \cdots k_J \end{pmatrix} p_1^{k_1} \cdots p_J^{k_J}$ and the prior is the Dirichlet density $\frac{\Gamma(\alpha_1 + \cdots + \alpha_J)}{\Gamma(\alpha_1) \cdots \Gamma(\alpha_J)} p_1^{\alpha_1 - 1} \cdots p_J^{\alpha_J - 1}$, then the posterior is Dirichlet with parameters $\alpha_j + k_j$.

[2]If the likelihood is formed from n draws from a normal $N(\mu, \Sigma^{-1})$ density with unknown μ or Σ, the prior for μ is normal, or the prior for Σ^{-1} is independently inverse Wishart, then the posterior distribution for μ is multivariate normal and for Σ^{-1} is independently inverse Wishart. There are alternative parameterizations and conjugate priors for Σ that may have better computational and statistical properties. For example, an LU decomposition expresses Σ as $\Sigma = LL'$, where L is lower triangular with a positive diagonal, with a normal/variance–gamma conjugate prior. For analyses of alternative distributions to the inverse Wishart assumption, see Song *et al.* (2018) and Akinc and Vanderbroek (2018).

tractable functions. Even so, except for special cases of conjugate priors, it is usually impossible to obtain analytically the mean $\int_\theta \theta \cdot K(\theta|\mathbf{y})d\theta$ or other key characteristics of the posterior density, or even to compute easily the scaling factor $A = \int_\theta L(y|\theta)k(\theta)d\theta$ that makes $K(\theta|y) = L(y|\theta)k(\theta)/A$ integrate to one. However, with contemporary computers it is practical in most CBC applications to use some combination of numerical integration and simulation methods to estimate features of the posterior density such as its mean. There is a large literature on methods for these calculations and their computational efficiency, see for example Abramowitz and Stegum (1964, chap 25.4), Press *et al.* (2007), Quarteroni *et al.* (2007, Chap 9), Neal (2011), Carpenter *et al.* (2017), and Geweke (1997). This monograph will describe several basic methods, particularly those that have proven useful for Bayes estimation, but we refer readers interested in mathematical and computational details back to these sources.

Deterministic numerical integration is often practical when the dimension of θ is low. Consider integrals

$$A \cdot \mathbf{E}(\theta^j|\mathbf{y}) \equiv \int_\theta \cdot \theta^j \cdot L(\mathbf{y}|\theta)k(\theta)d\theta$$

$$\equiv \int_\theta \cdot \theta^j \cdot \left\{ \frac{L(y|\theta)k(\theta)}{h(\theta)} \right\} h(\theta)d\theta, \qquad (6.6)$$

for $j = 0, 1$, with the last formula obtained by multiplying and dividing by a positive function $h(\theta)$ selected to facilitate computation. For example, h may be a probability density with a CDF that is easy to compute. *Quadrature* is a standard approach to computing Equation (6.6): For a finite collection of points θ_i, $i = 1, \ldots, I$, and associated weights h_i, approximate Equation (6.6) by $\sum_{i=1}^I \theta_i^j \cdot \left\{ \frac{L(y|\theta_i)k(\theta_i)}{h(\theta_i)} \right\} h_i$. For example, the θ_i might enumerate the centroids of a finite grid that partitions θ-space, with h_i the probability that a random vector with density h falls in the partition with centroid θ_i. This is the computational analogue of Reimann integration, with accuracy limited by the degree of variation of the integrand $L(y|\theta)k(\theta)/h(\theta)$ on the partition; e.g., the accuracy with which the integrand can be approximated by step functions. Computation efficiency is improved by selecting a practical h such that the integrand varies slowly with θ. There are a variety of

mathematical methods that exploit smoothness features of the integrand to improve computational efficiency, e.g., Quarteroni *et al.* (2007, 9.1–9.7). A general idea behind quadrature is that various series of functions other than step functions, such as orthogonal polynomials or splines, have analytic integrals with respect to h, and points and weights can be selected so that the finite approximation to Equation (6.6) is exact for the initial terms in the specified series. For example, a widely used method termed *Gaussian quadrature* selects points and weights such that the approximation is exact for the family of Hermite orthogonal polynomials up to a given order. This can be computationally quite efficient, but two drawbacks are that the number of functional evaluations required rises to the power of the dimension of θ and quickly becomes intractable, and that the approximation is poor when the integrand varies rapidly relative to the approximating Hermite polynomials.

A simple simulation method for approximating Equation (6.6) is *importance sampling*. Start from a density $h(\theta)$ from which it is computationally easy to draw random deviates and for which $L(y|\theta)k(\theta)/h(\theta)$ is bounded for all θ. An average of the integrand in Equation (6.6) over a sample of simulated independent random deviates from h will by a law of large numbers converge to the integral.[3] The primary drawback of importance sampling is that unless the bracketed integrand is quite smooth and slowly varying in θ, the estimates will be relatively inefficient (as measured by precision achieved per unit of computational resources). However, variance reduction methods such as quota sampling from partitions of θ-space and use of antithetic variates that exploit symmetries in h can improve efficiency, e.g., Rubinstein *et al.* (1985). There are a variety of simulation methods that may improve substantially on the computational efficiency of importance sampling, some

---

[3]When $\theta^j \cdot \left\{ \frac{L(y|\theta)k(\theta)}{h(\theta)} \right\}$ is bounded by a constant C for $j = 0, 1$, Hoeffding's inequality (Pollard, 1984, Appendix B) establishes that $\text{Prob}\left(\left| \frac{1}{I} \sum_{i=1}^{I} \theta_i^j \cdot \left\{ \frac{L(\mathbf{y}|\theta_i)k(\theta_i)}{h(\theta_i)} \right\} - A \cdot \mathbf{E}(\theta^j|\mathbf{y}) \right| > \eta \right) < 2 \cdot e^{-I\eta^2/2C^2}$. Then with some manipulation, one can show that for $\eta < A/2$, $\text{Prob}\left(\left| \frac{\sum_{i=1}^{I} \theta_i \left\{ \frac{L(\mathbf{y}|\theta_i)k(\theta_i)}{h(\theta_i)} \right\}}{\sum_{i=1}^{I} \left\{ \frac{L(\mathbf{y}|\theta_i)k(\theta_i)}{h(\theta_i)} \right\}} - \mathbf{E}(\theta|\mathbf{y}) \right| > \eta \right) < 4 \cdot e^{-I\eta^2/32 \cdot \max(C^4, 1)}$, so convergence is exponential.

utilizing mathematical analogies to conservation of energy in physics, tempering and annealing in metallurgy, and minimization of entropy in probability theory. We explain several simple methods that are widely used in Bayesian analysis, and conclude with a survey of more advanced methods.

One simple alternative to the i.i.d. sampling from the density h used in importance sampling is termed *acceptance/rejection* (A/R) sampling. This method gives a sample from the posterior $K(\theta|\mathbf{y})$ without requiring estimation of the normalizing scale factor $A = \int_\theta L(y|\theta)k(\theta)d\theta$. Let $f(\theta|\mathbf{y}) \equiv L(\mathbf{y}|\theta)k(\theta) \propto K(\theta|\mathbf{y})$, and find a comparison density $h(\theta)$ from which it is easy to make pseudo-random draws, with the property $Bh(\theta) \geq f(\theta|\mathbf{y})$ for some positive scalar B.[4] Have the computer draw a sequence of pairs of pseudo-random numbers (θ_i, u_i), where θ_i is a draw from h and u_i is a draw from a uniform density on $[0,1]$. Accept and retain θ_i if $u_i Bh(\theta_i) \leq f(\theta_i|y)$; otherwise, reject and discard it. Then

$$\text{Prob}(\theta|\text{accept}) = \frac{\text{Prob}(\theta \,\&\, \text{accept})}{\int \text{Prob}(\theta' \,\&\, \text{accept})d\theta'} = \frac{h(\theta) \cdot f(\theta|\mathbf{y})/Bh(\theta)}{\int h(\theta') \cdot f(\theta'|\mathbf{y})/Bh(\theta')d\theta'}$$

$$= \frac{f(\theta|\mathbf{y})/B}{\int f(\theta'|\mathbf{y})/Bd\theta'} = K(\theta|\mathbf{y}), \tag{6.7}$$

and the sequence of accepted points θ_i will be a random sample from $K(\theta|\mathbf{y})$ as claimed. To illustrate, consider a choice experiment in which $n = 1, \ldots, N$ subjects are asked if they will purchase a product at price x_n, and d_n is an indicator that is one if the response is "Yes", zero otherwise; then $y_n \equiv (x_n, d_n)$. Suppose the probability model is logistic, $\text{Prob}(y_n|x_n) = \exp(-d_n x_n \theta)/(1+\exp(-x_n\theta))$, and the prior is standard exponential, e^θ for $\theta < 0$. Then $K(\theta|\mathbf{y}) \propto f(\theta|\mathbf{y}) \equiv \prod_{n=1}^{N} \frac{\exp(-y_n x_n \theta)}{1+\exp(-x_n\theta)}$ ·

---

[4]In the R core statistical package, vectors of draws from the beta, binomial, Cauchy, Chi-squared, Exponential, F, Gamma, Hypergeometric, Logistic, Log-normal, Multinomial, negative binomial, normal, Poisson, T, uniform, and Weibull distributions are obtained from the respective functions rbeta, rbinom, rcauchy, rchisq, rexp, rf, fgamma, rhyper, rlogis, rlnorm, rmultinom, rnbinom, rnorm, rpois, rt, runif, rweibull, with arguments specifying distribution parameters. These functions can also be used, with prefixes "d", "p", or "q" substituted for "r", to get densities, CDFs, or quantiles, respectively. Additional distributions are available in packages from CRAN.
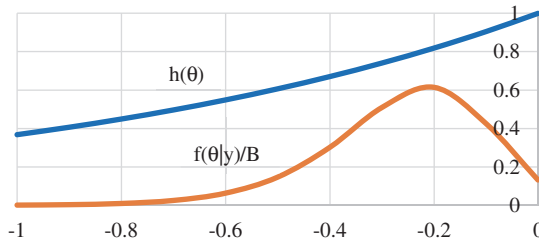
Figure 6.1: Illustration of acceptance/rejection method.

$e^\theta \mathbf{1}(\theta < 0)$, an expression whose integral does not have a closed form. Note that $f(\theta|\mathbf{y}) \le e^\theta \mathbf{1}(\theta < 0)$; take $h(\theta) \equiv e^\theta \mathbf{1}(\theta < 0)$ and $B = 1$.

Figure 6.1 illustrates how the acceptance/rejection method works in this case. In this figure, points are drawn with equal probability from the area below the blue curve that plots $Bh(\theta)$, extending indefinitely to the left, and are accepted if they lie below the red curve that plots $f(\theta|\mathbf{y})$, with the scale factor B set so that the red curve is never higher than the blue curve. The efficiency or yield of the procedure, the share of sampled points accepted, is given by the ratio of the area under the red curve to the area under the blue curve.

A limitation of the simple acceptance/rejection method is that it can have low efficiency unless the h curve and scale factor B can be tuned to $f(\theta|y)$; this is particularly true when the parameter vector is of high dimension. There are several modifications to A/R methods that may attain higher yields. First, in acceptance/rejection sampling, it may be possible to partition the support of $K(\theta|y)$ into a finite number of intervals or rectangles A_j, and define a value B_j that is a fairly tight upper bound on $f(\theta|y)/h(\theta)$ for $\theta \in A_j$. Define weights $w_j = \frac{h(A_j)/B_j}{\sum_{j'} h(A_{j'})/B_{j'}}$. Draw θ from h and u from uniform $[0, 1]$, determine the rectangle A_j in which θ lies, and accept θ with weight $1/w_j$ if $uB_jf(\theta|y) \le h(\theta)$. Then the weighted accepted sample has an empirical weighted CDF that converges to the CDF of $K(\theta|y)$, and the yield is high when the bounds within each rectangle are good. Second, the B_j can be adjusted adaptively. If a "burn-in" sample of θ's falls in rectangle j, then the maximum of $f(\theta|y)/h(\theta)$ in this sample is a consistent estimator of the best, or least upper bound, B_j. Using this

best estimate may under-accept θ from regions where the accepted subsample maximum is less than the true least upper bound, but adaptively adjusted B_j will converge very fast (e.g., at a $1/N$ rather than a $1/\sqrt{N}$ rate), so the bias will be negligible in simulations of reasonable size. Third, the partition into rectangles A_j can be refined recursively and adaptively to reduce the gap between the maximum and the minimum of $f(\theta|y)/h(\theta)$ in each rectangle. If these refinements are made as a rate that is asymptotically negligible relative to the accepted sample size, they will introduce only asymptotically negligible noise into the simulation. A variant of A/R sampling is termed *slice sampling.* It introduces an auxiliary variable τ with a uniform density on $[0, L(\mathbf{y}|\theta)k(\theta)]$ and a density $\pi(\theta|\tau) = \mathbf{1}(\tau \le L(\mathbf{y}|\theta)k(\theta))$. Starting from θ_{-1}, the sampler first draws τ from $[0, L(\mathbf{y}|\theta_{-1})k(\theta_{-1})]$, next places a random rectangle in θ-space around the point (τ, θ_{-1}), steps out this rectangle until it covers the set $\{\theta|\tau \le L(\mathbf{y}|\theta)k(\theta)\}$, and then samples θ from the resulting rectangle until $\tau \le L(\mathbf{y}|\theta)k(\theta)$ is satisfied. In Figure 6.1, this corresponds to drawing horizontal strips at random, and then sampling from the portion of these strips that lie under the curve. An interpretation of slice sampling is that it is a stochastic version of Lebesgue integration, while simple A/R sampling is a stochastic version of Reimann integration.

Additional efficiency may be achievable using *Monte Carlo Markov Chain* (MCMC) methods, see Tierney (1994), Albert and Chib (1996), and Chib and Greenberg (1996), and McFadden (1996). A *Metropolis–Hastings sampler* is a Markov process that recursively draws points from the parameter space Θ using a positive conditional density $q(\theta|\theta_{-1})$ from which it is convenient to sample. Given θ_{-1}, draw a trial θ^* from $q(\theta^*|\theta_{-1})$ and define $\gamma(\theta^*, \theta_{-1}|\mathbf{y}) \equiv \min\left(1, \frac{f(\theta^*|\mathbf{y}) \cdot q(\theta_{-1}|\theta^*)}{f(\theta_{-1}|\mathbf{y}) \cdot q(\theta^*|\theta_{-1})}\right)$. A key feature of the Metropolis–Hastings sampler is the symmetry of the function $\gamma(\theta^*, \theta_{-1}|\mathbf{y})$ that enters the acceptance criterion. This makes the sampler reversible in time, a defining characteristic of a successful sampler. This critical value reduces to $\min\left(1, \frac{f(\theta^*|\mathbf{y})}{f(\theta_{-1}|\mathbf{y})}\right)$ when $q(\theta|\theta_{-1})$ is symmetric, the original, more restrictive Metropolis sampler; e.g., specify $\theta - \theta_{-1}$ multivariate normal with mean zero. If a random scalar u from a uniform density on $[0, 1]$ satisfies $u < \gamma(\theta^*, \theta_{-1}|\mathbf{y})$ set $\theta = \theta^*$;

otherwise set $\theta = \theta_{-1}$. This defines a Markov transition probability

$$P(\theta|\theta_{-1}) = \begin{cases} q(\theta|\theta_{-1})\gamma(\theta,\theta_{-1}|\mathbf{y}) & \text{if } \theta \neq \theta_{-1} \\ 1 - R(\theta_{-1}) & \text{if } \theta = \theta_{-1} \end{cases} \tag{6.8}$$

where $R(\theta_{-1}) = \int_{\theta' \in \Theta} q(\theta'|\theta_{-1}) \cdot \gamma(\theta', \theta_{-1}|\mathbf{y}) d\theta'$ is the probability of acceptance. The posterior $K(\theta|y) = A \cdot f(\theta|y)$ for some constant A is an invariant of this Markov process. If θ_{-1} has density $A \cdot f(\theta_{-1}|y)$, the joint density of θ and θ_{-1} is $P(\theta|\theta_{-1}) \cdot A \cdot f(\theta_{-1}|y)$. Then, the marginal density of θ is

$$A \cdot \int_{\theta_{-1} \in \Theta} f(\theta_{-1}|\mathbf{y}) q(\theta|\theta_{-1}) \cdot \gamma(\theta, \theta_{-1}|\mathbf{y}) d\theta_{-1} + \mathbf{1}(\theta = \theta_{-1})$$

$$\cdot A \cdot f(\theta_{-1}|\mathbf{y}) \cdot (1 - R(\theta_{-1}))$$

$$= A \cdot \int_{\theta_{-1} \in \Theta} f(\theta|\mathbf{y}) q(\theta_{-1}|\theta) \cdot \min\left(1, \frac{f(\theta_{-1}|\mathbf{y}) \cdot q(\theta|\theta_{-1})}{f(\theta|\mathbf{y}) \cdot q(\theta_{-1}|\theta)}\right) d\theta_{-1}$$

$$+ \int_{\theta_{-1} \in \Theta} \mathbf{1}(\theta = \theta_{-1}) \cdot A \cdot f(\theta_{-1}|\mathbf{y}) \cdot (1 - R(\theta_{-1}))$$

$$= A \cdot f(\theta|\mathbf{y}) \cdot R(\theta) + A \cdot f(\theta|\mathbf{y}) \cdot (1 - R(\theta_{-1})) = A \cdot f(\theta|\mathbf{y}),$$

$$\tag{6.9}$$

and $K(\theta|y)$ is an invariant density. In Equation (6.14), $\mathbf{1}(\theta = \theta_{-1})$ denotes a unit probability on the set of points in $\Theta \times \Theta$ that meet the condition. The requirement that $q(\theta|\theta_{-1}) > 0$ on $\Theta \times \Theta$, plus technical conditions on the tails of $q(\theta|\theta_{-1})$ relative to $f(\theta|y)$ when Θ is not compact, are sufficient to assure that there is a unique invariant density and the Markov process converges to this invariant density.[5] A drawback of MCMC methods is that they can have high serial correlation, so that very large numbers of draws and substantial "burn-in" iterations are needed to get reliable empirical approximations to the posterior.

It is fairly common in applications to require a sample from a computationally intractable multivariate distribution whose one-dimensional

---

[5]The conditions that must be met on $q(\theta|\theta_{-1})$ and Θ are *irreducibility* (i.e., the process can always go from one point in Θ to any other point in a finite number of steps), *acyclicity* (i.e., the process does not always return to its starting point in a finite number of steps), and *positive recurrent* (i.e., the expected time until a return to a subset of positive measure is finite), see Tierney (1994), and McFadden (1996).

conditional distributions are relatively manageable. A specialization of the Metropolis–Hastings sampler called *Gibbs sampling* can be used in this case. Suppose θ is of dimension d, and $\theta = (\theta_1, \ldots, \theta_d)$ has a multivariate density $g(\theta_1, \ldots, \theta_d)$ with univariate conditionals $g_i(\theta_i|\theta_1, \ldots, \theta_{i-1}, \theta_{i+1}, \ldots, \theta_d)$ from which it is computationally tractable to sample using, say, the quantile transformations of these densities, acceptance/rejection sampling, or a MCMC procedure. From an initial vector $\theta' = (\theta'_1, \ldots, \theta'_d)$, the sampler loops through draws of θ''_i from $g_i(\theta_i|\theta''_1, \ldots, \theta''_{i-1}, \theta'_{i+1}, \ldots, \theta'_d)$ for $i = 1, \ldots, d$. The procedure then proceeds recursively, starting next from the vector θ''. The sampler can be started from any initial θ vector. A leading example is a truncated multivariate normal distribution g whose one-dimensional conditionals g_i are (truncated) univariate normal. To illustrate, in the R statistical package draws from g_i subject to a truncation $\theta_i \in [c_i, +\infty)$ are given by qnorm(runif(1,pnorm($c_i, \mu_i, \sigma_i, 1, 0$), 1)).

There has been considerable study of the conditions under which Metropolis–Hastings and Gibbs sampling methods perform poorly, see Wu and Fitzgerald (1996), Neal (2011), Hoffman and Gelman (2014), and Betancourt and Girolami (2013). When $K(\theta|\mathbf{y})$ has a strong gradient relative to the conditional density $q(\theta|\theta_{-1})$ in the Metropolis–Hastings sampler, or large differences between conditional and marginal variances in Gibbs sampling, sampler output can resemble local random walks that are highly serially correlated and fail to span the θ-space well in a reasonable number of iterations. The problem is particularly critical when $K(\theta|\mathbf{y})$ has several isolated sharp peaks. One straightforward method of dealing with this problem is called *annealing*, a term borrowed from metallurgy where a material is repeatedly heated above its crystallization temperature to allow atoms to migrate, and then slowly cooled. The analogy here replaces the target $f(\theta|\mathbf{y}) \equiv L(\mathbf{y}|\theta)k(\theta)$ by $f_\tau(\theta|\mathbf{y}) = f(\theta|\mathbf{y})^\tau$, which is flatter than $f(\theta|\mathbf{y})$ for $\tau < 1$, and then over the course of the iterations periodically allows τ to drop below one for a few iterations. Used in a Metropolis–Hastings sampler with θ's selected from parts of the chain where $\tau = 1$, this will lead a broader span of accepted values. However, the frequency, magnitude, and duration of $\tau < 1$ deviations, and "burn-in" times following these deviations, all have to be selected to tune the sampler to the application.

The most sophisticated versions of Monte Carlo Markov Chain methods currently in wide use are Hamiltonian Monte Carlo methods, particularly the No-U-Turn Sampler (NUTS) of Carpenter *et al.* (2017). These methods accomplish the same things as well-tuned annealing, and utilize a combination of mathematical theory and tinkering to automate some or all of the tuning process. A good introduction to these methods is Neal (2011). Hamilton Monte Carlo methods introduce auxiliary variables, in this case a vector ω that is of the same dimension as θ, two functions $V(\theta) = -\log(L(\mathbf{y}|\theta)k(\theta))$ and $W(\omega) = \omega' M^{-1}\omega/2$, where M is a positive definite matrix, and dynamic equations of motion in continuous time,

$$d\theta/dt = \partial W(\omega)/\partial \omega \equiv M^{-1}\omega \quad \text{and}$$

$$d\omega/dt = -\partial V(\theta)/\partial \theta \equiv \partial \log(L(y|\theta))/\partial \theta + \partial \log(k(\theta))/\partial \theta. \quad (6.10)$$

Interpret ω as a momentum vector, $W(\omega)$ as kinetic energy, and $V(\theta)$ as potential energy. This dynamic system is time reversible and preserves the Hamiltonian $H(\theta, \omega) \equiv V(\theta) + W(\omega)$; i.e., $dH(\theta, \omega)/dt \equiv 0$. This is analogous to conservation of energy in physical systems. The dynamic system moves rapidly out of regions where potential energy (i.e., posterior probability) is low, and slowly out of regions where potential energy is high. It has a strong mathematical property, called *symplecticness*, that the Jacobian of the transformation from (θ, ω) at time t to time $t + s$ is one, see Neal (2011, page 116). An implication of this property is that this dynamic system is probability preserving; i.e., it will move along contours of constant probability density. In computational approximation to this dynamic system, continuous time is replaced by discrete time moving in small steps Δ. Start from $\theta(t) = \theta_{-1}$ and a new independent draw $\omega(t) = M^{-1/2}\xi$, where ξ is a draw of a standard normal vector, independent of the previous state. Advance θ and ω for $s = 0, \ldots, S - 1$ steps using what is termed the *leapfrog method*:

$$\omega(t + (s + \tfrac{1}{2})\Delta) = \omega(t + s\Delta) - (\Delta/2) \cdot \partial V(\theta(t + s\Delta))/\partial \theta,$$

$$\theta(t + (s + 1)\Delta) = \theta(t + s\Delta) + \Delta \cdot M\omega(t + (s + \tfrac{1}{2})\Delta),$$

$$\omega(t + (s + 1)\Delta) = \omega(t + (s + \tfrac{1}{2})\Delta) - (\Delta/2)$$
$$\cdot \partial V(\theta(t + (s + 1)\Delta))/\partial \theta. \quad (6.11)$$

The half-steps have to be computed only for steps 1 and S, since in between the final half-step for s combines with the initial half-step for $s+1$. With this method, the discrete time approximation retains the time-reversibility and symplecticness properties of the continuous time process. At the end of the S steps, one has a proposed state $(\theta(t+S\Delta), \omega(t+S\Delta))$. This proposal is accepted as the next state if a uniform u drawn from $[0,1]$ satisfies

$$u < \min\{1, \exp(V(\theta(t)) + W(\omega(t)) - V(\theta(t+S\Delta))$$
$$- W(\omega(t+S\Delta))\}; \tag{6.12}$$

otherwise the process starts again from the previous value θ_{-1}.[6] The primary limitation of this Hamiltonian Monte Carlo method is that the step size Δ and path length S have to be specified through tuning of the sampler to the application; Neal (2011) gives an extensive discussion of methods and properties in his Section 5.4, and of alternatives to and variations on Hamiltonian Monte Carlo in his Section 5.5. The NUTS procedure of Carpenter *et al.* (2017) is an extension of Hamiltonian Monte Carlo that makes the specification of S automatic and adaptive. The key idea is to stop the leapfrog iteration (8.1) when the Euclidean distance between $\theta(t)$ and $\theta(t+s\Delta)$ stops increasing at step s, a "U-turn". To preserve time reversibility, NUTS uses an algorithm that runs forward or backward one step, then forward or backward two steps, and so on until a U-turn occurs. At that point, it samples from the points collected through the iteration in a way that preserves symmetry, and applies a criterion like Equation (6.12) to determine if the selected point is accepted. Carpenter *et al.* (2017) give a detailed discussion of this method, including proof of its time reversibility properties, recommendations and code for its implementation, a method for adaptively setting the step size Δ, and empirical experience with NUTS for Bayesian analysis of some standard test data. On the basis of the reported performance and our experience with this algorithm, and its availability in the open-software, R-compatible STAN statistical

---

[6]The symmetry of the Hamiltonian process requires that the sign of $\omega(t+S\Delta)$ be reversed if the proposed state is accepted. However, this step is not needed unless this process is interwoven with other Monte Carlo methods.

package developed by the authors and associates (Carpenter *et al.*, 2017), we believe it to be the best general-purpose method currently available for Bayesian analysis, particularly hierarchical Bayes analysis of CBC studies. However, it is out-performed by the Allenby–Train method described below in applications where normality assumptions allow use of efficient Gibbs sampling steps. The code for an R harness for organizing CBC data and accessing the STAN package can be downloaded from https://eml.berkeley.edu/~train/foundations_R.txt. This site also contains code for other data and procedures used in this monograph.

6.3 Hierarchical Bayes Methods

In marketing, a widely used model for CBC data is a mixed (or random coefficients) multinomial logit model like Equation (3.8) for the choice probabilities, with a Bayesian prior $k(\theta)$ on the deep parameters, see Rossi *et al.* (2005), Bąk and Bartłomowicz (2012), Hofstede *et al.* (2002), and Lenk *et al.* (1996). Let $L_n(d_n|x_n, p_n, \zeta)$ denote the likelihood, from Equation (5.2), that respondent n will state a vector of choices d_n from the offered menus. Then the posterior density on these parameters is

$$K(\theta|\mathbf{x}, \mathbf{p}) \propto \prod_{n=1}^{N} \int_{\zeta} L_n(d_n|x_n, p_n, \zeta) \cdot F(d\zeta|s_n, \theta) \cdot k(\theta). \qquad (6.13)$$

To illustrate the use of Equation (6.13) to produce Bayesian estimates of its parameters, consider a specification with a fairly conventional relatively uninformative prior k and a density $f(\zeta|s_n, \theta)$ that is multivariate normal except for a monotonic transformation (e.g., lognormal) or censoring in some components, as described by Train and Sonnier (2005) and summarized in the discussion around Equation (5.6). Let (σ, β) be a transformation of a commensurate latent vector ζ of normally distributed terms with mean μ and covariance matrix Ω, so that the parameters to be estimated are $\theta = (\mu, \Omega)$. Alternately, write $\Omega = \Lambda\Lambda'$, where Λ is a lower triangular matrix. In this case, $\theta = (\mu, \Lambda)$. For the illustration, assume that s_n does not shift the mean μ. If t is the dimension of (σ, β), then the dimension of θ is $t(t+3)/2$; this is also the dimension of ζ. The procedure will repeatedly use the following

operation for drawing (σ, β), referred to hereafter as the *taste parameter draw*: Given a trial value θ and a vector ω of i.i.d. standard normal deviates, let $\zeta = \mu + \Omega^{1/2}\omega \equiv \mu + \Lambda\omega$ be a $t \times 1$ multivariate normal vector. Each component of (σ, β) that is unrestricted in sign is set equal to the corresponding component of ζ; components like σ that are necessarily positive are set to the exponent of the corresponding element of ζ; and components that are censored to be non-negative are set to the maximum of 0 and the corresponding element of ζ. Other transformations are of course possible. We describe two widely used procedures that have been applied for this specification.

Allenby–Train Procedure: (Allenby, 1997, Train, 2009, pages 299–305):[7] Assume that the prior $k(\theta)$ is the product of (a) multivariate normal density for μ with zero mean and sufficiently large variance, so that the density is flat from a numerical perspective, and (b) an inverted Wishart for Ω with T degrees of freedom and parameter $T\mathbf{I}_T$, where $\mathbf{I}_T$ is a T-dimensional identity matrix. Draws from the posterior distribution are obtained by iterative Gibbs sampling from three conditional posteriors, with a Monte Carlo Markov Chain algorithm used for one of the conditional posteriors. Let μ_i, Ω_i, and ζ_{ni} be values of the parameters in iteration i, with the taste parameter draw implied by ζ_{ni}. Recall that N is the number of subjects. Gibbs sampling from the conditional posteriors for each parameter is implemented in three layers:

$\mu|\Omega, \zeta_n$ for all n: Given Ω_i and ζ_{ni} for all n, the conditional posterior on μ is $N\left(\bar{\zeta}_i, \frac{\Omega_i}{N}\right)$ where $\bar{\zeta}_i = \frac{1}{N}\sum_n \zeta_{ni}$. A new draw of μ is obtained as $\mu_{i+1} = \bar{\zeta}_i + \mathbf{\Psi}_i\omega$ where $\mathbf{\Psi}_i$ is the lower-triangular Cholesky factor of Ω_i/N and ω is a draw of standard normal deviates.

$\Omega|\mu, \zeta_n$ for all n: Given μ_{i+1} and v_{ni} for all n, the conditional posterior on Ω is Inverted Wishart with degrees of freedom $T + N$ and parameter $TI + N\bar{V}$, where $\bar{V} = \frac{1}{N}(\zeta_i - \mu_{i+1})(\zeta_i - \mu_{i+1})'$ Take

---

[7]Matlab, Gauss, and R codes to estimate mixed logits in preference-space by the Allenby–Train procedure are available on Train's website at https://eml.berkeley. edu/~train/software.html. The codes can be readily adapted for mixed logits in WTP-space. Sawtooth Software offers codes with this procedure for mixed logit in preference-space with normally distributed coefficients and the option of sign restrictions.

$T + N$ draws of T-dimensional vectors of i.i.d. standard normal deviates, labeled $\eta_r \; r = 1, \ldots, T+N$. A new draw of Ω is obtained as $\Omega_{i+1} = (\sum_r (\Gamma \eta_r)(\Gamma \eta_r)')^{-1}$ where Γ is the Cholesky factor of $(TI + N\bar{V})^{-1}$.

$\zeta_n | \mu, \Omega$: Given μ_{i+1} and Ω_{i+1}, for taste parameters σ_n, β_n, δ_n implied by ζ_n, the conditional posterior of $\zeta_{n,i+1}$ is proportional to $P^n(\zeta_{n,i+1})\varphi(\zeta_{n,i+1}|\mu_{i+1}, \Omega_{i+1})$, where φ denotes the multivariate normal density with mean μ_{i+1} and covariance matrix Ω_{i+1}, and P^n is defined following Equation (5.2). A draw of $\zeta_{n,i+1}$ is obtained by the following Metropolis–Hastings procedure: Draw ω_n from a jumping distribution, which we specify as normal with mean zero and variance $m\Omega$, where scalar m is chosen as described below. Calculate a trial value of ζ_n as $\tilde{\zeta}_n = \zeta_{ni} + \omega_n$. Using the taste parameters $(\tilde{\sigma}_n, \tilde{\beta}_n)$ implied by $\tilde{\zeta}_n$, calculate the probability $P^n(\tilde{\zeta}_n)$. If a $[0, 1]$ uniform random variate u satisfies

$$u < \frac{P^n(\tilde{\zeta}_n)\varphi(\tilde{\zeta}_n|\mu_{i+1}, \Omega_{i+1})}{P^n(\tilde{\zeta}_{ni})\varphi(\zeta_{ni}|\mu_{i+1}, \Omega_{i+1})}, \qquad (6.14)$$

then the new value of $\zeta_{n,i+1}$ is this trial $\tilde{\zeta}_n$; otherwise, discard the trial $\tilde{\zeta}_n$ and retain the old value as the new value, $\zeta_{n,i+1} = \zeta_{ni}$ The value of m is adjusted over iterations to maintain an acceptance rate near 0.30 over n and iterations.

For a model with intra-consumer as well as inter-consumer heterogeneity, the only change in the HB procedure described is to introduce an additional layer of multivariate normal intermediate parameters ρ_{mn} that vary by menu as well as subject and have a density centered at ζ_n, with a covariance matrix adjusted in an additional Gibbs layer.

NUTS procedure: This procedure uses the NUTS sampler developed by Carpenter *et al.* (2017) and implemented in the R-STAN statistical package. This is open-source software, with both R and STAN available in versions that run on Linux, Windows, and Mac platforms. The key acceptance stage of the sampler utilizes a Metropolis–Hastings algorithm, and the sampler incorporates an adaptive adjustment of Hamiltonian discrete time step size, and the interactive determination

of number of steps to optimize acceptance rates. The sampler is flexible in specification of priors and likelihoods, and is not limited to the multivariate normal model described above. A drawback of this flexibility is that STAN may run more slowly for this model than procedures that use computationally efficient Gibbs samplers for part of the multivariate normal iteration. Suppose an intermediate parameter vector ζ that is multivariate normal with mean μ and a covariance matrix $\Omega = D\Gamma\Gamma'D$, where D is diagonal and Γ is the lower triangular Cholesky factor of the correlation matrix (so that Γ is characteristically a lower triangular matrix with rows whose sums of squared elements are one). Recommended priors are a relatively diffuse multivariate normal prior for μ, relatively diffuse exponential densities for the elements of D, and a LKJ(1) density for the correlation matrix $\Gamma\Gamma'$, see Lewandowski *et al.* (2009).[8]

Terminal Steps: Consider sampler data μ_i, Ω_i, and v_{ni} for subjects $n = 1, \ldots, N$ and iterations $i = 1, \ldots, I$ after discarding a few thousand "burning in" iterations; I is typically at least 10,000, and perhaps as large as 1 million. The Bayes estimates of the deep parameters are the posterior means $\hat{\mu}$ and $\hat{\Omega}$ formed by averaging μ_i and Ω_i respectively over $i = 1, \ldots, I$. A first question in applications of sampling schemes like the one described above is whether one has achieved convergence to a stationary distribution that spans θ-space, with sufficiently strong mixing so that serial correlations between distant iterations are small and sampler means are accurate estimates of true posterior means. The recommendations of the developers of these samplers are to start them from alternative parameter values, vary burn-in times and tuning parameters, and jitter tuning parameters to avoid accidental cyclicity. In the special case of test data with known deep parameters, one can examine bias and mean square error as a function of sampler

---

[8]All correlation matrices can be written as $\Gamma\Gamma'$, where Γ is a lower triangular matrix with the property that the squares of the elements in each row sum to one. Then the elements $(\Gamma_{i1}, \ldots, \Gamma_{ii})$ are points in a unit sphere of dimension i, and the method of Muller (1959) and Marsaglia (1972) can be used to draw points uniformly from this sphere, i.e., draw i i.i.d. standard normal deviates, and scale the vector of these deviates by dividing by the square root of their sum of squares.

characteristics.[9] Econometric time series diagnostic tests for stationarity and for mixing conditions can be applied more generally to the sampler output, see Azizzadeh and Rezakhah (2014), Delgado (2005), Domowitz and El-Gamal (1993), Corradi *et al.* (2000), and Ljung and Box (1978).

From a classical perspective, the posterior means $\hat{\mu}$ and $\hat{\Omega}$ are point estimates of true parameter values μ_0 and Ω_0. Suppose the analyst wishes to treat these as classical estimators and carry out conventional steps such as associating standard errors with the estimates and providing confidence bounds or hypothesis tests. Due to the asymptotic equivalence of MLE and hierarchical Bayes estimators, this can be done using the formulas (5.2)–(5.7). To improve the asymptotic approximation, we recommend executing one or more BHHH steps starting from the HB posterior means, using the result from Equation (5.4) to obtain MLE-equivalent estimates, and using Equation (5.5) to estimate the covariance matrix of the resulting estimators. In most cases, it is practical to draw new intermediate parameter vectors ζ_r for Equation (5.7), but it is also possible to reuse the array of vectors ζ_{ni} from iterations in the hierarchical Bayes process in the calculation of Equations (5.2)–(5.6).

6.4 A Monte Carlo example

To demonstrate the procedures given in the previous two sections, consider a Monte Carlo example where a known decision-making process is used to generate an artificial CBC data set, and the methods above are then applied to estimate the parameters behind these "observed" choices. For this exercise, suppose we have CBC data on preferences for table grapes, the example discussed earlier in Section 2.6. Suppose 1,000 subjects are each presented with eight menus, and on each menu they have the option of choosing one of three bunches of grapes with

---

[9]In test data where the ζ for each subject n are drawn from a known normal distribution $N(\mu, \Omega)$, and choice data are then generated from a random utility model with these taste parameters, the mean and covariance matrix used for these comparisons should be $\mu_0 = \frac{1}{N} \sum_{n=1}^{N} v_n$ and $\Omega_0 = \frac{1}{N} \sum_{n=1}^{N} (v_n - \mu_0)(v_n - \mu_0)'$ rather than the underlying μ and Ω; as the target of the analysis is recovery of characteristics of realized tastes, and the analysis cannot identify the process that generated the realized tastes.

Table 6.1: Table grape CBC attributes and levels.

Attribute	Symbol	Levels
Price	P	$1.00 to $4.00
Sweetness	S	Sweet (1) or Tart (0)
Crispness	C	Crisp (1) or Soft (0)
Size	L	Large (1) or Small (0)
Organic	O	Organic (1) or Non-organic (0)

various attributes and prices, or choosing to buy no grapes on this occasion. Assume that the subjects are prompted to think of this as the opportunities they will face on their next trip to the supermarket, so that existing stocks of grapes and purchasing opportunities outside the experiment are similar to those faced by these subjects in the real market. The product attributes and levels in the experiment are given in Table 6.1.

The gender (G) or each subject is observed, coded $G = 1$ for female and $G = 0$ for male. Let I_n denote disposable income, which is not measured, but which drops out of the utility maximization. Define the interactions $SC \equiv S^*C$ and $SG \equiv S^*G$. Index a menu's alternatives by $j = 1, 2, 3$ for the three offered bunches, and $j = 4$ for the no purchase option. Let B_{jmn} be a dummy variable that is one for purchase alternatives ($j = 1, 2, 3$) and zero for the no-purchase alternative $j = 4$. Assume the true utility has the WTP form

$$U_{jmn} \equiv I_n - P_{jmn} + S_{jmn}\beta_{Sn} + C_{jmn}\beta_{Cn} + L_{jmn}\beta_{Ln} + O_{jmn}\beta_{On}$$
$$+ SC_{jmn}\beta_{SCn} + SG_{jmn}\beta_{SGn} + B_{jmn}\beta_{qn} + \sigma_n\varepsilon_{jn}, \quad (6.15)$$

where the ε_{jn} are i.i.d. standard EV1 distributed, and $\zeta_n = (\beta_{Sn}, \beta_{Cn}, \beta_{Ln}, \beta_{On}, \beta_{SCn}, \beta_{SGn}, \beta_{Bn}, \log(\sigma_n))$ are the preference parameters. Define $D_{jmn} = 1$ for the alternative within a menu that maximizes utility, $D_{jmn} = 0$ otherwise. Assume that the true parameters are distributed multivariate normal in the population with the means, standard deviations, and correlation/covariance matrix given in Table 6.2; R code that generates the data is available at https://eml.berkeley.edu/~train/foundations_R.txt.

Table 6.2: True parameter distribution β_S, β_C, β_L, β_O, β_{SC}, β_{SG}, β_{qn}, $\log(\alpha_n)$.

	Mean	Std. Dev.	Covariance (bottom) and correlation (top) matrices							
			β_S	β_C	β_L	β_O	β_{SC}	β_{SG}	δ_n	$\log(\sigma_n)$
β_S	1.0	0.3	0.09	0.6	0	0	0.3	0	0	0
β_C	0.3	0.1	0.018	0.01	0.48	0	0.42	0	0	0
β_L	0.2	0.1	0	0.0048	0.01	0	0.42	0	0	0
β_O	0.1	0.2	0	0	0	0.04	0.3	0	0	0
β_{SC}	0.0	0.05	0.0045	0.0021	0.0021	0.003	0.0025	0	0	0
β_{SG}	0.1	0.2	0	0	0	0	0	0.04	0	0
β_{Bn}	2.0	1.0	0	0	0	0	0	0	1	0
$\log(\sigma_n)$	−0.5	0.3	0	0	0	0	0	0	0	0.09

Model (6.15) is estimated with independence of $\log(\sigma_n)$ and the remaining parameters imposed, but all other covariances allowed. Table 6.3 give estimates by hierarchical Bayes using the Allenby–Train procedure (HB-AT), hierarchical Bayes using the NUTS procedure (HB-NUTS), and maximum simulated likelihood (MSL). Note that the true population parameter values given in Table 6.1 are reproduced closely but not exactly by their sample counterparts for the 1,000 subjects in the CBC study; these are given in the third column of the table, and are the relevant values for comparison with estimates. The maximum simulated likelihood estimates use 250 Halton draws to approximate the terms in Equations (5.5) and (5.6), and use Newton–Raphson iteration to the estimator. Starting values were obtained by estimating a model with fixed coefficients, using the estimated WTPs as starting values in a model with random but uncorrelated WTPs, and then using these estimates as starting values in the model with correlated WTPs. The three steps combined took about three hours of run time. The HB-AT estimates are obtained with 100,000 iterations used to tune and burn-in the sampler, and every tenth draw taken from 100,000 subsequent draws to form the posterior means given in the table.[10] It has been observed that this procedure tends to inflate variances that are numerically very small (see, e.g., Balcombe and Fraser, 2009); to counteract this tendency, we divided the nonprice variables by ten and then rescaled the estimates accordingly. Estimation took 12 minutes. The HB-NUTS estimates are obtained from 11,000 iterations, with the first 5,500 discarded for "burn-in", and every 10th draw from the remainder used to obtain the estimates in Table 6.3.

For the NUTS sampler, we also calculated posterior means (not reported) using all iterations including burn-in iterations, using every iteration rather than every 10th iteration, and using every 100th iteration including burn-in iterations. The results were very close, indicating that in this application NUTS seems to show rapid convergence with relatively little drift. However, time-series graphs do show some long

---

[10]We set T, the degrees of freedom for the inverted Wishart prior, equal to the number of random coefficients in the model. We have found little numerical effect of using a larger degrees of freedom.

Table 6.3: Estimates from table grapes CBC data.

Parameter	True values		HB-AT		HB-NUTS		MSL	
	Population	Sample	Mean	Std. Dev.	Mean	Std. Dev.	Estimate	Std. Error
Means								
β_S	1.0	1.005	0.9682	(0.0445)	0.9784	(0.0407)	0.9694	(0.0457)
β_C	0.3	0.299	0.3495	(0.0358)	0.3610	(0.0342)	0.3554	(0.0391)
β_L	0.2	0.197	0.1781	(0.0251)	0.1898	(0.0258)	0.1760	(0.0238)
β_O	0.1	0.109	0.0910	(0.0238)	0.0964	(0.0224)	0.09988	(0.0248)
β_{SC}	0.0	0.001	-0.0450	(0.0485)	-0.0515	(0.0464)	-0.0579	(0.0490)
β_{SG}	0.1	0.093	0.1369	(0.0540)	0.1446	(0.0469)	0.1523	(0.0536)
β_{Bn}	2.0	2.012	2.0049	(0.0498)	2.0159	(0.0583)	1.9893	(0.512)
$\log(\sigma_n)$	-0.5	-0.481	-0.5277	(0.0253)	-0.5038	(0.0228)	-0.5687	(0.281)

Panel A

(Continued)

Hierarchical Bayes Estimation

Table 6.3: (*Continued*)

Parameter	True values		HB-AT		HB-NUTS		MSL	
	Population	Sample	Mean	Std. Dev.	Mean	Std. Dev.	Estimate	Std. Error
Std. Devs.								
β_S	0.3	0.303	0.3086	(0.0633)	0.3299	(0.0610)	0.3044	(0.8889)
β_C	0.1	0.099	0.2149	(0.0354)	0.1299	(0.0539)	0.2661	(0.1085)
β_L	0.1	0.101	0.1621	(0.0370)	0.1325	(0.0497)	0.1496	(0.0517)
β_O	0.2	0.210	0.2157	(0.0378)	0.1863	(0.0489)	0.2440	(0.0467)
β_{SC}	0.05	0.051	0.2227	(0.0516)	0.0974	(0.0738)	0.2564	(0.1511)
β_{SG}	0.2	0.202	0.2518	(0.0758)	0.2960	(0.1325)	0.3955	(0.1084)
B_{Bn}	1.0	0.988	0.7924	(0.0764)	0.9288	(0.0548)	0.8917	(0.0606)
$\log(\sigma_n)$	0.3	0.296	0.4220	(0.0257)	0.2757	(0.0415)	0.3406	(0.0415)

<div align="center">(Continued)</div>

Table 6.3: (*Continued*)

Panel C

Parameter	True values		HB-AT		HB-NUTS		MSL	
	Population	Sample	Mean	Std. Dev	Mean	Std. Dev	Estimate	Std. Error
Covariances								
β_S,β_C	0.018	0.0185	0.0348	0.0183	0.0044	(0.0119)	0.0547	(0.0375)
β_S,β_L	0	0.0007	-0.0054	0.0166	-0.0020	(0.0125)	-0.0038	(0.0169)
β_S,β_O	0	-0.0001	0.0077	0.0169	0.0012	(0.0148)	0.0003	(0.0176)
β_S,β_{SC}	0.0045	0.0046	-0.0244	0.0237	0.0000	(0.0151)	-0.02867	(0.0449)
β_S,β_{SG}	0	-0.0004	0.0002	0.0230	-0.0150	(0.0439)	-0.0152	(0.0411)
β_S,β_B	0	-0.0032	0.1430	0.0566	0.0289	(0.0432)	0.1028	(0.0453)
β_C,β_L	0.0048	0.0050	0.0067	0.0095	0.0026	(0.0052)	0.0165	(0.0190)
β_C,β_O	0	0.0006	0.0161	0.0119	0.0012	0.0079	0.0108	(0.0193)
β_C,β_{SC}	0.0021	0.0022	-0.0291	0.0154	-0.0030	(0.0061)	-0.0579	(0.0639)
β_C,β_{SG}	0	0.0000	-0.0129	0.0177	0.0099	(0.0172)	-0.0038	(0.0449)
β_C,β_B	0	-0.0027	0.1073	0.0402	0.0112	(0.0378)	0.0439	(0.0533)
β_L,β_O	0	0.0009	0.0145	0.0114	0.0121	(0.0093)	0.0315	(0.143)
β_L,β_{SC}	0.0021	0.0023	-0.0052	0.0116	0.0002	(0.0052)	-0.0083	(0.0210)
β_L,β_{SG}	0	0.0001	-0.0105	0.0148	-0.0043	(0.0151)	-0.0001	(0.0217)
β_L,β_B	0	-0.0040	-0.0013	0.0332	-0.0094	(0.0253)	-0.0169	(0.0236)
β_O,β_{SC}	0.0030	0.0032	-0.0142	0.0131	0.0006	0.0084	0.0050	(0.0240)
β_O,β_{SG}	0	-0.0016	-0.0153	0.0202	-0.0076	(0.0182)	-0.0213	(0.0263)
β_O,β_B	0	-0.0002	0.0492	0.0331	0.0396	(0.0268)	0.0076	(0.0326)
β_{SC},β_{SG}	0	-0.0002	0.0212	0.0194	0.0027	(0.0140)	0.0160	(0.0471)
β_{SC},β_B	0	-0.0020	-0.1083	0.0513	-0.0045	(0.0290)	-0.0420	(0.0668)
β_{SG},β_B	0	0.0008	-0.0417	0.0615	-0.0145	(0.0508)	-0.0869	(0.0747)

cycles in the parameter estimates. The NUTS procedure required computation overnight; this may be due to storage of posterior intermediate parameters for each subject in each iteration, output that may be useful for some marketing applications such as segmentation, but is not needed for the primary task of estimating the deep parameters. Comparing the HB-AT, HB-NUTS, and MSL estimates with the true parameter values and with each other, we find that all the methods do a good job of recovering the true utility model.

7

An Empirical CBC Study Using MSL and HB Methods

In this section, we provide an example of a money-metric random utility model estimated using both Maximum Simulated Likelihood and Hierarchical Bayes methods. We utilize the data from a conjoint experiment designed and described by Glasgow and Butler (2017) for consumers' choice among video streaming services. Their experiments included the monthly price of the service and the following non-price attributes:

Each choice experiment included four alternative video steaming services with specified price and attributes plus a fifth alternative of not subscribing to any video streaming service. Each respondent was presented with 11 choice experiments. Glasgow and Butler obtained choice from 300 respondents and implemented an estimation procedure that accounts for "protestors" (mainly consumers who never chose a service that shared data). In our use of their data, we do not include the 40 respondents that they identified as protestors, so that our sample consists of 260 respondents.

Table 7.2 gives a model estimated in WTP-space by hierarchical Bayes, using the Allenby–Train procedure with 10,000 burn-in iterations, 10,000 iterations after burn-in from which, to reduce serial correlation,

91

every 10th draw from the sampler was retained to calculate the estimates. WTP for each non-price attribute is assumed to be normally distributed over consumers; the price/scaling coefficient $(1/\alpha)$ is assumed to be log-normally distributed; and correlation is allowed among the WTPs as well as the price/scaling coefficient. The point estimate of the population mean in the mean of the draws from the posterior distribution, and the standard error of this estimate is the standard deviation of the draws. Similarly for the standard deviation in the population, the estimate is the mean of the draws from its posterior and the standard error is the standard deviation of these draws. The point estimates of the correlations are shown in the second part of the table. The results in Table 7.1 indicate that people are willing to pay \$1.57 per month on average to avoid commercials. Fast availability is valued highly, with an

Table 7.1: Non-price attributes.

Attribute	Levels
Commercials shown between content	Yes ("commercials") No (baseline category)
Speed of content availability	TV episodes next day, movies in 3 months ("fast content") TV episodes in 3 months, movies in 6 months (baseline category)
Catalogue	10,000 movies and 5,000 TV episodes ("more content") 2,000 movies and 13,000 TV episodes ("more TV/fewer movies") 5,000 movies and 2,500 TV episodes (the baseline category)
Data-sharing policies	Information is collected but not shared (baseline category) Usage information is shared with third parties ("share usage")[a] Usage and personal information are shared with third parties ("share usage and personal")

Note: [a]Glasgow and Butler use the terms "non-personally identifiable information (NPPI)" and "personally identifiable information (PII)" for what we are labelling "share usage" and "share usage and personal".

average WTP of \$3.50 per month in order to see TV shows and movies soon after their original showing. On average, people prefer having a mix with more TV shows and fewer movies, but the mean is not significant. Average willingness to pay for more content of both kinds is \$2.57 per month. Interestingly, people who want fast availability tend to be those who prefer more TV shows and fewer movies: the correlation between these two WTPs is 0.51, while the correlation between WTP for fast availability and more content of both kinds is only 0.04. Apparently, the desire for fast availability mainly applies to TV shows.

Consider now the data-sharing policies of services. The point estimate implies that consumers have a WTP of 43 cents per month to avoid having their usage data shared in aggregate form; however, the hypothesis of zero average WTP cannot be rejected. Consumers are much more concerned about their personal information being shared along with their usage information: the average WTP to avoid such sharing is \$1.83 per month. The correlation between WTP to avoid the two forms of sharing is a substantial 0.82.

Table 7.3 gives the same model estimated by the method of maximum simulated likelihood (MSL). We used STATA's module for mixed logit in wtp-space, "mixlogitwtp." We specified 100 Halton draws per person, which has been shown to be more accurate for mixed logit estimation than 1,000 pseudo-random draws per person (Bhat, 2001; Train, 2000; Train, 2009). We used the HB estimates (Table 6.3) as starting values. At the HB estimates, the log-likelihood was -3994.64 and rose to -3903.47 at convergence.[1]

The MSL estimates for the mean WTPs are fairly similar to the HB estimates. In particular, the MSL estimates of mean WTP to: avoid commercials is \$1.56 compared to \$1.57 by HB; obtain fast availability is \$3.95 compared to \$3.50 by HB; obtain more content is \$2.96 compared to \$2.57 by HB; to avoid having no service is \$27.26 compared to \$28.59

---

[1]The mixlogitwtp module allows estimation by Newton–Raphson (NR), BHHH, DFP, and BFGS. Using NR, one iteration raised the log-likelihood from -3994.65 to -3969.31. The additional rise to convergence at -3903.47 took 29 iterations. Run time was about 90 minutes. The NR option utilizes numerical gradients and Hessian. The other procedures utilize numerical gradients but alternative forms of the Hessian and can be expected to run faster.

Table 7.2

Panel A: Hierarchical Bayes model of WTPs for video streaming service

	Population Mean		Std. Dev. in Population	
	Estimate	Std. Error	Estimate	Std. Error
Log($1/\sigma$)	−1.890	0.1188	1.401	0.2497
WTP for:				
Commercials	−1.571	0.5547	3.4882	0.4206
Fast availability	3.496	0.7182	3.6517	0.4189
More TV/fewer movies	0.227	0.5237	3.6706	0.3602
More content	2.569	0.5170	2.6593	0.3532
Share usage only	−0.432	0.3015	1.9504	0.3170
Share personal and usage	−1.831	0.6325	4.7189	0.6012
No service	−28.591	3.2753	20.752	2.9424

Panel B: Correlations: Point estimates

	Log($1/\sigma$)	Commercials	Fast availability	More TV, fewer movies	More content	Share usage	Share personal and usage	No service
Log($1/\sigma$)	1.0000	−0.3010	−0.1484	−0.1414	0.0072	−0.0050	−0.1431	0.3865
Commercials		1.0000	0.2238	0.1493	−0.1912	−0.0799	−0.0204	−0.3639
Fast availability			1.0000	0.5094	0.0418	−0.5126	−0.5135	0.2197
More TV, fewer movies				1.0000	−0.4523	−0.1151	−0.2787	0.3266
More content					1.0000	−0.0057	0.1650	0.1884
Share usage						1.0000	0.8155	−0.2866
Share personal and usage							1.0000	−0.4535
No service								1.0000

Table 7.3

Panel A: MSL Estimates of WTPs for Video Streaming Services

	Population Mean		Std. Dev. in Population	
	Estimate	Std. Error	Estimate	Std. Error
$\log(1/\sigma)$	−2.002	0.0.945	1.0637	0.0755
WTP for:				
Commercials	−1.562	0.4214	3.940	0.5302
Fast Availability	3.945	0.4767	3.631	0.4138
More TV, fewer movies	−0.6988	0.4783	4.857	0.5541
More content	2.963	0.4708	2.524	0.4434
Share usage only	−0.6224	0.4040	2.494	0.4164
Share personal and usage	−2.705	0.5844	6.751	0.7166
No service	−27.26	2.662	19.42	2.333

Panel B: Correlations: Point Estimates (*denotes significance at 5% level)

	Commercials	Fast Availability	Mostly TV	Mostly movies	Share usage	Share personal and usage	No service
$\log(1/\sigma)$	−0.5813*	−0.1371	0.0358	0.0256	0.0022	−0.1287	0.2801*
Commercials	1.0000	0.1172	−0.3473*	0.0109	−0.2562	−0.0079	−0.4108*
Fast Availability		1.0000	0.8042*	−0.4019*	−0.3542*	−0.4206*	0.2391*
Mostly TV			1.0000	−0.5890*	−0.1695	−0.3328*	0.4616*
Mostly movies				1.0000	0.5141*	0.5181*	−0.0147
Share usage					1.0000	0.9370*	−0.0563
Share personal and usage						1.0000	−0.0975
No service							1.0000

by HB. The mean WTPs for more TV shows/fewer movies and to avoid sharing of usage data are insignificant under both MSL and HB. The only substantial difference is the mean WTP to avoid sharing of personal and usage information, which is estimated to be $2.71 by MSL and $1.83 by HB.

The estimated standard deviations of WTP are practically the same by MSL and HB for the following attributes: to avoid commercials, obtain fast service, obtain more content, and avoid not having service. Standard deviations of WTP for more TV/fewer movies and to avoid either form of data sharing are estimated to be higher by MSL than HB. The correlations that were discussed above in relation to the HB estimates are also evidenced with MSL. However, other correlations differ in magnitude and even sign between MSL and HB.

A preference-space model without the money-metric scaling was also estimated with the price coefficient assumed log-normally distributed and the non-price coefficients assumed normal (not shown.)[2] In general, the estimated means and standard deviations of WTP were considerably higher and less plausible. For example, for sharing of usage and personal data, the model in WTP-space above gives a mean WTP to avoid sharing of $2.71 with a standard deviation of $6.75. For the model in preference space, the derived distribution of WTP has a mean $6.99 and standard deviation of $98.95. The high values, especially for the standard deviation, are due to the fact that WTP in the preference space model is calculated as the ratio of the attribute's coefficient divided by the price coefficient, such that values of the price coefficient that are arbitrarily close to zero (which are allowed by the distribution of the price coefficient) imply arbitrarily large WTP. Models in WTP-space avoid this issue by specifying and estimating the distribution of WTP directly. However, the log-likelihood of the model in preference space is higher than that in wtp-space: -3863.9 compared to -3903.5. These results, namely, more plausible estimates of the distribution of WTP accompanied by a lower log-likelihood, mirror those of Train and Weeks (2005). Other studies making this comparison (e.g., Scarpa *et al.*, 2008)

---

[2]As noted in section 3.1, a model in preference-space with the assumption of a log normal price coefficient and normal non-price coefficients is not equivalent to a WTP-space model with normally distributed WTP for non-price attributes.

have found no such tradeoff on their datasets, with their model in WTP-space obtaining a better fit as well as more plausible distributions of WTP.

As discussed above, the means and standard deviations of the WTP estimates obtained by MSL and HB-AT methods generally agree on sign and magnitude, but are not close enough to conclude that these methods have converged to asymptotically equivalent estimates. This may reflect finite sampling deviations from the asymptotic limit; there is no guarantee that the hierarchical Bayes and MSL estimates will be close in finite samples. However, both methods use simulations, and the hierarchical Bayes methods in particular are known to converge slowly to their posterior limiting distribution when their tuning is imperfect. Then, some of the differences in Tables 7.2 and 7.3 might disappear with more complete tuning and more simulation draws.

8

An Application with Inter and Intra-consumer Heterogeneity

This section extends the Hierarchical Bayes method to models with inter- and intra-consumer heterogeneity. These models account for intra- and inter-consumer heterogeneity with three levels of parameters:

1. Population-level parameters μ and Ω^b: average tastes/ preferences in the population and the inter-consumer variance–covariance matrix, respectively.

2. Individual-level parameters ζ_n and Ω^w: average tastes/preferences of a specific individual and the intra-consumer variance–covariance matrix, respectively.

3. Menu-level parameters ρ_{mn}: to reflect menu-specific (choice specific) taste perturbations.

Such models have been previously estimated by Hess and Train (2011), and Hess and Rose (2009) using a maximum simulated likelihood estimator. Below we describe an extension of the Allenby–Train estimator that is based on the same Normality assumptions. For more details, see

Becker *et al.* (2018). The extended HB procedure includes the following five Gibbs sampling layers.

Five-Step Gibbs Sampling

$\mu|\Omega^b$, ζ_n, Ω^w, ρ_{mn}: Given Ω_i^b and ζ_{ni} for all n, the conditional posterior on μ is $N\left(\bar{\zeta}_i, \frac{\Omega_i^b}{N}\right)$ where $\bar{\zeta}_i = \frac{1}{N}\sum_n \zeta_{ni}$. A new draw of μ is obtained as $\mu_{i+1} = \bar{\zeta}_i + \Psi_i\omega$ where Ψ_i is the Cholesky factor of Ω_i^b/N and ω is a draw of the multivariate standard normal.

$\Omega^b|\mu$, ζ_n, Ω^w, ρ_{mn}: Given μ_{i+1} and ζ_{ni} for all n, the conditional posterior on Ω^b is Inverted Wishart with degrees of freedom T and parameter $TI + N\bar{V}^b$, where T is the number of unknown parameters and $\bar{V}^b = \frac{1}{N}(\zeta_{ni} - \mu_{i+1})(\zeta_{ni} - \mu_{i+1})'$. Take $T + N$ draws of T-dimensional vectors of i.i.d. standard normal deviates, labeled v_r, $r = 1, \ldots, T + N$. A new draw of Ω^b is obtained as $\Omega_{i+1}^b = (\sum_r (\Gamma^b v_r)(\Gamma^b v_r)')^{-1}$ where Γ^b is the Cholesky factor of $(TI + N\bar{V}^b)^{-1}$.

$\Omega^w|\mu$, ζ_n, Ω^b, ρ_{mn}: Given $\rho_{mn,i}$ and ζ_{ni} for all n, the conditional posterior on Ω^w is Inverted Wishart with degrees of freedom $T+M_T$ and parameter $TI + M_T\bar{V}^w$, where M_T represents the total number of menus faced by all individuals, and $\bar{V}^w = \frac{1}{M_T}\sum_{m=1}^M (\rho_{mn,i} - \zeta_i)(\rho_{mn,i} - \zeta_i)'$. Take $T + M_T$ draws of T-dimensional vectors of i.i.d. standard normal deviates, labeled v_s, $s = 1, \ldots, T + M_T$. A new draw of Ω^w is obtained as $\Omega_{i+1}^w = (\sum_s (\Gamma^w v_s)(\Gamma v_s)')^{-1}$ where Γ^w is the Cholesky factor of $(TI + M_T\bar{V}^w)^{-1}$. Here we assume a single variance–covariance matrix for all individuals; due to the small number of choice situations faced by each individual in a typical stated preference survey, it is not possible to estimate a variance–covariance matrix for each individual.

$\zeta_n|\Omega^b$, μ, Ω^w, ρ_{mn}: Given μ_i, ρ_{mni} and Ω_i^b for all n, the conditional posterior on ζ_n with $N(\mu_i, \Omega_i^b)$ as a prior is $N(\tilde{\zeta}_n, \Sigma_{\zeta n})$ where $\tilde{\zeta}_n = ([\Omega_{i+1}^b]^{-1} + M[\Omega_{i+1}^w]^{-1})^{-1}([\Omega_{i+1}^b]^{-1}\mu_i + M[\Omega_{i+1}^w]^{-1}\frac{1}{M}\sum_{m=1}^M \rho_{mni})$, and $\Sigma_{\zeta n} = ([\Omega_{i+1}^b]^{-1} + M[\Omega_{i+1}^w]^{-1})^{-1}$. A new draw of μ is obtained

as $\mu_{i+1} = \tilde{\zeta}_n + \Psi_i \omega$ where Ψ_i is the Cholesky factor of $\Sigma_{\zeta n}$ and ω is a draw of the multivariate standard normal.

$\rho_{mn}|\mu, \zeta_n, \Omega^b, \Omega^w$: Given ζ_{i+1} and Ω^w_{i+1}, the conditional posterior of $\rho_{mn,i+1}$ is proportional to $P^{mn}(\eta_{mn,i+1})\varphi(\eta_{mn,i+1}|\zeta_{n,i+1},\Omega^w_{i+1})$, where φ denotes the multivariate normal density with mean $\zeta_{n,i+1}$ and covariance matrix Ω^w_{i+1}, and $P^{mn}(\eta) =$

$$\prod_{j=0}^{J}\left[\frac{\exp\left(\frac{x_{jmn}\rho_{mn}-p_{jmn}}{\sigma(\rho_{mn})}\right)}{\sum_{i=0}^{J}\exp\left(\frac{x_{imn}\rho_{mn}-p_{imn}}{\sigma(\rho_{mn})}\right)}\right]^{d_{jmn}}$$

. A draw of $\rho_{mn,i+1}$ is ob-

tained by the following Metropolis–Hastings procedure: Draw ω_n from a jumping distribution, which we specify as normal with mean zero and variance $m\Omega^w$, where scalar m is adjusted over iterations to maintain an acceptance rate of 0.3. Calculate a trial value of $\rho_{mn,i+1}$ as $\tilde{\rho}_{mn} = \rho_{mn,i} + \omega_n$. Calculate the probability $P^{mn}(\tilde{\rho}_{mn})$. If a $[0,1]$ uniform random variate u satisfies

$$u < \frac{P^{mn}(\tilde{\rho}_{mn})\varphi(\tilde{\rho}_{mn}|\zeta_{n,i+1},\Omega^w_{i+1})}{P^{mn}(\rho_{mn,i})\varphi(\rho_{mn,i}|\zeta_{n,i+1},\Omega^w_{i+1})}, \tag{8.1}$$

then the new value of $\rho_{mn,i+1}$ is this trial $\tilde{\rho}_{mn}$; otherwise, discard the trial $\tilde{\rho}_{mn}$ and retain the old value as the new value, $\rho_{mn,i+1} = \rho_{mn,i}$.

The description of the estimator above assumes that all coefficients have inter- and intra-consumer distributions. However, the estimator also includes coefficients with only inter-consumer heterogeneity, or coefficients without any heterogeneity, as follows:

1. If some of the coefficients do not vary across subjects or menus, the procedure described above is only applied to coefficients with inter-consumer heterogeneity, while the sampling of these coefficients is obtained by an additional Metropolis–Hastings procedure.

2. Similarly, if some of the coefficients do not vary across menus, the procedure described above is only applied to coefficients with intra-consumer heterogeneity, and an additional Metropolis–Hastings procedure is used to sample inter only coefficients.

Similarly, the estimator is also able to impose zero covariance restrictions on Ω^b and Ω^w. If these matrices are specified to be block-diagonal, then sampling is performed by drawing from the different blocks separately using the steps specified above. For more detail see Becker *et al.* (2018).

A Monte Carlo Example

To demonstrate this HB procedure consider again the Monte Carlo example from Section 6.4. Intra-consumer heterogeneity is introduced to the table grape CBC data for the coefficients of "size" and "organic". Data is generated using the procedure sourced in part C of the codes at https://eml.berkeley.edu/~train/foundations_R.txt. In addition, an intra-consumer covariance matrix is obtained by defining a Cholesky factor of the correlation matrix, and intra-consumer standard deviations. The data generation process is presented in Figure 8.1. The correlation between the coefficients of "size" and "organic" (β_L and β_O) is specified to be -0.3. Additionally, three levels of intra-consumer heterogeneity are considered as presented in Table 8.1.

Population means and inter-consumer covariances are the same as those presented in Table 6.2. The values in the samples for intra-consumer heterogeneity are presented in Tables 8.2. The corresponding true values are presented in the Appendix.
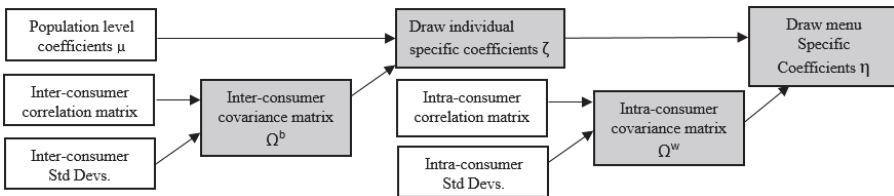

Figure 8.1: Data Generation Process — Inputs in White Boxes and Outputs in Shaded Boxes

Table 8.1: Levels of intra-consumer heterogeneity.

Heterogeneity level	β_L Std. Dev.	β_O Std. Dev.
Low	0.2	0.1
Medium	0.6	0.3
High	1	0.5

Table 8.2

	Std. Devs.	Covariances and Correlation in top right corner	
		β_L	β_O
Panel A: Low heterogeneity — values in sample.			
β_L	0.201	0.040	-0.291
β_O	0.099	-0.006	0.010
Panel B: Medium heterogeneity — values in sample			
β_L	0.602	0.362	-0.291
β_O	0.297	-0.052	0.088
Panel C: High heterogeneity — values in sample			
β_L	1.003	1.006	-0.291
β_O	0.496	-0.145	0.246

Model Estimation

The utility equation for each menu is similar to that presented in Equation (9.4), but with menu-specific coefficients β_{Lmn} and β_{Omn}:

$$U_{jmn} \equiv I_n - P_{jmn} + S_{jmn}\beta_{Sn} + C_{jmn}\beta_{Cn} + L_{jmn}\beta_{Lmn} + O_{jmn}\beta_{Omn}$$

$$+ SC_{jmn}\beta_{SCn} + SG_{jmn}\beta_{SGn} + B_{jmn}\beta_{qn} + \sigma_n\varepsilon_{jmn} \qquad (8.2)$$

The model is estimated using the five-step procedure described above. After 100,000 burn-in iterations, every tenth draw is taken from 100,000 subsequent draws for the estimation of the parameters. The estimation results are presented in Table 8.3. Models with only inter-consumer heterogeneity were also estimated using data sets with intra-consumer heterogeneity. The results are presented in the Appendix.

Table 8.4 presents the differences in the mean and standard deviation of the scale parameter $\text{Log}(\sigma_n)$ between models estimated with and without intra-consumer heterogeneity (using the data generated for the cases of low heterogeneity, medium heterogeneity, and high heterogeneity). The results indicate that ignoring intra-consumer heterogeneity results in smaller means and standard deviations of the scale parameter compared to the cases where intra-consumer heterogeneity is accounted

Table 8.3: Estimates from table grapes CBC data with intra-consumer heterogeneity.

Parameter	True values		No heterogeneity		Low heterogeneity		Medium heterogeneity		High heterogeneity	
Means	Population	Sample	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
					Panel A					
β_S	1	1.005	0.973	0.036	0.995	0.049	1.027	0.058	1.002	0.043
β_C	0.3	0.299	0.358	0.033	0.286	0.036	0.291	0.032	0.287	0.043
β_L	0.2	0.197	0.181	0.024	0.213	0.025	0.223	0.027	0.238	0.031
β_O	0.1	0.109	0.094	0.027	0.120	0.024	0.146	0.026	0.128	0.028
β_{SC}	0	0.001	−0.059	0.039	0.075	0.046	0.055	0.042	0.046	0.048
β_{SG}	0.1	0.093	0.144	0.046	0.053	0.051	0.044	0.074	0.077	0.045
β_{BN}	2	2.012	2.003	0.051	2.009	0.053	1.969	0.050	1.981	0.054
$\mathrm{Log}(\sigma_n)$	−0.5	−0.481	−0.526	0.024	−0.486	0.025	−0.486	0.027	−0.478	0.033
					Panel B					
Inter-consumer Std. Devs.										
β_S	0.3	0.304	0.268	0.058	0.200	0.044	0.213	0.046	0.227	0.050
β_C	0.1	0.099	0.167	0.037	0.190	0.057	0.157	0.043	0.194	0.055
β_L	0.1	0.101	0.137	0.033	0.165	0.039	0.168	0.045	0.171	0.047
β_O	0.2	0.210	0.195	0.034	0.159	0.041	0.205	0.043	0.192	0.044
β_{SC}	0.05	0.051	0.171	0.049	0.221	0.056	0.191	0.054	0.235	0.074
β_{SG}	0.2	0.202	0.202	0.059	0.199	0.066	0.208	0.057	0.213	0.069
β_{BN}	1	0.988	0.844	0.059	0.922	0.076	0.887	0.066	0.852	0.072
$\mathrm{Log}(\sigma_n)$	0.3	0.296	0.327	0.032	0.285	0.031	0.287	0.033	0.301	0.037

(Continued)

Table 8.3: (*Continued*)

Panel C

Intra-consumer Std. Devs. and covariance	No heterogeneity			Low heterogeneity			Medium heterogeneity			High heterogeneity		
	True value	Mean	Std. Dev.	True value	Mean	Std. Dev.	True value	Mean	Std. Dev.	True value	Mean	Std. Dev.
β_L std. dev.	0	0.186	0.045	0.2	0.213	0.096	0.6	0.681	0.072	1	1.063	0.068
β_O std. dev.	0	0.128	0.056	0.1	0.135	0.082	0.3	0.206	0.095	0.5	0.382	0.120
β_L, β_O Cov	0	0.010	0.012	−0.006	0.002	0.017	−0.054	−0.066	0.051	−0.15	−0.106	0.070

(*Continued*)

Table 8.3: (*Continued*)

Covariances	True values		No heterogeneity		Low heterogeneity		Medium heterogeneity		High heterogeneity	
	Population	Sample	Mean	Std. Dev.	Covariance	Std. Dev.	Covariance	Std. Dev.	Covariance	Std. Dev.
					Panel D					
β_S, β_C	1.8	0.0185	0.0141	0.1824	0.0978	0.1167	0.0095	0.1474	0.0123	0.1955
β_S, β_L	0	0.0007	−0.0060	0.0130	0.0001	0.0153	0.0089	0.0145	0.0080	0.0155
β_S, β_O	0	−0.0002	−0.0009	0.0204	0.0055	0.0093	0.0064	0.0152	0.0127	0.0158
β_S, β_{SC}	0.45	0.0046	0.0016	0.0191	−0.0027	0.0157	−0.0027	0.0146	−0.0091	0.0248
β_S, β_{SG}	0	−0.0004	0.0124	0.0181	0.0059	0.0181	0.0105	0.0149	0.0020	0.0168
β_S, β_{BN}	0	−0.0032	0.0975	0.0437	0.0623	0.0507	0.0631	0.0496	0.0850	0.0659
β_C, β_L	0.48	0.0050	0.0012	0.0691	0.0701	0.1352	0.0016	0.1064	0.0054	0.1150
β_C, β_O	0	0.0006	0.0069	0.1023	0.0264	0.0941	−0.0071	0.1453	0.0058	0.1408
β_C, β_{SC}	0.21	0.0022	−0.0143	0.1361	−0.1322	0.1606	−0.0050	0.1225	−0.0260	0.2507
β_C, β_{SG}	0	0.0000	−0.0023	0.1163	0.0738	0.1631	0.0054	0.1308	−0.0054	0.1795
β_C, β_{BN}	0	−0.0027	0.0786	0.4008	0.3574	0.5944	0.0300	0.4170	0.0759	0.6075
β_L, β_O	0	0.0009	0.0115	0.0084	−0.0048	0.0114	0.0098	0.0117	0.0116	0.0126
β_L, β_{SC}	0.21	0.0023	−0.0015	0.0075	−0.0020	0.0182	−0.0061	0.0141	−0.0074	0.0147
β_L, β_{SG}	0	0.0001	−0.0060	0.0111	−0.0046	0.0147	0.0005	0.0143	−0.0026	0.0163
β_L, β_{BN}	0	−0.0040	0.0046	0.0274	−0.0253	0.0378	0.0269	0.0363	0.0355	0.0397
β_O, β_{SC}	0.3	0.0032	−0.0056	0.0120	−0.0064	0.0134	−0.0121	0.0169	−0.0148	0.0176
β_O, β_{SG}	0	−0.0016	−0.0105	0.0151	0.0013	0.0105	−0.0101	0.0129	−0.0049	0.0213
β_O, β_{BN}	0	−0.0002	0.0535	0.0320	0.0509	0.0342	0.0497	0.0414	0.0677	0.0356
β_{SC}, β_{SG}	0	−0.0002	0.0120	0.0117	0.0047	0.0215	0.0138	0.0158	0.0121	0.0234
β_{SC}, β_{BN}	0	−0.0020	−0.0607	0.0465	−0.0943	0.0697	−0.0371	0.0555	−0.0781	0.0913
β_{SG}, β_{BN}	0	0.0081	−0.0402	0.0592	0.0137	0.0654	−0.0287	0.0648	−0.0482	0.0731

Table 8.4: Differences in the mean and standard deviation of the scale parameter between models with and without intra-consumer heterogeneity.

		Values in sample	No heterogeneity	Data Low heterogeneity	Medium heterogeneity	High heterogeneity
Estimator	**Means**					
	With intra-consumer heterogeneity	−0.481	−0.526 (109%)	−0.486 (101%)	−0.486 (101%)	−0.478 (99%)
	Without intra-consumer heterogeneity	−0.481	−0.516 (107%)	−0.468 (97%)	−0.402 (84%)	−0.301 (63%)
	Std. Devs.					
	With intra-consumer heterogeneity	0.296	0.327 (110%)	0.285 (96%)	0.287 (97%)	0.301 (101%)
	Without intra-consumer heterogeneity	0.296	0.323 (109%)	0.279 (94%)	0.253 (85%)	0.247 (83%)

Note: *Percentages represent deviations from the values in the sample.

Table 8.5: Differences in the inter-consumer standard deviations of the variables with intra-consumer heterogeneity between models with and without intra-consumer heterogeneity.

Std. Dev.(β_L)+Std. Dev.(β_O)	Values in sample	No heterogeneity	Low heterogeneity	Medium heterogeneity	High heterogeneity
With intra-consumer heterogeneity	0.311	0.333 (107%)	0.324 (104%)	0.373 (120%)	0.363 (117%)
Without intra-consumer heterogeneity	0.311	0.358 (115%)	0.382 (123%)	0.409 (132%)	0.438 (141%)

Note: *Percentages represent deviations from the values in the sample.

for. Smaller values for the scale parameter indicate the bias is due to a greater degree of unobserved effects. In addition, when ignoring intra-consumer heterogeneity, increasing values of the standard deviations of coefficients with intra-consumer heterogeneity are observed as indicated in Table 8.5. Thus, the unspecified intra-consumer heterogeneity results in over-estimation of inter-consumer heterogeneity. This Monte Carlo example demonstrated the application of the extended Inter-/Intra-HB estimator and indicated that if intra-heterogeneity is present then an inter only estimator may overstate the level of inter-consumer heterogeneity.

A careful analysis of the HB Markov chains generated by the Monte Carlo examples above and in Section 6 indicated two issues that require further attention. The first issue is that the Inverted Wishart prior was found to be informative. The examples required data scaling to reduce the resulting bias caused by properties of the Inverted Wishart. For more details and alternatives to inverse Wishart, see Song *et al.* (2018) and Akinc and Vanderbroek (2018). The second issue is the need for unusually long Markov chains in cases of weak identification. This indicated a need for systematic procedures to determine the required length (200K iterations in the examples) and thinning (1 every 10 in the examples) of the Markov chains.

To allow more flexible distributions than the Normality assumption, a straightforward extension is a discrete mixture of Normals obtained by adding a Multinomial to the likelihood function and a Dirichlet prior. For more details, see Danaf *et al.* (2018).

9

Policy Analysis

Stated preference elicitations and modeling of consumer tastes and choices are often undertaken for the purpose of quantifying the impacts of alternative policies on suppliers, consumers, and markets. Clients for such policy analysis can include businesses interested in the financial impacts of product innovations, governments who need to know the impacts of market regulations on transactions and product standards, and litigants who want to determine the harm caused by "as is" policy compared with "but for" policy that corrects improper conduct. Economic cost–benefit and business strategic analyses are usually *prospective*, forecasting the impacts of alternative policies on future outcomes for broad classes of consumers, e.g., the effect on consumers of food and drug product safety regulations, or the impact on suppliers of new product offerings. In contrast, analyses relevant to antitrust, patent, environmental, and tort law are primarily *retrospective*, determining the consequences for individuals or specific consumer classes of alternatives to actual past policy, e.g., the harm to buyers of cars with unsafe ignition switches, or to homeowners in a neighborhood with groundwater contamination. However, the separation between prospective and retrospective analysis is imperfect, as realized outcomes and *ex*

post equity are also of concern to economists, while jurisprudence is also concerned with the *ex ante* incentive effects of remedies for harm and with costs incurred by consumers to mitigate potential harm, see Rawls (1971), Hammond (1981), Kaplow and Shavell (2002), and Fleurbaey and Maniquet (2011). Revealed and stated preference studies can provide forecasts of average per capita market demands under alternative policies for classes of consumers in either prospective or retrospective analysis. There is no additional information in prospective analysis on the choices that consumers would actually make under each policy, and no direct information on differences between anticipated and realized outcomes. However, in retrospective analysis, the analyst also observes some realized outcomes such as revealed product defects, and observes consumers' choices of products under actual market conditions, which often define relevant consumer classes. In this section, we first describe general-purpose simulation methods that line up well the use of hierarchical Bayes estimates and can be used for a variety of policy purposes. Following this, we discuss policy impacts on suppliers, a motivation for the development of CBC methods in market research, and conclude with a discussion of measuring consumer well-being.

9.1 Policy simulations

Suppose consumers face policy alternatives $m = 1, 2$, and note that this is a different use of the index m than in Sections 4–6 where m indexed menus presented to each subject in a CBC experiment. Suppose consumers have observed characteristics (W, A, T, s), and face menus containing alternatives $j = 0, \ldots, J_m$ with attributes (z_{jm}, p_{jm}). For welfare analysis later, it will be useful to also introduce a net incremental income transfer t from the consumer to the government. Assume utilities are of a fully neoclassical money-metric form Equation (3.1), $u_{jm}(t, \zeta, \varepsilon) = T - t + v_{jm}(t, \zeta) + \sigma(\zeta)\varepsilon_j$, with *net values* $v_{jm}(t, \zeta) \equiv X(W, A, T - t - p_{jm}, s, z_{jm})\beta(\zeta) - p_{jm}$, tastes ζ, i.i.d. EV1 psychometric noise ε_j, a parametric taste distribution $F(\zeta|s, \theta)$, and WTP parameters $\sigma(\zeta)$ and $\beta(\zeta)$ that are predetermined functions of ζ. These consumers are fully neoclassical since ζ and the ε_j are fixed across policy regimes. Giving a branded product distinct indices in $\mathbf{J}_1$ and $\mathbf{J}_2$,

which may be appropriate in some applications, allows psychometric effects across policies. However, this device rules out forms of dependence in noise across policies intermediate between total independence and total dependence. The treatment of this specification can have significant effects on the distribution of policy impacts across classes of consumers making different "as is" choices.[1]

When preferences are heterogeneous and choices are discrete, revealed and stated preference studies in practice provide insufficient information to predict reliably the demands and utilities of individual consumers. It is necessary to use bounds, approximations, or statistical averages over classes of consumers, conditioned on what can be usefully observed and distinguished among them, to obtain reliable estimates. When consumer choices revealed in the market or stated in CBC experiments allow estimation using the methods of Sections 4–6 of choice probabilities of the form (3.8), then the estimated underlying taste distribution $F(\zeta|s, \hat{\theta})$, Equation (3.9) can be used, along with the menus available to consumers under alternative policies, to forecast demands under various product or public policy innovations. Further, $F(\zeta|s, \hat{\theta})$ can be used to predict the distribution of money-metric decision utilities of consumers for these menus of products.

A straightforward but computationally intensive approach to policy analysis is to construct a synthetic population with the distributions of utilities and tastes described above, simulate their choices and utilities for the menus they face under each policy, and then process the results to obtain measures of interest. Start with a representative sample $n = 1, \ldots, N$ of consumer characteristics, (W_n, A_n, T_n, s_n), drawn from a general population survey, or from a CDC sample if it is representative. For each n, make R draws $\zeta_{1n}, \ldots, \zeta_{Rn}$ from $F(\zeta|s_n, \hat{\theta})$, and Q i.i.d. EV1 draws of psychometric noise vectors $\varepsilon_{qn}, \ldots, \varepsilon_{qn}$.[2] If hierarchical Bayes

---

[1]Giving a branded product distinct indices in $\mathbf{J}_1$ and $\mathbf{J}_2$ reintroduces variation in psychometric effects across policies, which may be appropriate in some applications. Ruled out by this device are forms of dependence in noise across policies intermediate between total independence and total dependence.

[2]The application will determine whether ε_{qj1n} and ε_{qj2n} are independent, identical, or exhibit some other form of dependence, and this specification can have

estimation of $F(\zeta|s,\hat{\theta})$ is used, then the values $\zeta_{1n},\ldots,\zeta_{Rn}$ may be obtained directly from the estimation process. Each combination of draws (n,r,q) defines a synthetic consumer. For policy m, construct the available menu of alternatives (z_{jm},p_{jm}) for $j=1,\ldots,J_m$. As noted earlier in the table grape CBC example, this step may be non-trivial, particularly if some attributes are ambiguous, non-generic, or brand-specific, or are stated in "consumer-friendly" terms that may be difficult to translate from technical attributes controlled by the product supplier, or if there are questions of how much "brand loyalty" and current consumer information will transfer to new menus. Further, product prices and attributes in scenario m may not be predetermined, but rather determined through market equilibrium. Then, the analyst must model producer behavior, and solve for the fixed point in prices and attributes that balance demand and supply. In many applications, this requires repetition of the steps below to obtain demands for a grid of product prices and attributes, calculation of supplies at each grid point, and search for the grid point that approximates equilibrium, with grid refinements as necessary.

For each synthetic consumer (n,r,q), calculate $\sigma(\zeta_{rn})$ and $\beta(\zeta_{rn})$, the net value functions $v_{jmn}(t,\zeta_{rn}) \equiv X(W_n,A_n,T_n - t - p_{jmn},s_n,z_{jmn})\beta(\zeta_{rn}) - p_{jmn}$, and the utilities $u_{jmn}(t,\zeta_{rn},\varepsilon_{qn}) = T_n - t + v_{jmn}(t,\zeta_{rn}) + \sigma(\zeta_{rn})\varepsilon_{jqn}$ When $t \neq 0$, this calculation will be repeated for each t of interest. Finally, calculate the choice that maximizes utility, indicated by $d_{jmn}(t,\zeta_{rn},\varepsilon_{qn})$ for each consumer and policy, the maximum money-metric utility $u_{mn}(t,\zeta_{rn},\varepsilon_{qn}) \equiv \max_{j\in J_m} u_{jmn}(t,\zeta_{rn},\varepsilon_{qn})$, the marginal utility of transfer income given choice j, $\mu_{jmn}(t,\zeta_{rn}) = 1 - \partial v_{jmn}(t,\zeta_{rn})/\partial t$, and the unconditional marginal utility of transfer income $\mu_{mn}(t,\zeta_{rn},\varepsilon_{mqn}) \equiv \sum_{j\in J_m} d_{jmn}(t,\zeta_{rn},\varepsilon_{mqn})\mu_{jmn}(t,\zeta_{rn})$.

With the information in hand for this synthetic population, it is a mechanical exercise to form statistical averages to estimate measures of interest. For example, one can estimate demands for each product and the revenues generated under each policy, and if it is a target of the investigation, the policy impacts on selected sub-populations. When

---

significant effects on the distribution of policy impacts across classes of consumers making different "as is" choices.

necessary, this can be combined with a model of producer behavior to determine market equilibrium prices and attributes under each policy. Using measures described in Section 9.3, one can also estimate consumer welfare effects of policy changes. Confidence intervals can be obtained by a Krinsky and Robb (1986) and Krinsky and Robb (1990) procedure with the estimated covariance of $\hat{\theta}$, or by a block bootstrap in which repeated estimates of the deep parameters are obtained by resampling from the survey respondents.

Both the Allenby–Train and NUTS procedures produce intermediate parameter draws ζ_{ni} at iteration i whose empirical distribution over retained trials estimates the posterior density $f(\zeta_n|s_n, \theta)$ conditioned on the choices d_n stated by respondent n in a CBC sample. Given the limited data a CBC study provides on individual respondents, these estimates will not be individually statistically consistent, but averages of smooth functions of the empirical distribution will be statistically well-behaved. Then, if the CBC sample is representative and is used to construct the synthetic population, these empirical distributions can be used to estimate expressions such as changes in levels of maximized money-metric utilities for classes of consumers that are large enough for laws of large numbers to yield reliable averages.

9.2 Demand analysis

Consider a prospective policy intervention that changes the menus of products that consumers face. For example, for a product like solar panels, one could forecast the impact of introduction of a new product features by suppliers, or of new government subsidies. Assume that consumers know product attributes and prices when they make choices in the future market. Suppose consumers are described as in Section 9.1, with tastes that are fixed across policies and fully neoclassical. Let "$\mathbf{E}_n$" denote an expectation over the consumer population; this is equivalent to indexing by n the observations (W, A, T, s) on the characteristics of consumers and (z_{jm}, p_{jm}) on the attributes and prices of each product j under an incumbent policy $m = 1$ and a replacement policy $m = 2$. Assume that $t = 0$. Let $(z_{j\lambda}, p_{j\lambda})$ for $\lambda \in [1, 2]$ denote a rectifiable path between policies for product attributes and prices. In the case that a

product is unavailable under policy m, we set its price to a large positive number so that it effectively drops out of the choice set. The estimated (per capita) demand D_j and revenue R_j for product j along this path are given by

$$
\begin{aligned}
D_j(\lambda) &= \mathbf{E}_n \mathbf{E}_{\zeta|s} P_j(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta) \\
R_j(\lambda) &= \mathbf{E}_n E_{\zeta|s} p_{j\lambda} P_j(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta).
\end{aligned}
\tag{9.1}
$$

where P_j is the MNL model (3.5) evaluated at $\rho_\lambda = \zeta$. Note that P_j declines exponentially for large $p_{j\lambda}$, so that $R_j(\lambda)$ is bounded and eventually decreasing as $p_{j\lambda} \to +\infty$. The change in demand along $\lambda \in [1, 2]$ is

$$
\begin{aligned}
\frac{dD_j(\lambda)}{d\lambda} &= \mathbf{E}_n \mathbf{E}_{\zeta|s} P_j(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta) \\
&\quad \times \frac{\delta_{jk} - P_k(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta)}{\alpha(\zeta)} \left\{ \frac{dx_{k\lambda n}}{d\lambda} \cdot \beta(\zeta) - \frac{dp_{k\lambda n}}{d\lambda} \right\},
\end{aligned}
\tag{9.2}
$$

where δ_{jk} is one if $j = k$, zero otherwise. This formula is easy to interpret in special cases. For example, if the price of good k is the same for all n and the policy changes only this price, and no $\mathbf{x}$ attributes or other prices, then Equation (9.2) gives the cross-elasticity of demand for good j with respect to the price of good k,

$$
\begin{aligned}
\frac{\partial \log D_j(\lambda)}{\partial \log p_{k\lambda}} &\equiv \frac{p_{k\lambda}}{D_j(\lambda)} \frac{dD_j(\lambda)/d\lambda}{dp_{k\lambda}/d\lambda} \\
&= -\mathbf{E}_n \mathbf{E}_{\zeta|s} \frac{(\delta_{jk} - P_k(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta)) p_{k\lambda}}{\alpha(\zeta)} \\
&\quad \cdot \frac{P_j(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta)}{D_j(\lambda)},
\end{aligned}
\tag{9.3}
$$

a population-and-taste-weighted average (i.e., over n and ζ) of MNL price elasticities.

A typical economic application of Equation (9.2) is to calculate the change in profit for the producer of product j resulting from a change Δp_{j1} in its product price and Δz_{j1} in its attributes. The change Δz_{j1} induces a change that may span multiple components of $x_{j1}: \Delta x_{j1} =$

$X(W, A, T - p_{j1} - \Delta p_{j1}, s, z_{j1} + \Delta z_{j1}) - X(W, A, T - p_{j1}, s, z_{j1})$. Assume that Δx_{j1} and Δp_{j1} are the same for all consumers in the population. Along a linear path $x_{j\lambda} = x_{j1} + (\lambda - 1)\Delta x_{j1}$ and $p_{j\lambda} = p_{j1} + (\lambda - 1)\Delta p_{j1}$ for $\lambda \in [1, 2]$, the estimated changes in demand and revenue are

$$
\begin{aligned}
\frac{dD_j(\lambda)}{d\lambda} &= \mathbf{E}_n \mathbf{E}_{\zeta|s} \left(\frac{(1 - P_j(W_n, A_n, T_n, s_n, \mathbf{x}_{\lambda n}, \mathbf{p}_{\lambda n}, \zeta)) p_{j\lambda 1}}{\alpha(\zeta)} \right) \\
&\quad \times (\Delta x_{j1} \beta(\zeta) - \Delta p_{j1}) P_j(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta), \\
\frac{dR_j(\lambda)}{d\lambda} &= \mathbf{E}_n \mathbf{E}_{\zeta|s} \left(1 - \frac{(1 - P_j(W_n, A_n, T_n, s_n, \mathbf{x}_{\lambda n}, \mathbf{p}_{\lambda n}, \zeta)) p_{j\lambda 1}}{\alpha(\zeta)} \right) \\
&\quad \times (\Delta x_{j1} \beta(\zeta) - \Delta p_{j1}) P_j(W, A, T, s, \mathbf{x}_\lambda, \mathbf{p}_\lambda, \zeta)
\end{aligned}
\tag{9.4}
$$

population-and-taste-weighted averages of MNL elasticity terms times $(\Delta x_{j1}\beta(\zeta) - \Delta p_{j1})$.[3] Combined with estimates of the marginal cost of product modifications and assumptions (e.g., Nash) on the response of other suppliers to its innovations, Equation (9.4) allows estimation of the incremental profitability of the change. Note that when $\beta = \beta(\zeta)$ is homogeneous, demand and revenue are unchanged when $\Delta x_{j1}\beta - \Delta p_{j1} = 0$, and the innovation increases profit if and only if it is cost-saving. For innovations that are not uniform in their incidence on consumers, when $\beta(\zeta)$ is heterogeneous, or when $\Delta x_{j1}\beta - \Delta p_{j1} \neq 0$, the profit analysis requires full calculation of Equation (9.2). Note that the formulas (9.1)–(9.4) can all be estimated using the synthetic population of consumers $n = 1, \ldots, N$ with a distribution of tastes ζ_{rn} constructed in Section 8.1, without sampling from the EV1 distributions, and that this alternative will reduce sampling variance relative to complete reliance on the full synthetic population (n, r, q).

9.3 Consumer welfare analysis

Suppose the population of consumers with nearly neoclassical decision utilities described in Section 9.1, and consider an incumbent/default/"as is" policy $m = 1$ and a replacement/candidate/"but for"

---

[3]If the supplier makes multiple products, the sum of revenues in (38) over its products leads to an extension of (41) that incorporates cross-elasticities between its products and allows analysis of the effects of product positioning and cannibalization.

policy $m = 2$. McFadden (2018) and McFadden and Train (2018) develop the consumer theory foundations and practical formulas for applied welfare analysis. This analysis can be applied to various combinations of stated and revealed preference data. In particular, the welfare consequences of policy changes can be calculated from the utility function of each simulated consumer in Section 9.1, and then aggregated to the population, or to classes that are sufficiently large so that the aggregates are statistically reliable.

There are several issues, detailed in McFadden (2018), that need to be resolved in setting up these calculations. First, welfare applications may be either prospective, examining policy alternatives that might be implemented in the future, or retrospective, calculating the benefit or harm to consumers from past policies. For example, determining the impact on consumer welfare of a proposed merger of suppliers requires a prospective analysis, while assessing the harm to consumers from past collusion between suppliers requires a retrospective analysis. Second, while retrospective analysis is usually conducted for the purpose of computing transfers to be fulfilled to "make consumers whole", the transfers in prospective analysis are often hypothetical, intended to determine whether the policy change could in principle be a Pareto improvement, rather than fulfilled to ensure an actual Pareto improvement. Third, market and stated preference studies usually provide only partial observability, insufficient to identify reliably the tastes of individual consumers, and at best identifying the distribution of tastes conditioned on observed histories and socioeconomic status. Then, any system of fulfilled transfers based on observed characteristics will generally lead to gainers and losers. Consequently, compensation calculations must consider both overall welfare and the impact of imprecision in individual compensating transfers. Fourth, experienced-utility of consumers may differ from the decision-utility that determines market demands, as the result of resolution of contingencies regarding attributes of products and interactions with consumer needs, or as the result of inconsistencies in tastes and incomplete optimizing behavior. There is then a question of whether the target of welfare analysis is experienced utility, reflecting outcomes, or decision utility, reflecting opportunities.

10

Conclusions

This monograph has reviewed methods for stated preference elicitation, concentrating on the most successful and widely applied approach, choice-based conjoint (CBC) analysis. It has detailed recommendations for the design of CBC studies, including sampling, subject training, incentive alignment, and reconciliation and validation of CBC data against revealed preference data. It has identified the conditions under which CBC data provide a reliable basis for forecasting demand for new or modified products, and cautioned readers about the sensitivity of CBC respondents to context and framing, the importance of accounting for the manipulability of consumers in both real and hypothetical markets, and in realistic hypothetical markets mimicking the manipulations present in real markets.

This monograph has reviewed the choice-theory underpinnings of CBC, and recommends modeling utility in money-metric or WTP-space form, introducing flexible generic attributes and taste heterogeneity in the population, and controlling carefully how components of income enter utility, discrete choice probability models, and measures of consumer welfare. It recommends the use of mixed logit choice models adapted to the CDC panel data structure to model the behavior of

nearly neoclassical consumers who aside from psychometric noise have stable tastes across different CBC menus.

The monograph has discussed extensively the use of simulated maximum likelihood and hierarchical Bayes methods to estimate the deep parameters that characterize distributions of tastes and choices in the population, and described methods that facilitate computation of both types of estimators. Hierarchical Bayes methods, particularly the Allenby-Train and NUTS procedures with diffuse priors, are shown to be very fast, even for high-dimensional problems, and for CBC studies of practical size are shown to be very close to classical maximum likelihood estimators. Therefore, we recommend use of these estimators.

Our overall conclusion is that collection and analysis of stated preference data is a powerful tool for exploring and predicting consumer behavior, but both the data collection and analysis require considerable care and caution in inferring that the results are predictive and reliable for real market behavior. We urge economist and non-economists to evaluate realistically what stated preference methods can do now, avoid overselling results from stated preference experiments in circumstances where they have not been proven reliable, and to engage in a vigorous, wide-ranging, cross-disciplinary research effort to improve these methods, expand them to provide structural information on consumer behavior such as distinctions between perceptions, decision utility, and experienced utility, and systematically monitor the performance of these methods in applications.

Appendix

True values of intra-consumer heterogeneity and estimation results with inter-only estimator

Intra-consumer heterogeneity — True values:

Table A1: Low heterogeneity — population

	Std. Devs.	Covariances and Correlation in top right corner	
		β_L	β_O
β_L	0.2	0.040	−0.3
β_O	0.1	−0.006	0.010

Table A2: Medium heterogeneity — population

	Std. Devs.	Covariances and Correlation in top right corner	
		β_L	β_O
β_L	0.6	0.360	−0.3
β_O	0.3	−0.054	0.09

Table A3: High heterogeneity — population

	Std. Devs.	Covariances and Correlation in top right corner	
		β_L	β_O
β_L	1	1	-0.3
β_O	0.5	-0.15	0.25

Estimation results for models without intra-consumer heterogeneity

Table A4: Data with inter and intra-consumer heterogeneity (three levels of intra-consumer heterogeneity), model with only inter-consumer heterogeneity

Parameter Means	True Values Population	Sample	No Heterogeneity Mean	Std. Dev.	Low Heterogeneity Mean	Std. Dev.	Medium Heterogeneity Mean	Std. Dev.	High Heterogeneity Mean	Std. Dev.
					Panel A: Parameter estimates					
β_S	1	1.005	0.950	0.044	0.983	0.041	1.020	0.046	1.015	0.055
β_C	0.3	0.299	0.337	0.034	0.295	0.031	0.292	0.049	0.289	0.040
β_L	0.2	0.197	0.176	0.025	0.221	0.026	0.236	0.027	0.260	0.027
β_O	0.1	0.109	0.095	0.026	0.123	0.026	0.148	0.025	0.127	0.028
β_{SC}	0	0.001	−0.023	0.039	0.074	0.037	0.056	0.064	0.041	0.051
β_{SG}	0.1	0.093	0.142	0.062	0.084	0.042	0.059	0.055	0.069	0.051
β_{BN}	2	2.012	2.016	0.050	2.002	0.051	1.969	0.055	1.973	0.055
$\text{Log}(\alpha_n)$	−0.5	−0.481	−0.516	0.023	−0.468	0.023	−0.402	0.022	−0.301	0.022
				Panel B: Standard deviations for inter-consumer heterogeneity						
β_S	0.3	0.304	0.257	0.063	0.202	0.054	0.185	0.052	0.201	0.046
β_C	0.1	0.099	0.177	0.047	0.152	0.042	0.154	0.036	0.183	0.039
β_L	0.1	0.101	0.162	0.033	0.217	0.051	0.187	0.052	0.222	0.050
β_O	0.2	0.210	0.196	0.050	0.165	0.044	0.222	0.046	0.216	0.048
β_{SC}	0.05	0.051	0.181	0.058	0.180	0.051	0.212	0.048	0.214	0.048
β_{SG}	0.2	0.202	0.202	0.051	0.222	0.067	0.247	0.075	0.204	0.062
β_{BN}	1	0.988	0.834	0.064	0.970	0.071	0.926	0.077	0.898	0.064
$\text{Log}(\alpha_n)$	0.3	0.296	0.323	0.030	0.279	0.029	0.253	0.025	0.247	0.026

(*Continued*)

Table A4: (*Continued*)

Covariances	True Values		No Heterogeneity		Low Heterogeneity		Medium Heterogeneity		High Heterogeneity	
	Population	Sample	Mean	Std. Dev.	Covariance	Std. Dev.	Covariance	Std. Dev.	Covariance	Std. Dev.
Panel C: Covariances for inter-consumer heterogeneity										
β_S, β_C	1.8	0.0185	0.121	0.138	0.032	0.128	0.001	0.107	0.010	0.099
β_S, β_L	0	0.0007	−0.008	0.013	0.003	0.018	0.007	0.014	0.001	0.016
β_S, β_O	0	−0.0002	−0.007	0.014	0.007	0.015	0.011	0.015	0.006	0.012
β_S, β_{SC}	0.45	0.0046	0.002	0.016	0.005	0.015	0.003	0.015	−0.006	0.015
β_S, β_{SG}	0	−0.0004	0.013	0.017	−0.004	0.023	0.002	0.023	0.013	0.016
β_S, β_{BN}	0	−0.0032	0.097	0.045	0.016	0.053	0.021	0.069	0.073	0.051
β_C, β_L	0.48	0.0050	0.061	0.097	0.044	0.145	0.001	0.103	0.003	0.143
β_C, β_O	0	0.0006	0.075	0.104	0.000	0.092	−0.008	0.113	−0.002	0.145
β_C, β_{SC}	0.21	0.0022	−0.184	0.171	−0.080	0.144	−0.002	0.132	−0.017	0.142
β_C, β_{SG}	0	0.0000	−0.011	0.122	0.094	0.157	0.002	0.154	−0.001	0.156
β_C, β_{BN}	0	−0.0027	0.835	0.367	0.101	0.648	0.002	0.597	0.078	0.449
β_L, β_O	0	0.0009	0.016	0.012	0.000	0.014	0.010	0.016	0.011	0.020
β_L, β_{SC}	0.21	0.0023	−0.005	0.011	0.000	0.016	−0.002	0.013	−0.009	0.023
β_L, β_{SG}	0	0.0001	−0.010	0.015	0.005	0.025	0.002	0.024	−0.004	0.023
β_L, β_{BN}	0	−0.0040	−0.002	0.030	−0.042	0.039	0.016	0.042	−0.005	0.041
β_O, β_{SC}	0.3	0.0032	−0.006	0.013	−0.001	0.008	−0.013	0.018	−0.014	0.018
β_O, β_{SG}	0	−0.0016	−0.013	0.019	−0.003	0.017	−0.019	0.029	−0.008	0.017
β_O, β_{BN}	0	−0.0002	0.046	0.032	0.040	0.035	0.054	0.042	0.062	0.035
β_{SC}, β_{SG}	0	−0.0002	0.007	0.015	0.004	0.015	0.028	0.022	0.006	0.021
β_{SC}, β_{BN}	0	−0.0020	−0.059	0.041	−0.051	0.072	−0.008	0.074	−0.086	0.053
β_{SG}, β_{BN}	0	0.0081	−0.002	0.062	0.023	0.068	0.003	0.065	−0.012	0.052

Acknowledgements

References

Abramowitz, M. and L. Stegum (1964). *Handbook of Mathematical Functions*. National Bureau of Standards, 24.4.

Acito, F. and A. Jain (1980). "Evaluation of Conjoint Analysis Results: A Comparison of Methods". *Journal of Marketing Research*. 17: 106–112.

Akaah, I. and P. Korgaonkar (1983). "An Empirical Comparison of the Predictive Validity of Self-explicated, Huber-hybrid, Traditional Conjoint, and Hybrid Conjoint Models". *Journal of Marketing Research*. 20: 187–197.

Akinc, A. and M. Vanderbroek (2018). "Bayesian Estimation of Mixed Logit Models: Selecting an Appropriate Prior for the Covariance Matrix". *Journal of Choice Modeling*. 29: 133–151.

Albert, J. and S. Chib (1996). "Computation in Bayesian Econometrics: An Introduction to Markov Chain Monte Carlo". In: *Bayesian Computational Methods and Applications*. Ed. by C. Hill. London: JAI Press. 3–24.

Allais, M. (1953). "Le Comportement de l'Homme Rationnel Devant Le Risque: Critique Des Postulats Et Axiomes de l'école Américaine". *Econometrica*. 21: 503–546.

Allenby, G. (1997). "An Introduction to Hierarchical Bayesian Modeling". Tutorial Notes, Advanced Research Techniques Forum, American Marketing Association.

Allenby, G. and P. Lenk (1994). "Modeling Household Purchase Behavior with Logistic Normal Regressions". *Journal of the American Statistical Association.* 89: 1218–1231.

Allenby, G. and P. Rossi (1993). "A Bayesian Approach to Estimating Household Parameters". *Journal of Marketing Research.* 30: 171–182.

Allenby, G. and P. Rossi (2006). "Hierarchial Bayes Models: A Practitioner's Guide". In: *The Handbook of Marketing Research.* Ed. by R. Grover and M. Vriens. Los Angeles: Sage Publications.

Andrews, R., A. Ansari, and I. Currim (2002). "Hierarchical Bayes versus Finite Mixture Conjoint Analysis Models: A Comparison of Fit, Prediction, and Partworth Recovery". *Journal of Marketing Research.* 39: 87–98.

Ariely, D., G. Lowenstein, and D. Prelec (2006). "Tom Sawyer and the Construction of Value". *Journal of Economic Behavior and Organization.* 60: 1–10.

Arrow, K. (1971). "Exposition of the Theory of Choice Under Uncertainty". In: *Theory of Risk-Bearing.* Ed. by K. Arrow. Chicago: Markham. 44–89.

Azizzadeh, F. and S. Rezakhah (2014). "The CUSUM Test for Detecting Structural Changes in Strong Mixing Processes". *Communications in Statistics.* 43(17): 3733–3750.

Bąk, A. and T. Bartłomowicz (2012). "Conjoint Analysis Method and Its Implementation in Conjoint R Package". Working paper, Wroclaw University.

Balcombe, K. and I. Fraser (2009). "Dichotomous-choice Contingent Valuation with 'Don't Know' Responses and Misreporting". *Journal of Applied Econometrics.* 24.7.

Bateson, J., D. Reibstein, and W. Boulding (1987). "Conjoint Analysis Reliability and Validity: A Framework for Future Research". In: *Review of Marketing.* Ed. by M. Houston. 451–481.

Becker, F., M. Danaf, B. A. X. Song, and M. Ben-Akiva (2018). "Hierarchical Bayes Estimator for Logit Mixtures with Inter- and Intra-consumer Heterogeneity". *Transportation Research Part B: Methodological.* 117(A): 1–17.

Becker, G., M. DeGroot, and J. Marschak (1964). "Measuring Utility by a Single-response Sequential Method". *Behavioral Science*. 9: 226–32.

Ben-Akiva, M. and D. Bolduc (1996). "Multinomial Probit with a Logit Kernel and a General Parametric Specification of the Covariance Structure". Working paper, Department of Civil Engineering, MIT.

Ben-Akiva, M., D. Bolduc, and M. Bradley (1993). "Estimation of Travel Model Choice Models with Randomly Distributed Values of Time". *Transportation Research Record*. 1413: 88–97.

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

Ben-Akiva, M. and T. Morikawa (1990). "Estimation of Switching Models from Revealed Preferences and Stated Intentions". *Transportation Research A*. 24: 485–495.

Berndt, E., B. Hall, R. Hall, and J. Hausman (1974). "Estimation and Inference in Nonlinear Structural Models". *Annals of Economic and Social Measurement*. 3(4): 653–665.

Berry, S., J. Levinsohn, and A. Pakes (1995). "Automobile Prices in Market Equilibrium". *Econometrica*. 63(4): 841–890.

Berry, S., J. Levinsohn, and A. Pakes (1998). "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market". Working paper, Yale University.

Berry, S., J. Levinsohn, and A. Pakes (2004). "Differentiated Products Demand Systems from a Combination of Micro and Macro Data: The New Car Market". *Journal of Political Economy*. 112: 68–105.

Betancourt, M. and M. Girolami (2013). "Hamilton Monte Carlo for Hierarchical Models". Working paper, arXiv:1312.0906.

Bhat, C. (2001). "Quasi-random Maximum Simulated Likelihood Estimation of the Mixed Multinomial Logit Model". *Transportation Research B*. 35: 677–693.

Bhat, C., S. Astroza, R. Sidharthan, and P. Bhat (2014). "A Multivariate Hurdle Count Data Model with an Endogenous Multiple Discrete-continuous Selection System". *Transportation Research B: Methodological*. 63: 77–97.

Bishop, R. and T. Heberlein (1979). "Measuring Values of Extramarket Goods: Are Indirect Measures Biased?" *American Journal of Agricultural Economics*. 61(5): 926–930.

Blass, A., S. Lach, and C. Manski (2010). "Using Elicited Choice Probabilities to Estimate Random Utility Models for Electricity Reliability". *International Economic Review*. 51: 421–440.

Bleimer, M. and J. Rose (2009). "Designing Stated Choice Experiments: The State of the Art". In: *The Expanding Sphere of Travel Behaviour Research*. Ed. by R. Kitamura, T. Yoshi, and T. Yamamoto. Bingley, West Yorkshire: Emerald Group Publishing, Ch. 25. 499–538.

Bohm, P. (1972). "Estimating the Demand for Public Goods: An Experiment". *European Economic Review*. 3: 111–130.

Bohm, P., J. Linden, and J. Sonnegard (1997). "Eliciting Reservation Prices: Becker DeGroot-Marschak Mechanisms vs. Markets". *The Economic Journal*. 107: 1079–1089.

Boyce, D. and H. Williams (2015). *Forecasting Urban Travel Demand*. Cheltenham: Edward Elgar.

Brookshire, D., B. Ives, and W. Schmitz (1976). "The Value of Aesthetic Preferences". *Journal of Environmental Economics and Measurement*. 3(4): 325–346.

Brookshire, D., A. Randall, and J. Stoll (1980). "Valuing Increments and Decrements in Natural Resource Service Flows". *American Journal of Agricultural Economics*. 62(3): 478–488.

Brownstone, D. (2001). "Discrete Choice Modeling for Transportation". In: *Travel Behavior Research: The Leading Edge*. Ed. by D. Hensher. Oxford: Elsevier. 97–124.

Brownstone, D., D. Bunch, and K. Train (2000). "Joint Mixed Logit Models of Stated and Revealed Preferences for Alternative-fuel Vehicles". *Transportation Research, Part B*. 34: 315–338.

Brownstone, D. and K. Train (1999). "Forecasting New Product Penetration with Flexible Substitution Patterns". *Journal of Econometrics*. 89: 109–129.

Burda, M., M. Harding, and J. Hausman (2012). "A Poisson Mixture Model of Discrete Choice". *Journal of Econometrics*. 166(2): 184–203.

Camerer, C. and R. Hogarth (1999). "The Effects of Financial Incentives in Experiments: A Review and Capital-labor-production Framework". *Journal of Risk and Uncertainty*. 19: 7–42.

Cameron, T. and J. DeShazo (2013). "Demand for Health Risk Reductions". *Journal of Environmental Economics and Management*. 65(1): 87–109.

Carmone, F., P. Green, and A. K. Jain (1978). "Robustness of Conjoint Analysis: Some Monte Carlo Results". *Journal of Marketing Research*. 15: 300–303.

Caro, F., T. Ho, D. McFadden, A. Gottlieb, C. Yee, T. Chan, and J. Winter (2012). "Using the Internet to Administer more Realistic Vignette Experiments". *Social Science Computer Review*. 30: 184–201.

Carpenter, B., A. Gelman, M. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M. Brubaker, J. Guo, P. Li, and A. Riddell (2017). "STAN: A Probabilistic Programming Language". *Journal of Statistical Software*. 76(1).

Carson, R. (2012). "Contingent Valuation: A Practical Alternative When Prices Aren't Available". *Journal of Economic Perspectives*. 26(4): 27–42.

Carson, R. T., J. J. Louviere, D. A. Anderson, P. Arabie, D. S. Bunch, D. A. Hensher, R. M. Johnson, W. F. Kuhfeld, D. Steinberg, J. Swait, H. Timmermans, and J. B. Wiley (1994). "Experimental Analysis of Choice". *Marketing Letters*. 5(4): 351–368.

Carson, R. and M. Czajkowski (2013). "A New Baseline Model for Estimating Willingness to Pay from Discrete Choice Models". Working paper, University of California, San Diego.

Carson, R., N. Flores, and N. Meade (2001). "Contingent Valuation: Controversies and Evidence". *Environmental and Resource Economics*. 19: 173–210.

Carson, R. and J. Louviere (2011). "A Common Nomenclature for Stated Preference Elicitation Approaches". *Environmental and Resource Economics*. 49(4): 539–559.

Chandukala, S., J. Kim, T. Otter, P. Rossi, and G. Allenby (2007). "Choice Models in Marketing: Economic Assumptions, Challenges, and Trends". *Foundations and Trends in Marketing*. 2(2): 97–184.

Chapman, R. and R. Staelin (1982). "Exploiting Rank Ordered Choice Set Data within the Stochastic Utility Model". *Journal of Marketing Research*. 14: 288–301.

Chen, R. (2017). "Models of Count with Endogenous Choices". *Transportation Research Procedia*. 23C: 460–479.

Chesher, A. and J. Santos-Silva (2002). "Taste Variation in Discrete Choice Models". *Review of Economic Studies*. 69: 62–78.

Chib, S. and E. Greenberg (1996). "Markov Chain Monte Carlo Simulation Methods in Econometrics". *Econometric Theory*. 12(3): 409–431.

Chipman, J. and J. Moore (1980). "Compensating Variation, Consumer's Surplus, and Welfare". *American Economic Review*. 70: 933–348.

Chipman, J. and J. Moore (1990). "Acceptable Indicators of Welfare Change, Consumer's Surplus Analysis, and the Gorman Polar Form". In: *Preferences, Uncertainty, and Optimality: Essays in Honor of Leonid Hurwicz*. Ed. by D. McFadden and M. Richter. New York: Westview Press. 68–120.

Ciriacy-Wantrup, S. (1947). "Capital Returns from Soil-conservation Practices". *Journal of Farm Economics*. 29: 1181–1202.

Corradi, V., N. Swanson, and H. White (2000). "Testing for Stationarity-ergodicity and for Comovements between Nonlinear Discrete Time Markov Processes". *Journal of Econometrics*. 96: 39–73.

Daly, A., S. Hess, and K. Train (2012). "Assuring Finite Moments for Willingness to Pay in Random Coefficient Models". *Transportation*. 39(1): 19–31.

Danaf, M., B. Atasoy, and M. Ben-Akiva (2018). "Logit Mixture with Inter and Intra-consumer Heterogeneity and Flexible Mixing Distributions". Working paper, MIT.

Davis, R. (1963a). "Recreation Planning as an Economic Problem". *Natural Resources*. 3: 239–249.

Davis, R. (1963b). *The Value of Outdoor Recreation: An Economic Study of the Maine Woods*. PhD Dissertation, Harvard University.

Deaton, A. (2012). "The Financial Crisis and the Well-being of America". In: *Investigations in the Economics of Aging*. Ed. by D. Wise. Chicago: University of Chicago Press. 343–376.

Deaton, A. and J. Muellbauer (1980). *Economics and Consumer Behavior.* Cambridge: Cambridge University Press.

Delavande, A. and C. Manski (2015). "Using Elicited Choice Probabilities in Hypothetical Elections to Study Decisions to Vote". *Electoral Studies.* 38: 28–37.

Delgado, M. (2005). "Sign Tests for Long-memory Time Series". *Journal of Econometrics.* 128(2): 215–251.

Desarbo, W., V. Ramaswamy, and S. Cohen (1995). "Market Segmentation with Choice-based Conjoint Analysis". *Marketing Letters.* 6: 137–147.

Domowitz, I. and M. El-Gamal (1993). "A Consistent Test of Stationarity-ergodicity". *Econometric Theory.* 9(4): 589–601.

Dong, S., M. Ding, and J. Huber (2010). "A Simple Mechanism to Incentive-align Conjoint Experiments". *International Journal of Research in Marketing.* 27: 25–32.

Elrod, T., J. Louviere, and K. Davey (1992). "An Empirical Comparison of Ratings-based and Choice-based Conjoint Models". *Journal of Marketing Research.* 29: 368–377.

Fink, D. (1997). "A Compendium of Conjugate Priors". Working paper, Montana State University.

Fischhoff, B. and C. Manski (1999). *Elicitation of Preferences.* Kluwer: Boston.

Fisher, I. (1892). *Mathematical Investigations in the Theory of Value and Prices.* New Haven: Yale.

Fleurbaey, M. and F. Maniquet (2011). *A Theory of Fairness and Social Welfare.* Cambridge: Cambridge University Press.

Fosgerau, M., D. McFadden, and M. Bierlaire (2013). "Choice Probability Generating Functions". *Journal of Choice Modelling.* 8: 1–18.

Frisch, R. (1926). "Sur Un Probleme of Economie Pure. *Norsk Matematisk Forenings Sskrifter*, 16, 1–40. Translation". In: *Preferences, Utility, and Demand.* Ed. by J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein. New York: Harcourt.

Gelman, A., J. Carlin, H. Stern, and D. Rubin (2004). *Bayesian Data Analysis* (2nd edition). Boca Raton, FL: Chapman & Hall/CRC.

Geweke, J. (1997). *Posterior Simulators in Econometrics.* Econometric Society Monographs 28. New York: Cambridge University Press.

Glasgow, G. and S. Butler (2017). "The Value of Non-personally Identifiable Information to Consumers of Online Services: Evidence from a Discrete Choice Experiment". *Applied Economics Letters.* 24(6): 392–395.

Gorman, W. (1953). "Community Preference Fields". *Econometrica.* 21: 63–80.

Gorman, W. (1961). "On a Class of Preference Fields". *Metroeconomica.* 13: 53–56.

Green, D., K. Jacowitz, D. Kahneman, and D. McFadden (1998). "Referendum Contingent Valuation, Anchoring, and Willingness to Pay for Public Goods". *Resource and Energy Economics.* 20: 85–116.

Green, P. (1974). "On the Design of Choice Experiments Involving Multifactor Alternatives". *Journal of Consumer Research.* 1: 61–8.

Green, P., A. Krieger, and Y. Wind (2001). "Thirty Years of Conjoint Analysis: Reflections and Prospects". *Interfaces.* 31: S56–S73.

Green, P. and V. Srinivasan (1978). "Conjoint Analysis in Consumer Research: Issues and Outlook". *Journal of Consumer Research.* 5: 103–123.

Green, P. and V. Srinivasan (1990). "Conjoint Analysis in Marketing: New Developments with Implications for Research and Practice". *Journal of Marketing.* 54: 3–19.

Hajivassiliou, V. and D. McFadden (1998). "The Method of Simulated Scores for the Estimation of LDV Models". *Econometrica.* 66: 863–896.

Hajivassiliou, V. and P. Ruud (1994). "Classical Estimation Methods for LDV Models Using Simulation". In: *Handbook of Econometrics.* Ed. by R. Engle and D. McFadden. Amsterdam: North-Holland. 2383–2441.

Hammond, P. (1981). "Ex-ante and Ex-post Welfare Optimality Under Uncertainty". *Economica.* 48: 235–250.

Hanemann, M., J. Loomis, and B. Kanninen (1991). "Statistical Efficiency of Double-bounded Dichotomous Choice Contingent Valuation". *American Journal of Agricultural Economics.* 73: 1255–1263.

Hauser, J. and V. Rao (2002). "Conjoint Analysis, Related Modelling, and Applications". In: *Advances in Marketing Research: Progress and Prospects.* Ed. by Y. Wind and P. Green. New York: Springer. 141–168.

Hauser, J. and G. Urban (1979). "Assessment of Attribute Importances and Consumer Utility Functions: von Neumann-Morgenstern Theory Applied to Consumer Behavior". *Journal of Consumer Research.* 5: 251–262.

Hausman, J. (2012). "Contingent Valuation: From Dubious to Hopeless". *Journal of Economic Perspectives.* 26(4): 43–56.

Heckman, J. (1981). "The Initial Parameters Problem and the Problem of Initial Conditions in Estimating a Discrete Time-discrete Data Stochastic Process". In: *Structural Analysis of Discrete Data.* Ed. by C. Manski and D. McFadden. Cambridge: MIT Press. 179–195.

Hensher, D. and M. Bradley (1993). "Using Stated Response Data to Enrich Revealed Preference Discrete Choice Models". *Marketing Letters.* 4: 39–152.

Hensher, D. and W. Greene (2002). "Specification and Estimation of Nested Logit Model". *Transportation Research Part B.* 36: 1–17.

Hensher, D. and W. Greene (2003). "The Mixed Logit Model: The State of Practice and Warnings for the Unwary". *Transportation.* 30: 133–176.

Hensher, D. and J. Louviere (1983). "Identifying Individual Preferences for International Air Travel: An Application of Functional Measurement Theory". *Journal of Transport Economics and Policy.* 17: 225–245.

Hensher, D., J. Rose, and A. Collins (2011). "Identifying Commuter Preferences for Existing Modes and a Proposed Metro in Sydney, Australia, with Special Reference to Crowding". *Public Transport.* 3: 109–147.

Hess, S. and J. Rose (2009). "Allowing for Intra-respondent Variations in Coefficients Estimated on Repeated Choice Data". *Transportation Research, Part B.* 43(6): 708–719.

Hess, S. and K. Train (2011). "Recovery of Inter- and Intra-personal Heterogeneity Using Mixed Logit Models". *Transportation Research, Part B.* 45(7): 973–990.

Hicks, J. (1939). *Value and Capital.* Oxford: Clarendon Press.

Hoffman, M. and A. Gelman (2014). "The No-U-Turn sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo". *Journal of Machine Learning Research.* 15: 1351–1381.

Hofstede, F. T., Y. Kim, and M. Wedel (2002). "Bayesian Prediction in Hybrid Conjoint Analysis". *Journal of Marketing Research.* 36: 253–261.

Hong, H. and B. Preston (2012). "Bayesian Averaging, Prediction, and Nonnested Model Selection". *Journal of Econometrics.* 167(2): 358–369.

Huber, J. (1975). "Predicting Preferences on Experimental Bundles of Attributes: A Comparison of Models". *Journal of Marketing Research.* 12: 290–297.

Huber, J. (1987). "Conjoint Analysis: How We Got Here and Where We Are". In: *Proceedings of the Sawtooth Software Conference on Perceptual Mapping, Conjoint Analysis, and Computer Interviewing.* 237–252.

Huber, J. and K. Train (2001). "On the Similarity of Classical and Bayesian Estimates of Individual Mean Partworths". *Marketing Letters.* 12: 259–269.

Huber, J., D. Wittink, J. Fiedler, and R. Miller (1993). "The Effectiveness of Alternative Preference Elicitation Procedures in Predicting Choice". *Journal of Marketing Research.* 30: 105–114.

Huber, J. and K. Zwerina (1996). "The Importance of Utility Balance in Efficient Choice Designs". *Journal of Marketing Research.* 33: 307–317.

Hudomiet, P., M. Hurd, and S. Rohwedder (2018). "The Causal Effects of Economic Incentives, Health, and Job Characteristics on Retirement: Estimates Based on Subjective Conditional Probabilities". RAND working paper.

Hurd, M. and D. McFadden (1998). "Consumption and Savings Balances of the Elderly: Experimental Evidence on Survey Response Bias". In: *Frontiers in the Economics of Aging.* Ed. by D. Wise. Chicago: University of Chicago Press. 353–387.

Hurwicz, L. and H. Uzawa (1971). "On the Integrability of Demand Functions". In: *Preferences, Utility and Demand*. Ed. by J. Chipman, L. Hurwicz, M. Richter, and H. Sonnenschein. New York: Harcourt, Brace and Jovanovich. 114–148.

Ibragimov, I. and R. Has'minskii (1973). "On the Approximation of Statistical Estimators by Sums of Independent Variables". *Soviet Math. Dokl.* 14: 883–887.

Islam, T., J. J. Louviere, and P. F. Burke (2007). "Modeling the effects of vincluding/excluding attributes in choice experiments on systematic and randon components". *Intern. J. of Reserch in Markeing*. 24: 289–300.

Jain, A., F. Acito, N. Malhotra, and V. Mahajan (1979). "A Comparison of the Internal Validity of Alternative Parameter Estimation Methods in Decompositional Multiattribute Preference Models". *Journal of Marketing Research*. 16: 313–322.

Joe, H. (1997). *Multivariate Models and Dependence Concepts*. London: Chapman & Hall.

Johnson, R. (1974). "Trade-off Analysis of Consumer Values". *Journal of Marketing Research*. 11: 121–127.

Johnson, R. (1999). "The Joys and Sorrows of Implementing HB Methods for Conjoint Analysis". Working Paper, Sawtooth Software.

Johnson, R. and B. Orme (1996). "How Many Questions Should You Ask in Choice-based Conjoint Studies?" Sawtooth Software technical report.

Kahneman, D. and A. Krueger (2013). *Developments in the Measurement of Subjective Well-Being*. Cheltenham: Edward Elgar.

Kahneman, D. and A. Tversky (1979). "Prospect Theory: An Analysis of Decisions Under Risk". *Econometrica*. 47: 263–291.

Kahneman, D. and A. Tversky (2000). *Choices, Values, and Frames*. Cambridge: Cambridge University Press.

Kaplow, L. and S. Shavell (2002). *Fairness and Welfare*. Cambridge: Harvard University Press.

King, G., C. Murray, J. Salomon, and A. Tandon (2004). "Enhancing the Validity and Cross-cultural Comparability of Measurement in Survey Research". *American Political Science Review*. 98: 191–207.

Kling, C., D. Phaneuf, and J. Zhao (2012). "From Exxon to BP: Has Some Number Become Better Than No Number?" *Journal of Economic Perspectives.* 26(4): 3–26.

Koop, G. (2003). *Bayesian Econometrics.* Chichester: John Wiley and Sons.

Krinsky, I. and L. Robb (1986). "On Approximating the Statistical Properties of Elasticities". *The Review of Economics and Statistics.* 68: 715–719.

Krinsky, I. and L. Robb (1990). "On Approximating the Statistical Properties of Elasticities: a Correction". *The Review of Economics and Statistics.* 72: 189–190.

Kuhfield, W., R. Tobias, and M. Garratt (1994). "Efficient Experimental Design with Marketing Research Applications". *Journal of Marketing Research.* 31: 545–557.

Lenk, P., W. DeSarbo, P. Green, and M. Young (1996). "Hierarchical Bayes Conjoint Analysis: Recovery of Partworth Heterogeneity from Reduced Experimental Designs". *Marketing Science.* 15: 173–191.

Lewandowski, D., D. Kurowicka, and H. Joe. (2009). "Generating Random Correlation Matrices Based on Vines and Extended Onion Method". *Journal of Multivariate Analysis.* 100: 1989–2001.

List, J. and C. Gallet (2001). "What Experimental Protocol Influences Disparities Between Actual and Hypothetical Stated Values?" *Environmental and Resource Economics.* 20(3): 241–254.

Ljung, G. and G. Box (1978). "On a Measure of a Lack of Fit in Time Series Models". *Biometrika.* 65(2): 297–303.

Lleras, J., Y. Masatlioglu, D. Nakajima, and E. Ozbay (2017). "When More is Less: Limited Consideration". *Journal of Economic Theory.* 170: 70–85.

Louviere, J. (1988). *Analyzing Decision Making: Metric Conjoint Analysis.* Newbury Park: Sage.

Louviere, J., T. Flynn, and R. Carson (2010). "Discrete Choice Experiments Are Not Conjoint Analysis". *Journal of Choice Modeling.* 3(3): 57–72.

Louviere, J., D. Hensher, and J. Swait (2000). *Stated Choice Methods: Analytics and Applications.* Cambridge: Cambridge University Press.

Louviere, J. and G. Woodworth (1983). "Design and Analysis of Simulated Consumer Choice or Allocation Experiments: An Approach Based on Aggregate Data". *Journal of Marketing Research.* 20: 350–367.

Luce, D. (1959). *Individual Choice Behavior.* Hoboken: John Wiley and Sons.

Luce, D. and P. Suppes (1965). "Preferences, Utility and Subjective Probability". In: *Handbook of Mathematical Psychology.* Ed. by R. Luce, R. Bush, and E. Galanter. Hoboken: John Wiley and Sons. 249–410.

Luce, D. and J. Tukey (1964). "Simultaneous Conjoint Measurement: A New Type of Fundamental Measurement". *Journal of Mathematical Psychology.* 1: 1–27.

Manski, C. (1999). "Analysis of Choice Expectations in Incomplete Scenarios". *Journal of Risk and Uncertainty.* 19: 49–66.

Manski, C. (2004). "Measuring Expectations". *Econometrica.* 72: 1329–1376.

Manski, C. (2007). "Minimax-regret Treatment Choice with Missing Outcome Data". *Journal of Econometrics.* 139: 105–115.

Marsaglia, G. (1972). "Choosing a Point from the Surface of a Sphere". *Annals of Mathematical Statistics.* 43(2): 645–646.

Marschak, J. (1960). "Binary Choice Constraints on Random Utility Indicators". In: *Stanford Symposium on Mathematical Methods in the Social Sciences.* Ed. by K. Arrow. Palo Alto: Stanford University Press.

McFadden, D. (1974). "Conditional Logit Analysis of Qualitative Choice Behavior". In: *Frontiers of Econometrics.* Ed. by P. Zarembka. New York: Academic Press. 105–142.

McFadden, D. (1981). "Econometric Models of Probabilistic Choice". In: *Structural Analysis of Discrete Data.* Ed. by C. Manski and D. McFadden. Cambridge: MIT Press. 198–272.

McFadden, D. (1986). "The Choice Theory Approach to Market Research". *Marketing Science.* 5: 275–297.

McFadden, D. (1994). "Contingent Valuation and Social Choice". *American Journal of Agricultural Economics.* 76: 689–708.

McFadden, D. (1996). "Lectures on Simulation-assisted Statistical Inference". Conference Presentation, EC-squared Conference, Florence, Italy; and Working Paper, Department of Economics, University of California, Berkeley.

McFadden, D. (1999). "Rationality for Economists?" In: *Elicitation of Preferences*. Ed. by B. Fischoff and C. Manski. Boston: KAP. 73–106.

McFadden, D. (2001). "Disaggregate Behavioral Travel Demand's RUM Side: A Thirty Year Retrospective". In: *The Leading Edge of Travel Behavior Research*. Ed. by D. Hensher. Oxford: Pergamon Press.

McFadden, D. (2004). "Welfare Economics at the Extensive Margin: Giving Gorman Polar Consumers Some Latitude". Working Paper, University of California, Berkeley.

McFadden, D. (2006). "Revealed Stochastic Preference: A Synthesis". In: *Rationality and Equilibrium*. Ed. by C. Aliprantis, R. Matzkin, and D. McFadden. Berlin: Springer.

McFadden, D. (2012). "Economic Juries and Public Project Provision". *Journal of Econometrics*. 166(1): 116–126.

McFadden, D. (2014a). "An Economist's Perspective on Environmental Damages". *The Eighth Annual Advanced Conference on Litigating Natural Resource Damages*. Law Seminars International.

McFadden, D. (2014b). "The New Science of Pleasure: Consumer Behavior and the Measurement of Well-being". In: *Handbook of Choice Modelling*. Ed. by S. Hess and A. Daly. Cheltenham: Edward Elgar. 7–48.

McFadden, D. (2015). *A Primer on Bayes Method for Decision-making and Statistical Analysis*. Lecture notes, University of California, Berkeley.

McFadden, D. (2017). "Stated Preference Methods and their Use in Environmental Use and Non-use Valuation". In: *Contingent Valuation of Environmental Goods*. Ed. by D. McFadden and K. Train. Cheltenham: Edward Elgar. 153–187.

McFadden, D. (2018). "Parametric Approximations of Preference Fields". Working paper, University of California, Berkeley.

McFadden, D., M. Ben-Akiva, A. Goett, and D. Bolduc (1986). "The Choice Theory Approach to Analyze the Preferences and Choices of Electric Utility Consumers". EPRI Final Report RP2671–1.

McFadden, D., A. Goett, and C.-K. Woo (1988). "Estimating Household Value of Electric Service Reliability with Market Research Data". *Energy Journal: Special Electricity Reliability Issue.* 9: 105–120.

McFadden, D. and P. Ruud (1994). "Estimation by Simulation". *Review of Economics and Statistics.* 76: 591–608.

McFadden, D. and K. Train (2000). "Mixed MNL Models for Discrete Response". *Journal of Applied Econometrics.* 15: 447–470.

McFadden, D. and K. Train (2017). *Contingent Valuation of Environmental Goods: A Comprehensive Critique.* Cheltenham: Edward Elgar.

McFadden, D. and K. Train (2018). "Welfare Economics in Product Markets". Working paper, University of California, Berkeley.

Miller, K., R. Hofstetter, H. Krohmer, and J. Zhang (2011). "How Should Consumers' Willingness to Pay be Measured? An Empirical Comparison of State-of-the-art Approaches". *Journal of Marketing Research.* 48(1): 172–184.

Moore, W. (2004). "A Cross-validity Comparison of Rating-based and Choice-based Conjoint Analysis Models". *International Journal of Research in Marketing.* 21: 299–312.

Morikawa, T., M. Ben-Akiva, and D. McFadden (2002). "Discrete Choice Models Incorporating Revealed Preferences and Psychometric Data". In: *Econometric Models in Marketing.* Ed. by P. Franses and A. Montgomery. Amsterdam: Elsevier. 29–55.

Moscati, I. (2007). "Early Experiments in Consumer Demand Theory: 1930–1970". *History of Political Economy.* 39: 359–401.

Muller, M. E. (1959). "A Note on a Method for Generating Points Uniformly on N-dimensional Spheres". *Comm. Assoc. Comput. Mach.* 2: 19–20.

Natter, M. and M. Feurstein (2002). "Real World Performance of Choice-based Conjoint Models". *European Journal of Operations Research.* 117: 448–458.

Neal, R. (2011). "MCMC using Hamiltonian Dynamics". In: *Handbook of Markov Chain Monte Carlo*. Ed. by S. B. *et al.* Ch. 5. London: Chapman and Hall.

Neslin, S. (1981). "Linking Product Features to Perceptions: Self-stated versus Statistically Revealed Importance Weights". *Journal of Marketing Research*. 18: 80–86.

Orme, B. (1999). "ACA, CBC, or Both? Effective Strategies for Conjoint Research". Working paper, Sawtooth Software.

Pesaran, H. and M. Weeks (2001). "Non-nested Hypothesis Testing: An Overview". In: *A Companion to Theoretical Econometrics*. Ed. by B. Baltagi. Oxford: Blackwell. 277–309.

Pollard, D. (1984). *Convergence of Stochastic Processes*. New York: Springer.

Press, W., W. Vetterling, S. Teukolsky, and B. Flannery (2007). *Numerical Recipes: The Art of Scientific Computing*. Cambridge University Press.

Quarteroni, A., R. Sacco, and F. Saleri (2007). *Numerical Mathematics*. New York: Springer.

Raghavarao, D., J. Wiley, and P. Chitturi (2010). *Choice-Based Conjoint Analysis: Models and Designs*. Boca Raton: CRC Press.

Ramsey, F. (1931). "Truth and Probability". In: *The Foundations of Mathematics and Other Logical Essays*. Ed. by R. Braithwaite. London: Kegan, Paul, Trench, Trubner and Co. 156–98.

Rao, V. (1977). "Conjoint Measurement in Marketing Analysis". In: *Multivariate Methods for Market and Survey Research*. Ed. by J. Sheth. Chicago: American Marketing Association. 257–286.

Rao, V. (2014). *Applied Conjoint Analysis*. Berlin: Springer.

Rawls, J. (1971). *A Theory of Justice*. Cambridge: Harvard University Press.

Reibstein, D., J. Batesonm, and W. Boulding (1988). "Conjoint Analysis Reliability: Empirical Findings". *Marketing Science*. 7: 271–286.

Revelt, D. and K. Train (1998). "Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level". *The Review of Economics and Statistics*. LXXX(4): 647–657.

Rose, J. and M. Bleimer (2009). "Constructing Efficient Stated Choice Experimental Designs". *Transport Reviews*. 29(5): 587–617.

Rose, J., S. Hess, M. Bliemer, and A. Daly (2015). "The Impact of Varying the Number of Repeated Choice Observations on the Mixed Multinomial Logit Model". Working paper, University of Leeds.

Rossi, P. (1979). "Vignette Analysis: Uncovering the Normative Structure of Complex Judgments". In: *Qualitative and Quantitative Social Research: Papers in Honor of Paul Lazarsfeld*. Ed. by R. Merton, J. Coleman, and P. Rossi. New York: Macmillan. 175–188.

Rossi, R., G. Allenby, and R. McCulloch (2005). *Bayesian Statistics and Marketing*. Hoboken: John Wiley & Sons Ltd.

Rubinstein, R., G. Samorodnitsky, and M. Shaked (1985). "Antithetic Variates, Multivariate Dependence and Simulation of Stochastic Systems". *Management Science*. 31(1): 66–77.

Samuelson, P. (1947). *Foundations of Economic Analysis*. Cambridge: Harvard University Press, 1983.

Samuelson, P. (1950). "The Problem of Integrability in Utility Theory". *Economica*. 17: 355–385.

Savage, L. (1954). *The Foundations of Statistics*. Hoboken: John Wiley & Sons Ltd.

Scarpa, R., M. Thiene, and K. Train (2008). "Utility in Willingness to Pay Space: A Tool to Address Confounding Random Scale Effects in Destination Choice to the Alps". *American Journal of Agricultural Economics*. 90: 994–1010.

Schkade, D. and J. Payne (1993). "Where Do the Numbers Come From? How People Respond to Contingent Valuation Questions". In: *Contingent Valuation: A Critical Assessment*. Ed. by J. Hausman. Amsterdam: North-Holland. 271–293.

Schultz, H. (1925). "The Statistical Law of Demand as Illustrated by the Demand for Sugar". *Journal of Political Economy*. 33: 481–504.

Schultz, H. (1928). *Statistical Laws of Demand and Supply with Special Application to Sugar*. Chicago: University of Chicago Press.

Segal, M. (1982). "Reliability of Conjoint Analysis: Contrasting Data Collection Procedures". *Journal of Marketing Research*. 19: 139–143.

Song, X., F. Becker, M. Danaf, B. Atasoy, and M. Ben-Akiva (2018). "Enhancement of Hierarchical Bayes Procedure for Logit Mixture". Working paper, MIT.

Srinivasan, V. (1988). "A Conjunctive-compensatory Approach to the Self-explication of Multiattributed Preferences". *Decision Sciences.* 19: 295–305.

Strasser, H. (1975). "The Asymptotic Equivalence of Bayes and Maximum Likelihood Estimates". *Journal of Multivariate Analysis.* 5: 206–226.

Swait, J. and W. Adamowicz (2001). "The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching". *Journal of Consumer Research.* 28: 135–148.

Thurstone, L. (1927). "A Law of Comparative Judgment". *Psychological Review.* 34: 273–286.

Thurstone, L. (1931). "The Indifference Function". *Journal of Social Psychology.* 2: 139–167.

Tierney, L. (1994). "Markov Chains for Exploring Posterior Distributions". *Annals of Statistics.* 22: 1701–1728.

Train, K. (1986). *Qualitative Choice Analysis.* Cambridge: MIT Press.

Train, K. (1999). "Mixed Logit Models for Recreation Demand". In: *Valuing Recreation and the Environment.* Ed. by J. Herriges and C. Kling. Northampton: Edward Elgar.

Train, K. (2000). "Halton Sequences for Mixed Logit". Working paper, University of California, Berkeley. URL: https://eml.berkeley.edu/wp/train0899.pdf.

Train, K. (2001). "A Comparison of Hierarchical Bayes and Maximum Simulated Likelihood for Mixed Logit". Working paper, University of California, Berkeley. URL: https://eml.berkeley.edu/~train/compare.pdf.

Train, K. (2007). "A Recursive Estimator for Random Coefficient Models". Working paper, University of California, Berkeley. URL: http://eml.berkeley.edu/~train/re.pdf.

Train, K. (2009). *Discrete Choice Methods with Simulation* (2nd edition). Cambridge: Cambridge University Press.

Train, K. and G. Sonnier (2005). "Mixed Logit with Bounded Distributions of Correlated Partworths". In: *Applications of Simulation Methods in Environmental and Resource Economics.* Ed. by A. Alberini and R. Scarpa. Dordrecht: Springer. 117–134.

Train, K. and M. Weeks (2005). "Discrete Choice Models in Preference Space and Willingness-to-pay Space". In: *Applications of Simulation Methods in Environmental and Resource Economics*. Ed. by A. Alberini and R. Scarpa. Dordrecht: Springer. 1–17.

Tversky, A. (1972). "Elimination by Aspects". *Psychological Review*. 79(4): 281–299.

Urban, G., J. Hauser, W. Qualls, B. Weinberg, J. Bohlmann, and R. Chicos (1997). "Validation and Lessons from the Field: Applications of Information Acceleration". *Journal of Marketing Research*. 34(1): 143–153.

Urban, G., J. Hauser, and J. Roberts (1990). "Prelaunch Forecasting of New Automobiles: Models and Implementation". *Management Science*. 36: 401–421.

Walker, J. L., Y. Wang, M. Thorhauge, and M. Ben-Akiva (2018). "D-efficient or Deficient? A Robustness of SP Experimental Design in a VOT Estimation Context". *Theory and Decision*. 84(2): 215–238.

Walker, J., M. Ben-Akiva, and D. Bolduc (2007). "Identification of Parameters in Normal Error Component Logit-mixture (NECLM) Models". *Journal of Applied Econometrics*. 22: 1095–1125.

Wallis, A. and M. Friedman (1942). "The Empirical Derivation of Indifference Functions". In: *Studies in Mathematical Economics and Econometrics*. Ed. by O. Lange, F. McIntyre, T. Yntema, and H. Schuktz. Chicago: University of Chicago Press. 175–89.

Wittink, D. and T. Bergestuen (2001). "Forecasting with Conjoint Analysis". In: *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Ed. by J. Armstrong. New York: Springer. 147–167.

Wittink, D. and D. Montgomery (2009). "Models of Job Preference for Stanford MBA's '78". Stanford University Graduate School of Business Research Paper No. 2018.

Wright, P. and M. Kriewall (1980). "State-of-mind Effects on Accuracy with Which Utility Functions Predict Marketplace Utility". *Journal of Marketing Research*. 17: 277–293.

Wu, M. and W. Fitzgerald (1996). "Bayesian Multimodal Evidence Computation by Adapti Tempering MCMC". In: *Maximum Entropy and Bayesian Methods.* Ed. by K. Hanson and R. Silver. Dordrecht: Springer.

Yadav, L. (2007). *Hypothetical Bias in Contingent Valuation.* Mountain View: Proquest.

Yanez, M., E. Cherchi, B. Heydecker, and J. Dios-Ortuzar (2011). "On the Treatment of Repeated Observations in Panel Data: Efficiency of Mixed Logit Parameter Estimates". *Networks and Spatial Economics.* 11(3): 393–418.

Yang, M. (2008). "Normal Log-normal Mixture: Leptokurtosis and Skewness". *Applied Economics Letters.* 15: 737–742.

Zellner, A. and P. Rossi (1984). "Bayesian Analysis of Dichotomous Quantal Response Models". *Journal of Econometrics.* 25: 365–394.

Web Sites

https://github.com/stan-dev/rstan/wiki/RStan-Getting-Started
http://mc-stan.org/
http://www.slideshare.net/surveyanalytics/how-to-run-a-conjoint-analysis-project-in-1-hour
http://artax.karlin.mff.cuni.cz/r-help/library/conjoint/html/00Index.html
http://cran.r-project.org/web/packages/conjoint/index.html
http://www.lawseminars.com/detail.php?SeminarCode=14NRDNM#agenda
https://eml.berkeley.edu/~train/foundations_R.txt
https://eml.berkeley.edu/~train/software.html
https://www.sheffield.ac.uk/economics/people/hole/stata