Durham Research Online

Econometrics of Stated Preferences

Denzil G. Fiebig[a] and Hong Il Yoo[b]
[a] School of Economics, University of New South Wales
[b] Durham University Business School, Durham University

Summary:
Stated preference methods are used to collect individual level data on what respondents say they would do when faced with a hypothetical but realistic situation. The hypothetical nature of the data has long been a source of concern amongst researchers as such data stand in contrast to revealed preference data which record the choices made by individuals in actual market situations. But there is considerable support for stated preference methods as they are a cost-effective means of generating data that can be specifically tailored to a research question and, in some cases, such as gauging preferences for a new product or non-market good, there may be no practical alternative source of data. While stated preference data come in many forms, the primary focus in this chapter will be data generated by discrete choice experiments and thus the econometric methods will be those associated with modelling binary and multinominal choices with panel data.

Key words:
Stated preference methods
Discrete choice experiments
Choice modelling
Ranked data
Best worst scaling
Mixed logit models
Latent class models
Combining preference data

Introduction

Stated preference (SP) methods are a useful source of individual level data on choices that individuals make or are likely to make. SP data record what respondents say they would do when faced with a hypothetical but realistic situation. Such data stand in contrast to revealed preference (RP) data which record the choices made by individuals in actual market situations. SP should be interpreted as a generic term that signals a type of survey data collection distinguished by its comparison with RP data. Notice that both SP and RP refer to data from which preferences can be inferred rather than representing actual preferences. While SP data come in many forms, the primary focus in this chapter will be data generated by discrete choice experiments (DCEs) and thus the econometric methods will be those associated with modelling binary and multinominal choices with panel data.

The hypothetical nature of SP data has long been a source of concern amongst researchers but there is considerable support for a more balanced appraisal as indicated by Manski (2004):

> "Economists have long been hostile to subjective data. Caution is prudent but hostility is not warranted."

Studies such as List et al. (2006), Vossler et al. (2012) and Kesternich et al. (2013) provide validation of SP methods and together with Layton and Levine (2003), Small et al. (2005) and Sándor and Frances (2009) illustrate the breadth of applications using SP data and confirm that there is considerable acceptance of their use across a range of disciplines.

It may seem somewhat curious that SP data collection is so popular in the Big Data environment of 2018. The deluge of raw material for potential input into producing empirical evidence needs to be balanced by the recognition that more data does not necessarily mean better data. It is invariably true that the data from new and rapidly expanding sources have not been collected with research in mind. They are not well suited to answer more nuanced and substantive questions because there is a mismatch between key concepts and available data or what's available suffers from sample selection problems. This is where SP methods have a comparative advantage because they are a cost-effective means of generating data that can be specifically tailored to the research questions. In some cases, such as gauging preferences for a new product or non-market good, there may be no practical alternative source of data.

Because SP data are generated with a particular research question in mind, there is considerable scope for innovative SP methods relating to research design including combining SP and RP data. As such the position of Carson and Hanemann (2005) seems entirely appropriate:

> "Rather than seeing an inherent conflict between revealed and stated preference techniques, it is more productive to view the two approaches as complementary but having different strengths and weaknesses".

McFadden (2001) provides a similar view:

> "There will always be questions about how closely cognitive tasks in a hypothetical setting can match those in a real decision-making environment. Good experimental technique can remove the most obvious sources of incongruity, but calibration and validation using RP data is usually needed."

By necessity there is a need to limit what is covered in this chapter. The aim is to provide an overview of SP methods with the primary focus on the econometric methods that are ultimately used once the SP data have been collected. While there is considerable overlap in the econometric methods that are used to analyse both SP and RP data there are differences and these will guide the selection of topics to be covered. There will be little discussion of the important issues relating to the development and implementation of the choice survey that precedes data analysis. For general issues of survey design see Groves et al. (2009) and for experimental design see Street and Burgess (2007). Nor will there be any discussion of contingent valuation which is another form of SP data that is especially prevalent in environmental economics. These methods are well covered in Carson and Hanemann (2005). Even with these restrictions there is a considerable amount of material associated with the econometrics of stated preferences and an incomplete list of references that would complement this chapter includes Louviere et al. (2000), Train (2009), Ben-Akiva et al. (2016) and Lancsar et al. (2017).

Overview of Stated Preference Methods

Stated preference methods are used to elicit an individual's preferences for *alternatives* (goods, services, jobs) expressed in a survey context. They involve multiple dimensions that include logistics of data collection and questionnaire design underpinned by experimental design to define alternatives. All of this precedes, but should not be separate from, the ultimate analysis of the data and interpretation of the results. In contrast, the collection of RP data is typically divorced from the analysis stage and is the source of multiple data (modelling) problems. Griliches (1986) has argued that econometric methodology has evolved, in large part, to solve problems such as endogeneity and sample selection and to develop methods that extract meaningful inferences from non-experimental data. SP methods provide an opportunity to avoid many of these problems and in doing so better understand the behavior of economic agents that is often difficult with RP data.

In the case of DCEs, the survey questions are couched in terms of a realistic context that maps into the research question. Respondents are faced with a *choice set* of discrete and mutually exclusive alternatives defined in terms of *attributes*, and individuals are assumed to value these characteristics in coming to an evaluation of the alternative as a whole. Respondents are then required to answer one or more questions reflecting their evaluation of these alternatives. The same respondent then provides multiple outcomes for a sequence of different choice occasions or *scenarios* thus ensuring a cost-effective process of data collection. As an example, consider a representative choice scenario taken from Doiron et al. (2014) and displayed as Figure 1.

Scenario 1			
Features of Job	Job A	Job B	Job C
1. Location	Private hospital	Private hospital	Public hospital
2. Clinical rotations	Three	Three	None
3. Work hours	Part-time or fulltime	Fulltime only	Part-time or fulltime
4. Rostering	Flexible, usually accommodating requests	Inflexible, does not allow requests	Flexible, usually accommodating requests
5. Staffing levels	Usually well-staffed	Frequently short of staff	Usually well-staffed
6. Workplace culture	Supportive management and staff	Supportive management and staff	Unsupportive management and staff
7. Physical environment	Well equipped and maintained facility	Well equipped and maintained facility	Poorly equipped and maintained facility
8. Professional development and progression	Nurses encouraged	No encouragement for nurses	Nurses encouraged
9. Parking	Abundant and safe	Limited	Abundant and safe
10. Responsibility	Appropriate responsibility	Appropriate responsibility	Too much responsibility
11. Quality of care	Excellent	Poor	Poor
12. Salary	$1,250	$800	$1,100
Considering these three jobs:			
Q1. Which would you MOST like to get?	○ Job A	○ Job B	○ Job C
Q2. Which would you LEAST like to get?	○ Job A	○ Job B	○ Job C

Figure 1: Example of a scenario describing three alternative nursing jobs

The respondents in this survey are students or recent graduates from undergraduate nursing programs and the focus is on understanding preferences for attributes of nursing jobs. The choice set comprises three alternative jobs described in terms or attributes such

as salary and type of hospital. The *levels* of these attributes are then varied over scenarios according to an experimental design to provide different choice sets and facilitate efficient estimation. In each scenario respondents are required to first choose their most preferred alternative. In this survey they are then asked a second question requiring them to choose the worse of the two jobs remaining after their initial choice.

Some of the flexibility and opportunities one has in designing a SP survey can be illustrated by reference to this example. Including just the first question to determine the preferred alternative is possibly the most common way to generate choice outcomes and our discussion will concentrate on this case. The addition of the second question provides an example of what is called best-worst scaling that is becoming increasingly popular because of the extra preference information provided at low marginal cost; see for example Louviere et al. (2015). In this example, responses to these two questions together provide a complete ranking of the three alternatives.

The hypothetical alternatives in Figure 1 are fully described by their attributes and hence are denoted by generic titles, Job A, B and C. They are said to be unlabelled alternatives. Sometimes it is more appropriate to provide a descriptive name for the alternatives that constitute the choice set. For example, the choice could have been a choice between two jobs, one of which was always designated as private hospital and the other public hospital. Also, there could be an opt out option where respondents after the first question asking for their most preferred choice are asked whether they would actually choose that option. Or there could be a status quo option for respondents who already have a job, so this alternative would have attributes populated by the levels that describe that job and the investigator is determining which hypothetical alternative would be attractive enough to make respondents switch jobs.

Choices depend on the environment or context in which they are made. In designing a SP survey, the choice context plays a major role in making the hypothetical choice realistic. In the nurse's example, choosing an entry-level job in a hospital is a realistic and salient context for these respondents. Context can also be manipulated as part of the experimental design by defining different hypothetical environments in which the choice is to be made and then allocating respondents to these context treatments and by including context variables as attributes. For example, retention of nurses within the profession is a serious concern for policy makers and the study could have been extended by allocating respondents to context treatments that provided different information about foreshadowed government plans to change the working conditions of nurses.

While not included in Figure 1, respondent characteristics would also be collected as part of the survey. For example, the level of within hospital experience is likely to vary quite considerably between nursing students and early graduates and such experience may very well be one possible covariate that helps to explain variation in the relative valuation of job attributes across respondents. While valuable for analysis purposes, typical concerns about endogeneity of such covariates are less of a concern here because of the exogenous manipulation of the job attributes.

Ultimately Doiron et al. (2014) are aiming to draw policy implications from a better understanding of the heterogeneity of preferences for different job attributes. This is a case where some RP data would be available from say a survey of nurses. Here though it is unlikely that such data would include information on the choice set and instead would typically include attributes of just the chosen job. Even if information on jobs in the choice set could be obtained, it will often be the case that there is not sufficient variation in the important attributes to allow estimation of relevant preferences.

Layton and Levine (2003) is a case where use of SP data is advocated because no market data exists. They explore preferences for alternative climate change mitigation policies to reveal people's willingness to pay for measures to alleviate the impact of future climate change. But it need not be an either-or situation. Small et al. (2005) investigate the distribution of driver preference for reliable highway travel to inform road pricing policies using a combination of SP and RP data. This flexibility to use SP methods to provide information to compliment other data sources is a big part of the appeal of SP methods.

This initial overview has introduced several features of SP data that will eventually impact model specification and estimation. As a starting point consider a base case where over a sequence of scenarios, respondents choose a preferred option from a choice set containing two or more discrete and mutually exclusive alternatives.

Econometric Models for Choice Data

The Random Utility Model and MNL

The Random Utility Model (RUM) is the basis for model specification providing a framework within which to formulate families of probabilistic discrete choice models. Assume the utility that respondent i derives from choosing alternative j in choice scenario s is given by

(1) $U_{isj} = \boldsymbol{\beta}' \boldsymbol{x}_{isj} + \varepsilon_{isj}; i = 1, \cdots, N; s = 1, \cdots, S; j = 1, \cdots, J;$

where there are N respondents choosing amongst J alternatives across S scenarios. $\boldsymbol{x}_{isj}$ is a K-vector of observed attributes of alternative j faced by person i in scenario s, $\boldsymbol{\beta}$ is a conformable vector of utility weights, and ε_{isj} is the stochastic disturbance term representing characteristics unobservable by the analyst. $\boldsymbol{x}_{isj}$ could also include alternative specific constants (ASCs) and demographic characteristics of person i but for notational convenience these have not been explicitly included. The analyst also observes discrete choices, $y_{isj} = 1$ if i chooses alternative j in choice scenario s and zero otherwise.

The decision-maker chooses alternative j if it represents the highest utility in comparison with the utility associated with all other alternatives in the choice set. Thus, the probability of choosing alternative j is given by:

(2) $P_{isj} = Prob(y_{isj} = 1) = Prob(U_{isj} - U_{isl} > 0) \; \forall l \neq j.$

Econometric analysis now requires several specification issues to be resolved. Initially, consider multinomial logit (MNL) and its link to the RUM established by McFadden (2001). This remains a baseline for most extensions to more sophisticated models and for research on the theoretical underpinnings of decision-making in choice problems. MNL results from assuming the disturbance terms, ε_{isj}, are independently and identically distributed (*iid*) extreme values which leads to a computationally tractable model where the probability that individual i chooses j in scenario s is given by:

(3) $P_{isj} = \dfrac{exp(\lambda \boldsymbol{\beta}' \boldsymbol{x}_{isj})}{\sum_{l=1}^{J} exp(\lambda \boldsymbol{\beta}' \boldsymbol{x}_{isl})}$

where λ is the scale parameter that is inversely proportional to the standard deviation of the disturbance. In a standard MNL model, λ cannot be separately identified and is conventionally set to unity by assuming further that the disturbance terms are *iid* "type-I"

extreme values. Stating the presence of λ explicitly in (3) provides a useful basis for our subsequent discussion, nevertheless. To simplify notation later, it is also useful to write out the MNL likelihood of all observations on respondent i:

$$(4)\ L_i^{MNL}(\boldsymbol{\beta}) = \prod_{s=1}^{S} \prod_{j=1}^{J} \left[\frac{\exp(\boldsymbol{\beta}' \boldsymbol{x}_{isj})}{\sum_{l=1}^{J} \exp(\boldsymbol{\beta}' \boldsymbol{x}_{isl})} \right]^{y_{isj}}$$

which incorporates the conventional normalization of $\lambda = 1$.

While MNL is convenient, the *iid* extreme values assumption for the unobserved component of utility implies unrealistic substitution properties associated with the independence of irrelevant alternatives (IIA) and ignores the panel structure of the data. MNL also assumes homogenous tastes for the attributes of the alternatives which is not compatible with compelling evidence of pervasive heterogeneity in consumer tastes. Consequently, much recent research has been devoted to developing more flexible models that allow for taste heterogeneity.

Mixed Logit Models
Specifying a multinomial probit (MNP) model under an alternative assumption of multivariate normality for the random components of utility is one possible way to proceed. Computational demands have limited the use of this type of model and instead practitioners have preferred the heterogeneous or mixed logit (MIXL) family of models. Here the original specification in (1) is rewritten as:

$$(5)\ U_{isj} = \boldsymbol{\beta}_i' \boldsymbol{x}_{isj} + \varepsilon_{isj}; i = 1, \cdots, N; s = 1, \cdots, S; j = 1, \cdots, J;$$

which allows for unobserved individual specific deviations $\boldsymbol{\eta}_i$ around baseline utility weights $\boldsymbol{\beta}$ to produce individual specific utility weights $\boldsymbol{\beta}_i = \boldsymbol{\beta} + \boldsymbol{\eta}_i$. It is these random coefficients that capture taste heterogeneity and distinguish this approach from fixed coefficient specifications such as MNL. This form of heterogeneity is in addition to that captured by interactions between observables that are assumed to have already been included in $\boldsymbol{x}_{isj}$.

MIXL maintains the assumption that the ε_{isj} are distributed type-I extreme value. The model is completed by specifying the distribution for $\boldsymbol{\beta}_i$, called the mixing distribution. Part of the appeal of this class of models is that McFadden and Train (2000) show that by the appropriate choice of the mixing distribution one can approximate any random utility model. Their result is an existence proof that unfortunately does not help in the specific

selection of the mixing distribution. In most applications $\boldsymbol{\eta}_i$ is assumed to have a multivariate normal distribution, $\boldsymbol{\eta}_i \sim MVN(\mathbf{0}, \boldsymbol{\Sigma})$, and is denoted by N-MIXL.

Often what is of most interest is marginal willingness to pay (WTP), or more generally a ratio between marginal utility weights on two different attributes that measures the value of one attribute in terms of the other attribute. Suppose (5) is rewritten as:

(6) $U_{isj} = -\alpha_i p_{isj} + \boldsymbol{\theta}_i' \mathbf{z}_{isj} + \varepsilon_{isj}$

where p_{isj} is the price and $\mathbf{z}_{isj}$ contains the remaining elements of $\boldsymbol{x}_{isj}$. Under this "preference space" approach, utility weights α_i and $\boldsymbol{\theta}_i$ are core parameters and WTP is derived as $\boldsymbol{wtp}_i = \boldsymbol{\theta}_i / \alpha_i$. In contrast, the "WTP space" approach of Train and Weeks (2005) takes α_i and $\boldsymbol{wtp}_i$ as core parameters by re-parameterizing (6) as

(7) $U_{isj} = -\alpha_i p_{isj} + \alpha_i \boldsymbol{wtp}_i' \mathbf{z}_{isj} + \varepsilon_{isj}$

which allows the researcher to specify and estimate the joint distribution of $\boldsymbol{wtp}_i$ directly. The researcher should bear in mind that the same mixing distribution may produce substantively different MIXL models in the two spaces. For example, the multivariate normality of $\{\ln \alpha_i, \boldsymbol{\theta}_i\}$ is not equivalent to that of $\{\ln \alpha_i, \boldsymbol{wtp}_i\}$ since the ratio of a normal to a lognormal is not a normal.

As noted in discussion of (3), identification of these choice models requires a normalization of the scale parameter λ which is equivalent to multiplying (5) through by λ. But given the possibility that there is variation in tastes it seems logical to consider variation in scale or heteroskedasticity across individuals. Introducing a person-specific scale term into (5) yields:

(8) $U_{isj} = \lambda_i \boldsymbol{\beta}_i' \boldsymbol{x}_{isj} + \varepsilon_{isj}; i = 1, \cdots, N; s = 1, \cdots, S; j = 1, \cdots, J.$

In this form it is apparent that one possible explanation of the success of MIXL in fitting SP data is the presence of scale heterogeneity. Even in the absence of taste heterogeneity, variation in scale implies coefficient heterogeneity where the homogenous $\boldsymbol{\beta}$'s are either scaled up or down according to λ_i. This scale heterogeneity MNL (S-MNL) model is useful to consider because it represents a very parsimonious model of heterogeneity.

The generalized MNL (GMNL) of Fiebig et al. (2010) includes all the previously discussed models as special cases. It uses (5) but specifies the utility weights as:

(9) $\boldsymbol{\beta}_i = [\lambda_i\boldsymbol{\beta} + \gamma\boldsymbol{\eta}_i + (1-\gamma)\lambda_i\boldsymbol{\eta}_i]$

and λ_i is assumed to have a lognormal distribution, $ln(\lambda_i) \sim N(\bar{\lambda}, \tau^2)$. For identification the normalization $E(\lambda_i) = 1$ is applied and τ governs the extent of scale heterogeneity. The extra parameter γ does not appear in either S-MNL or N-MIXL, appearing only in the full GMNL. γ determines how the variance of the residual taste heterogeneity varies with scale when both appear in the model. For example, compare $\gamma = 1$ which implies $\boldsymbol{\beta}_i = [\lambda_i\boldsymbol{\beta} + \boldsymbol{\eta}_i]$ to $\gamma = 0$ implying $\boldsymbol{\beta}_i = \lambda_i[\boldsymbol{\beta} + \boldsymbol{\eta}_i]$.

While scale and preference heterogeneity are conceptually distinct concepts, the basic confound between them makes interpretation difficult. Specifically, finding improved fit from extending N-MIXL to allow for scale heterogeneity may simply be a reflection that the normal mixing distribution is inappropriate, and a more flexible distribution is needed. Hess and Train (2017) stress the importance of allowing for a full $\boldsymbol{\Sigma}$ matrix in a MIXL specification as scale heterogeneity induces correlation across parameters. Thus, N-MIXL with all random parameters specified as correlated can accommodate scale heterogeneity even though it is not explicitly specified. It does come at the cost of requiring many parameters to be estimated compared to a more parsimonious GMNL specification that allows the researcher to constrain the off-diagonal elements of $\boldsymbol{\Sigma}$ to 0s without losing the ability to account for scale heterogeneity.

MNL can be estimated by maximum likelihood, but the extensions introduced in the present section require simulation methods. Constructing a MIXL likelihood of observations on person i is conceptually straightforward since it involves evaluating the expected value of the MNL likelihood in (4) over the postulated distribution of $\boldsymbol{\beta}_i$: specifically, $E(L_i^{MNL}(\boldsymbol{\beta}_i)) = \int L_i^{MNL}(\boldsymbol{\beta}_i)f(\boldsymbol{\beta}_i)d\boldsymbol{\beta}_i$ where f(.) is the joint density of $\boldsymbol{\beta}_i$ given the postulated distribution. This multiple integral does not have a closed-form solution, but can be approximated by simulation. The simulated log-likelihood function for a sample of N individuals is given by

(10) $SLL = \sum_{i=1}^{N} ln\left(\frac{1}{R}\sum_{r=1}^{R} L_i^{MNL}(\boldsymbol{\beta}_i^r)\right)$

where $\boldsymbol{\beta}_i^r$ is the r-th draw for individual i from the distribution of $\boldsymbol{\beta}_i$, and the mean of R such draws inside ln(.) is the simulated counterpart of $E(L_i^{MNL}(\boldsymbol{\beta}_i))$. $\boldsymbol{\beta}_i^r$ is a combination of random components that vary from draw to draw and estimated parameters that remain constant (e.g. $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ in N-MIXL), but this combination takes a different form in each

MIXL model. The estimator obtained by maximizing SLL is known as the maximum simulated likelihood (MSL) estimator. See Train (2009) for further details.

Latent Class Models

While N-MIXL and GMNL use continuous distributions to capture population heterogeneity in parameters, the MIXL framework can accommodate discrete distributions as well. The most well-known example is the Latent Class Logit Model (LCL) of Kamakura and Russell (1989) which uses (5), and postulates that the random coefficient vector $\boldsymbol{\beta}_i$ follows a discrete distribution with C support points such that $\boldsymbol{\beta}_i \in \{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_C\}$ and $\Pr(\boldsymbol{\beta}_i = \boldsymbol{\beta}_c) = \pi_c$ for each $c = 1,2, \dots, C$. To put it simply, LCL assumes that each respondent belongs to one of C preference classes and each class c makes up fraction π_c of the population. Then, the likelihood of observations on respondent i can be computed by mixing the MNL likelihood in (4) over the discrete distribution, resulting in the sample log-likelihood function

$$(11)\ LL^{LCL}(\{\boldsymbol{\beta}_c\}_{c=1}^{C}, \{\pi_c\}_{c=1}^{C-1}) = \sum_{i=1}^{N} ln\left(\sum_{c=1}^{C} \pi_c \times L_i^{MNL}(\boldsymbol{\beta}_c)\right)$$

where C preference vectors and $C-1$ shares are parameters to be estimated, and the share of the last class, π_C, is constrained to satisfy the adding-up restriction $\sum_{c=1}^{C} \pi_c = 1$. To estimate LCL, one must pre-specify the total number of preference classes C. In the empirical literature, it is common practice to estimate LCL repeatedly for alternative values of C, and focus subsequent reporting and discussion on the results for an "optimal" value of C that leads to the best fit in terms of the Bayesian Information Criterion or the Akaike Information Criterion.

LCL does not require simulation-based methods because (11) is a closed-form expression. Moreover, the maximum likelihood estimator of LCL is invariant to re-parameterizations of $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \dots, \boldsymbol{\beta}_C$, such as that of MNL. Among other things, this means that the WTP derived from LCL in the preference space must be identical to the WTP estimated directly by specifying LCL in the WTP space.

How does LCL handle correlated tastes and scale heterogeneity? Note that there is nothing in the LCL structure that dictates how different or similar preference vector β_c for one class should be relative to another. In other words, LCL allows for any pattern of correlations among random parameters, including one that is observationally equivalent to scale heterogeneity. For instance, if the population comprises two classes and scale heterogeneity is the only form of heterogeneity present, LCL can easily capture this using

two vectors $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ that satisfy $\boldsymbol{\beta}_2 = q\boldsymbol{\beta}_1$ for some positive scalar q. LCL is therefore not a model that assumes away scale heterogeneity; it is a model that postulates discrete heterogeneity in composite random parameters $\lambda_i \boldsymbol{\beta}_i$. The Scale Adjusted Latent Class (SALC) Model of Magidson and Vermunt (2007) specifies a discrete mixture analogue to GMNL more explicitly, under the assumption that respondent i simultaneously belongs to one of C preference classes and one of S scale classes. While LCL does allow for scale heterogeneity, SALC may be useful to the extent that adding scale parameters is a more parsimonious way to account for scale heterogeneity than adding extra K-dimensional preference vectors.

The LCL log-likelihood function in (11) has lent itself to several well-known variants and extensions. Ben-Akiva and Boccara (1995) use the LCL framework to model the notion that different respondents may consider different subsets of J available alternatives for final choices. Their model operationalizes this heterogeneity in "consideration sets" by specifying each preference class to have an MNL likelihood function over a distinct subset of alternatives. The Endogenous Segmentation Model of Bhat (1997) allows population shares π_c to vary with the observed characteristics of respondent i, by placing a fractional MNL structure on the shares. Note that while this approach appears seemingly more general, in a finite sample the resulting model may not nest one's preferred LCL as a special case: adding a fractional MNL structure to an "optimal" LCL specification often leads to an empirically underidentified model, compelling the researcher to reduce the number of preference classes from what is "optimal" for LCL. Scarpa et al. (2009) use the LCL framework to model "attribute non-attendance", the notion that different respondents may attend to different subsets of product attributes. In their model, each preference class is assumed to ignore a distinct subset of attributes, and the corresponding elements in their preference vector $\boldsymbol{\beta}_c$ are constrained to 0s. Finally, Train (2008) estimates a hybrid model that combines LCL with N-MIXL by postulating that each preference class is a subpopulation of respondents who draw their preference vectors $\boldsymbol{\beta}_i$ from a multivariate normal distribution specific to that class.

Models for Ranked Data

General Models

As seen earlier in Figure 1, the SP survey may ask the respondent to state their preference ranking of alternatives in a choice set, instead of asking what they would like to choose from a choice set. What follows is a review of models that one may consider in the econometric analysis of ranked data. For brevity, the focus is on baseline specifications that do not address unobserved population heterogeneity. Extending each specification to accommodate population heterogeneity is straightforward and involves the same set of procedures as described in the previous sections on Mixed Logit Models and Latent Class

Models. In fact, the random parameter extensions of all models reviewed here have already appeared in the literature; see for example Yoo and Doiron (2013), Doiron et al. (2014) and Oviedo and Yoo (2017).

Many economists may find it natural to proceed with the RUM in (1) as a behavioral foundation and formulate models for ranked data by equating the stated preference ranking with the latent ranking of utilities. Indeed, this is the approach that Beggs et al. (1981) take to derive the Rank-Ordered Logit (ROL) Model, which results from assuming that the disturbances ε_{isj} are *iid* extreme values, just as under MNL. To facilitate discussion, suppose that respondent i faced $J = 4$ alternatives in scenario s, and ranked alternatives $4, 1, 3,$ and 2 as best, second-best, third-best and worst in that order. The ROL probability of this rank-ordering can be derived as $P_{is\{4,1,3,2\}} = Prob(U_{is4} > U_{is1} > U_{is3} > U_{is2})$, and is given by

$$(12) \quad P_{is\{4,1,3,2\}} = \frac{exp(\lambda\boldsymbol{\beta}'\boldsymbol{x}_{is4})}{[\sum_{l=1}^{4} exp(\lambda\boldsymbol{\beta}'\boldsymbol{x}_{isl})]} \times \frac{exp(\lambda\boldsymbol{\beta}'\boldsymbol{x}_{is1})}{[\sum_{l=1}^{3} exp(\lambda\boldsymbol{\beta}'\boldsymbol{x}_{isl})]} \times \frac{exp(\lambda\boldsymbol{\beta}'\boldsymbol{x}_{is3})}{[\sum_{l=2}^{3} exp(\lambda\boldsymbol{\beta}'\boldsymbol{x}_{isl})]}$$

where λ is conventionally set to unity to achieve identification. Assuming the disturbances follow a multivariate normal distribution produces the Rank-Ordered Probit (ROP) Model (Train, 2009, p.158), and a generalized extreme value distribution produces the Nested Rank-Ordered Logit (NROL) Model (Dagsvik and Liu, 2009). Each rank-ordered model directly inherits all the strengths and weaknesses of the corresponding choice model. For instance, ROL is like MNL in that it has a tractable functional form but exhibits the independence of irrelevant alternatives (IIA) property, whereas ROP is like MNP in that it may allow for more flexible substitution patterns but requires computer-intensive methods for estimation.

The product structure of ROL intuitively illustrates the primary benefit of using rank-ordered data relative to choice data, though it must be stressed that this structure is a unique feature of ROL and *is not* shared by other models. Specifically, rank-ordered data provide more information on latent dependent variables (such as $U_{is4} > U_{is1} > U_{is3} > U_{is2}$) than choice data (from which one can only learn the like of $U_{is4} > max\{U_{is1}, U_{is2}, U_{is3}\}$), allowing the researcher to estimate a RUM of interest more efficiently. The ROL probability in (12) is a product of MNL probabilities in (3), making the source of efficiency gain easy to see: a single rank-ordered observation contributes to the sample log-likelihood the same amount of information as several choice observations (in this case three) on progressively smaller choice sets. The product structure of ROL is an implication of the IIA property, however, and does not generalize

to any other rank-ordered choice model. For example, the ROP probability is not a product of MNP probabilities, and the NROL probability is not a product of nested logit probabilities. It nevertheless remains true that ROP and NROL allow the researcher to estimate the RUM of interest more efficiently than their multinomial choice counterparts.

The product structure also implies that ROL (again, neither ROP nor NROL) can be formulated by taking a fundamentally different approach. Instead of modelling the stated ranking as a latent utility ordering, the researcher may model it directly as a particular sequence of choices. Under this approach, continuing with the $J = 4$ example, the first choice is over all four alternatives, the second choice is over three alternatives except the first choice, and the third choice is over two alternatives excluding the first and second choices; the example extends to other cases in an obvious manner. ROL then results from assuming that each choice is independent of another and generated from MNL. Hausman and Ruud (1987) take this approach to formulate the Heteroskedastic ROL (HROL) Model that allows scale parameter λ to vary across decision stages in the choice sequence, and Ben-Akiva et al. (1992) generalize HROL further by allowing a subset of utility weights to vary across the decision stages. Since Chapman and Staelin (1982), the sequential choice interpretation of ROL has sustained the notion that in cases where the researcher finds a discrepancy between MNL on first choices and ROL, the researcher must consider that as a symptom of unreliable rankings data and focus on the MNL estimation results. While the present section is not intended as a critical review of the empirical literature, we note that if the disturbances are not *iid* extreme values, both MNL and ROL estimators are inconsistent and there is no reason why they must lead to similar estimates.

Fully ranking many alternatives from best to worst may be a task that most respondents find difficult, and SP surveys may be designed to elicit an incomplete ranking instead. For example, Layton and Levine (2003) ask the respondent to identify the best and worst out of five alternatives. When using ROL, it is easy to handle incomplete rankings where preferences are observed only up to the Cth best; the researcher simply needs to retain the first C MNL probabilities in the full ROL formula. For other types of models and incomplete rankings, the researcher may still formulate suitable econometric specifications by using the RUM in (1) to derive the probability of an incomplete ranking. But the resulting probability would become more cumbersome to evaluate than the corresponding probability of a complete ranking, because of the need to consider all possible permutations of missing preference orderings explicitly; see Vann Ophem et al. (1999) and Layton and Levine (2003) for examples in the context of ROL and ROP respectively.

Profile Case Best-Worst Scaling

In recent years, an incomplete ranking task that asks the respondent to identify the best and worst *attributes* of an alternative, instead of the best and worst alternatives in a choice set, has become increasingly popular. Figure 2, taken from Yoo and Doiron (2013), shows an example of this type of "profile case" Best Worst Scaling (BWS) task which originates from the same survey as the "multi-profile case" BWS task in Figure 1. Now, instead of identifying the best and worst of three profiles or alternatives, the respondent evaluates one profile and state the best and worst out of its twelve characteristics. Louviere et al. (2015) provide a book-length treatment of the design and econometric analysis of BWS tasks.

		Best Aspect	Worst Aspect
1. Location:	Private hospital	○	○
2. Clinical rotations:	Three	○	○
3. Work hours:	Fulltime only	○	○
4. Rostering:	Flexible, usually accommodating requests	○	○
5. Staffing levels:	Usually well-staffed	○	○
6. Workplace culture:	Unsupportive management and staff	○	○
7. Physical environment:	Poorly equipped and maintained facility	○	○
8. Professional development and progression:	No encouragement for nurses	○	○
9. Parking (The parking facilities):	Limited	○	○
10. Responsibility:	Appropriate responsibility	○	○
11. Quality of care:	Poor	○	○
12. Weekly Salary:	$950	○	○

Figure 2: Example of a scenario describing profile case Best Worst Scaling

While one may analyse profile case BWS using general econometric models for incomplete rankings, by far the most popular method is a purpose-built variant of MNL known as the Maximum-Difference (Max-Diff) Model (Marley and Louviere, 2005). When a profile case BWS respondent identifies the best and worst of K attributes, Max-Diff postulates that the respondent would evaluate $K \times (K-1)$ options where each option is a potential pair of best and worst attributes, and choose the option that maximizes their utility; in the context of Figure 2, {Best = Private Hospital, Worst = $950} is one option, and so are {Best = $950, Worst = Private Hospital} and other permutations of the job aspects. The Max-Diff probability takes a MNL functional form defined over such $K \times (K-1)$ options, where the index for each option measures the utility difference between the best attribute and the worst attribute in that pair; for example, options {Best = Private Hospital, Worst = fulltime only} and {Best = fulltime only, Worst = Private Hospital} would have index values of $(\theta^{private\ hospital} - \theta^{fulltime\ only})$ and $(\theta^{fulltime\ only} - \theta^{private\ hospital})$ respectively, where θ^k measures utility from attribute k and is a parameter to be estimated. The name Max-Diff originates from the assumption that the respondent would choose a pair that maximizes the utility difference between the two component attributes. Eliciting choices over attributes directly

has interesting implications for econometric identification and interpretation of utility parameters. For example, profile case BWS allows one to infer whether job aspect "private hospital" is preferred to another job aspect "fulltime only", whereas the SP survey eliciting choices over alternatives allows one to infer only whether changing one job aspect is preferred to changing another job aspect. See Yoo and Doiron (2013) for fuller comparisons.

Computing Strategies for Mixed Logit
Alternatives to Maximum Simulated Likelihood
In principle, once the (simulated) log-likelihood function of a MIXL model has been programmed, estimation of the model can proceed in the usual manner using any popular variant of gradient-based numerical optimization or "hill-climbing" techniques. In practice, however, researchers are likely to face at least two types of computational challenges. First, unless "good" starting values are selected, the numerical optimizer may terminate before finding a maximum, a situation that is often casually described as "the model failed to converge". Unlike MNL, MIXL has a non-concave log-likelihood function which may sometimes display several nearly flat surfaces, making it difficult for the optimizer to see which hill to climb. Second, the numerical optimization process tends to progress rather slowly, and even in the modern computing environment new users of MIXL will quickly become accustomed to waiting for several hours, if not days, before seeing their estimation results. For models like N-MIXL and GMNL that require simulation methods, the source of the computational demand is apparent. While LCL results in a closed-form expression, estimating a *C*-class LCL specification is much more computationally demanding than estimating MNL *C* times because the dimension of the (quasi-)Hessian matrix increases quadratically in the number of model parameters.

For proportional hazard models with discrete heterogeneity, Heckman and Singer (1984) have popularized the use of a fast and numerically stable computing strategy known as the expectation-maximization (EM) algorithms. Bhat (1997) develops a suitable version of the EM algorithm for LCL. The intuition behind this strategy is straightforward. Consider a fictional situation where the researcher can observe class membership dummies $\{d_{ic}\}_{c=1}^{C}$ directly alongside choices and relevant regressors, where $d_{ic} = 1$ if respondent i belongs to preference class c and 0 otherwise. Then MLE of each class share π_c is simply the sample mean of d_{ic} across all respondents, and MLE of preference vector $\boldsymbol{\beta}_c$ can be obtained from a MNL regression for respondents whose $d_{ic} = 1$. More formally, the sample log-likelihood of observing the class dummies and choices in this fictional case is given by

$$(13) \quad \kappa(\{\boldsymbol{\beta}_c\}_{c=1}^{C}, \{\pi_c\}_{c=1}^{C-1}) = \left(\sum_{i=1}^{N}\sum_{c=1}^{C} d_{ic} \, ln \, \pi_c\right) + \left(\sum_{i=1}^{N}\sum_{c=1}^{C} d_{ic} \, ln \, L_i^{MNL}(\boldsymbol{\beta}_c)\right)$$

where we use the same notation as earlier. Of course the fictional construct $\{d_{ic}\}_{c=1}^{C}$ is always missing in the real world, and Bhat's EM algorithm is operationalized by replacing each d_{ic} with its expected value conditional on choices, or "posterior class shares", which takes a tractable functional form of $w_{ic} = E(d_{ic}|choices) = \pi_c \times L_i^{MNL}(\boldsymbol{\beta}_c)/(\sum_{s=1}^{C}\pi_s \times L_i^{MNL}(\boldsymbol{\beta}_s))$. More specifically, at iteration t, the researcher evaluates w_{ic} using the estimates of $\{\boldsymbol{\beta}_c\}_{c=1}^{C}$ and $\{\pi_c\}_{c=1}^{C-1}$ obtained from the preceding iteration $t-1$; then, the researcher obtains updates to each π_c by taking the sample mean of the resulting w_{ic} across respondents, and $\boldsymbol{\beta}_c$ by running a weighted MNL regression where each respondent's choice observations are weighted by their own w_{ic}. Interestingly, repeating this simple procedure until the estimates do not change between iterations produces the estimates that maximize the usual sample log-likelihood in (11).

Train (2008) generalizes the EM algorithms to MIXL models incorporating continuous mixing distributions as well as discrete mixture of continuous distributions. At least for N-MIXL, the intuition behind this newer strategy is clear, though it is worthwhile emphasizing from the outset that the resulting Method of Simulated Moment (MSM) estimator is not identical to the MSL estimator when the number of simulated draws is fixed; in this respect, the present procedure contrasts with the EM algorithm for LCL that directly results in MLE. Now consider a fictional situation where the researcher can observe each respondent's $\boldsymbol{\beta}_i$, alongside their choices and relevant regressors. Then, MLE of the mean $\boldsymbol{\beta}$ and covariance $\boldsymbol{\Sigma}$ for the population multivariate-normal density are simply the sample mean and covariance of $\boldsymbol{\beta}_i$. The sample log-likelihood of observing $\boldsymbol{\beta}_i$ and choices is then given by

$$(14) \quad \kappa(\boldsymbol{\beta}, \boldsymbol{\Sigma}) = \left(\sum_{i=1}^{N} ln \, MVN(\boldsymbol{\beta}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma})\right) + \left(\sum_{i=1}^{N} ln \, L_i^{MNL}(\boldsymbol{\beta}_i)\right)$$

which shows why estimating $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ does not require estimating any MNL model: the second term does not depend on $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$. In practice, operationalizing this strategy requires replacing $\boldsymbol{\beta}_i$ with simulated draws $\{\boldsymbol{\beta}_i^r\}_{r=1}^{R}$, and weighting each draw $\boldsymbol{\beta}_i^r$ by its simulated "posterior density" $h_i^r = L_i^{MNL}(\boldsymbol{\beta}_i^r)/(\sum_{m=1}^{R} L_i^{MNL}(\boldsymbol{\beta}_i^m))$. Specifically, at iteration t, the researcher generates $\{\boldsymbol{\beta}_i^r\}_{r=1}^{R}$ using $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ obtained at iteration $t-1$, and evaluates $\{h_i^r\}_{r=1}^{R}$ using those draws; then, the researcher obtains updates to $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ by computing the weighted sample mean and covariance of those draws, using $\{h_i^r\}_{r=1}^{R}$ as

weights. This recursive procedure continues until the estimates of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ do not change between iterations. While this procedure still requires simulation of the likelihood function that appears in the denominator of $\{h_i^r\}_{r=1}^R$, it is a much simpler task than MSL that also requires computation of the gradient and (quasi-)Hessian of the simulated log-likelihood function with respect to $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$.

When it comes to continuous mixture models like N-MIXL, the Hierarchical Bayesian (HB) procedure of Train (2001) is a popular alternative to EM. The name originates from approaching estimation of MIXL as a task that involves placing a prior distribution on a prior distribution. Continuing with the N-MIXL example, at the top-level, the researcher places a prior distribution on unknown parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ of population density $MVN(\boldsymbol{\beta}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma})$. Then, at the next level, the researcher uses density $MVN(\boldsymbol{\beta}_i; \boldsymbol{\beta}, \boldsymbol{\Sigma})$ as a prior distribution on unknown parameters $\boldsymbol{\beta}_i$ in the MNL likelihood $L_i^{MNL}(\boldsymbol{\beta}_i)$. The HB procedure makes it easy to apply Gibbs sampling to simulate the joint posterior distribution of $\boldsymbol{\beta}$, $\boldsymbol{\Sigma}$, and $\boldsymbol{\beta}_i$. While an adequate summary of the technical details cannot be provided here, note that the procedure offers computational advantages that are similar to EM; generating draws of $\boldsymbol{\beta}_i$ from its posterior distribution is a much simpler simulation exercise than what MSL demands, and updating the posteriors of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ involve basic algebraic operations on those draws. As Train (2001) stresses, the researcher may choose to use the HB procedure as an alternative computing strategy to obtain MSL estimates of MIXL models, without adopting it as a method of drawing Bayesian inferences; the posterior mean of $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}$ are asymptotically equivalent to the MSL estimator, though in a finite sample they may not be identical.

A Comparison of Methods
What explains the continued popularity of ML and MSL estimation of MIXL models, when these faster computing strategies are available? One possible factor is adaptability. Within the ML and MSL framework, when the researcher plans to incorporate a new mixing distribution, they only need to reprogram the (simulated) log-likelihood function; the extra step of coding algebraic derivatives is optional. For the EM algorithms and the HB procedure, however, incorporating a new mixing distribution may require more substantive changes to the recursive steps. A second and possibly more important factor is that, ironically, the faster computing strategies tend to slow down and may even run slower than ML and MSL as soon as the researcher simplifies the model specification by constraining certain parameters to be non-random. For example, consider estimation of LCL using equation (13). The presence of just one parameter that is common to all classes immediately implies that one cannot break down the second term into *C* separate MNL models, and maximizing this term would require joint estimation of all class-specific parameters as well as the common parameter. Train (2009, p.308) reports an

illustrative N-MIXL application that involves 5 random parameters and 1 non-random parameter; in that case MSL finds a solution almost three times faster than the HB procedure. This kind of slowdown is a major drawback considering that the researcher may often want to estimate MIXL specifications with some non-random parameters, for example to distinguish observed heterogeneity in preferences from unobserved heterogeneity.

The non-concave log-likelihood functions of MIXL models may exhibit not only flat surfaces, but also several local maxima. Unless "good" starting values are selected, even when the numerical optimizer produces a solution, there is no reason why the solution should be a global maximum. Despite the textbook emphasis on the importance of checking the sensitivity of results to alternative starting values, most empirical studies using MIXL, including many of our own, rarely report which starting values have been used and explored over. The folklore suggests that most practitioners use estimation results for special cases of a final model as starting values for the final model. As an alternative to this conventional strategy, Hole and Yoo (2017) explore the use of population-based optimization heuristics to conduct a more global search for "good" starting values. The findings suggest that the heuristics-based strategy may locate better maxima than the conventional strategy, even in those instances where multiple special cases of a final model lead to an identical and hence seemingly global maximum.

Econometric Analysis with Multiple Data Sources
Combining SP Data
SP methods are especially amenable to various research design strategies that involve combining data. If the researcher conducts a SP survey with multiple samples drawn from the same population or similar populations, would they obtain comparable findings across the samples? Several studies have addressed this type of external validity question explicitly. For example, Capparos et al. (2008) investigate the robustness of preferences across data on choices and data on the best alternative from a ranking exercise, by randomizing respondents to either a choice format or a ranking format of an identical survey. Hall et al. (2006) compare the preferences for genetic tests for a general population sample and a sample where the choices are more salient. Fiebig et al. (2009) consider joint decision-making and explore differences in the preferences of women making choices and providers making recommendations in relation to cervical screening. Doiron and Yoo (2017) test the temporal stability of preferences by administering the SP survey in Figure 1 to the same group of respondents twice over a span of about 12 months. If data collection occurs before and after a policy intervention, then repeated SP surveys can be used as a method of policy evaluation such as in Johar et al. (2013).

Even within the MNL framework, testing the stability of preferences is not a simple matter of testing whether utility parameters $\boldsymbol{\beta}^A$ and $\boldsymbol{\beta}^B$ are identical across data A and data B. The confounding factor is a possible shift in the scale; even when the underlying utility parameters are identical at $\boldsymbol{\beta}$, the identified parameters $\boldsymbol{\beta}^A = \lambda^A \boldsymbol{\beta}$ and $\boldsymbol{\beta}^B = \lambda^B \boldsymbol{\beta}$ may diverge in case there is more behavioral noise in one data relative to the other. In the context of Doiron and Yoo (2017), for example, one may conjecture that there would be less behavioural noise in the repeat survey since having participated in the initial survey may make the respondents more familiar with the choice task. A popular way to allow for possible scale differences in evaluation of preference stability is to focus on WTP parameters, which are not affected by variations in the scale. But such tests on WTP must be applied with caution because a difference in the marginal utility of money (in the denominator) could induce the WTP parameters to diverge even when utilities associated with all other attributes (in the numerator) are identical. Indeed, Doiron and Yoo (2017) find more evidence of stability in direct comparisons of utility parameters $\boldsymbol{\beta}^A$ and $\boldsymbol{\beta}^B$ than in comparisons of WTP, due to a substantial temporal variation in the marginal utility of money. As far as using parameter ratios to cancel out the scale factor goes, any one of utility parameters can be used as the common denominator of the ratios, and sensitivity checks across alternative choices of the denominator utility would be appropriate.

Combined SP-RP Data

In most cases the primary motivation for collecting SP data is the lack of suitable RP data, but in applications involving consumer goods the researcher may have access to both sources of data. Two well-known examples are Ben-Akiva and Morikawa (1990) who analyse transport mode choices by commuters, and Brownstone et al. (2000) who analyse private vehicle choices by households, using data collected from both SP and RP surveys. Even in these instances, however, the type and detail of information on product attributes are likely to vary across the two surveys, arguably for good reasons; constraining the flexibility, or increasing the complexity, of a SP design in an effort to create a replica of a RP choice context defeats the purpose of collecting SP data. The SP survey by Brownstone et al. (2000), for example, incorporates alternative fuel vehicles that are not available on the market but does not include all varieties of conventional vehicle models that are available on the market.

In an econometric analysis of SP-RP data, an important consideration is therefore how best to exploit differences between the two data sources to the researcher's advantage. An oft-cited principle is to use the large amount of independent variation in product attributes in the SP data to improve the statistical precision of utility parameter estimates associated

with those attributes, and to use the actual market settings of the RP data to estimate alternative-specific constants that equate the predicted market shares of existing products with their observed market shares (Train, 2009, Ch 7). Obviously, before allowing the SP and RP components of the implied joint model to share the utility parameters, the researcher should allow for a potential shift in the scale between the two data sources. The scale variation is arguably a bigger issue in the present context than in the comparison of two SP data sets, because the composition of unobservables affecting RP choices can be fundamentally different from those affecting SP choices: for example, consider the influence of the "word of mouth" in the RP settings.

The basic template for the joint RUM of RP-SP data may be summarized as follows. Suppose that explanatory variables are categorized into $\boldsymbol{x}_{isj}^{Shared}$, $\boldsymbol{x}_{isj}^{SP}$ and $\boldsymbol{x}_{isj}^{RP}$, where $\boldsymbol{x}_{isj}^{Shared}$ have identical utility parameters between the two data sources subject to the scale difference, and $\boldsymbol{x}_{isj}^{SP}$ and $\boldsymbol{x}_{isj}^{RP}$ have distinct utility parameters specific to the superscripted data sources. For instance, $\boldsymbol{x}_{isj}^{Shared}$ may include product attributes that are observed in both settings, whereas $\boldsymbol{x}_{isj}^{SP}$ and $\boldsymbol{x}_{isj}^{RP}$ may include alternative-specific constants. Then the econometric model may be formulated from a system of RUM equations

$$(15)\ U_{isj}^{SP} = \boldsymbol{\delta}^{SP\prime}\boldsymbol{x}_{isj}^{SP} + \boldsymbol{\beta}^{SP\prime}\boldsymbol{x}_{isj}^{Shared} + \varepsilon_{isj}^{SP}$$

$$(16)\ U_{isj}^{RP} = \boldsymbol{\delta}^{RP\prime}\boldsymbol{x}_{isj}^{RP} + \lambda^{RP}\boldsymbol{\beta}^{SP\prime}\boldsymbol{x}_{isj}^{Shared} + \varepsilon_{isj}^{RP}$$

where the scale of the SP equation is normalized to 1 so that λ^{RP} measures the scale of the RP equation relative to the SP equation. As long as λ^{RP} is a non-random parameter, whether $\boldsymbol{\delta}^{RP}$ is subjected to scaling or not is a matter of re-parameterization that does not affect substantive results. As usual, assuming the *iid* type-I extreme value disturbances lead to MNL (Ben-Akiva and Morikwa, 1990) and modelling the utility parameters as multivariate-normal random coefficients lead to N-MIXL (Brownstone et al., 2000; Small et al., 2005). In practice, partitioning explanatory variables into $\boldsymbol{x}_{isj}^{Shared}$, $\boldsymbol{x}_{isj}^{SP}$ and $\boldsymbol{x}_{isj}^{RP}$ is informed as much by a pre-analysis based on separate models for SP and RP data as by a survey design. For instance, in case the SP and RP coefficients on a shared attribute does not take the same sign in the separate models, ascribing their difference to a shift in the scale would be inappropriate.

Conclusion

Modelling and understanding heterogeneity in how people make choices remains an active area of research. In these endeavours, stated preference methods are popular providing a cost-effective means of generating data that can address such questions, and in some cases, questions which are not amenable to analysis using conventional data sources. Their potential to be even more useful is likely to lie in clever research design strategies. In addition to the data combination examples discussed previously in this chapter, SP methods can be used in conjunction with conventional data collections; see Joyce et al. (2010) where DCEs have been imbedded in a longitudinal survey of doctors and King et al. (2007) where patient preferences for preventive asthma medications preferences were elicited using a DCE embedded in a randomized clinical trial. The econometric methodology for the appropriate analysis of such data has been reviewed but this is likely to be a fruitful area of future research as new challenges in combining data arise.

References

Beggs, S., Cardell, S., and Hausman, J. (1981). "Assessing the potential demand for electric cars", *Journal of Econometrics* 16, 1-19.

Ben-Akiva, M. and Boccara, B. (1995), "Discrete choice models with latent choice sets", *International Journal of Research in Marketing*, 12, 9-24.

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

Ben-Akiva, M., Morikawa, T., and Shiroishi, F. (1992), "Analysis of the reliability of preference ranking data", *Journal of Business Research* 24, 149-164.

Ben-Akiva, M., McFadden, D. and Train, K. (2016), "Foundations of stated preference elicitation: Consumer behavior and choice-based conjoint analysis", University of California, Berkeley, http://eml.berkeley.edu/train/foundations.pdf.

Bhat, C. R. (1997), "An endogenous segmentation mode choice model with an application to intercity travel", *Transportation Science*, 31, 34-48.

Brownstone, D., Bunch, D., and Train, K. (2000), "Joint mixed logit models of stated and revealed preferences for alternative-fuel vehicles", *Transportation Research Part B* 34, 315–338.

Caparros, A., Oviedo, J. L., and Campos, P. (2008), "Would you choose your preferred option? Comparing choice and recoded ranking experiments", *American Journal of Agricultural Economics* 90, 843-855.

Carson, R.T. and Hanemann, W.M. (2005), "Contingent valuation", in *Handbook of Environmental Economics*, Volume 2, Edited by K.-G Maler, and J.R. Vincent, Elsevier B.V. Chapter 17, pp 821-934.

Chapman, R. G. and Staelin R. (1982), "Exploiting rank ordered choice set data within the stochastic utility model", *Journal of Marketing Research* 19, 288-301.

Dagsvik, J. K. and Liu, G. (2009), "A framework for analyzing rank ordered data with application to automobile demand", *Transportation Research Part A* 43, 1-12.

Doiron, D. Hall, J., Kenny, P. and Street, D.J, (2014), "Job preferences of students and new graduates in nursing", *Applied Economics*, 46:9, 924-939.

Doiron, D. and Yoo, H. I. (2017), "Temporal stability of stated preferences: the case of junior nursing jobs", *Health Economics* 26, 802-809.

Fiebig, D.G., Haas, M., Hossain, I., Street, D.J. and Viney, R. (2009), "Decisions about Pap tests: What influences women and providers", *Social Science and Medicine* 68, 1766-1774.

Fiebig, D.G., Keane, M.P., Louviere, J.J. and Wasi, N. (2010), "The generalized multinomial logit model: Accounting for scale and coefficient heterogeneity", *Marketing Science* 29, 393-421.

Griliches, Z. (1986), "Economic data issues" Ch. 25 of Z. Griliches and M.D. Intriligator eds. *Handbook of Econometrics*, North Holland.

Groves, R.M., Fowler, F.J., Couper, M.P., Lepkowski, J.M. Singer, E. and Tourangeau, R. (2009), *Survey Methodology*, New Jersey: Wiley.

Hall, J.P., King, M.T., Fiebig, D.G., Hossain, I., and Louviere (2006), "What influences participation in genetic carrier testing? Results from a discrete choice experiment", *Journal of Health Economics*, 25, 520-537.

Hausman, J. A. and Ruud, P. A. (1987), "Specifying and testing econometric models for rank-ordered data", *Journal of Econometrics* 34, 83-104.

Heckman, J. and Singer, B. (1984), "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data", *Econometrica* 52, pp.271-320.

Hess, S. and Train, K. (2017), "Correlation and scale in mixed logit models", *Journal of Choice Modelling* 23, 1-8.

Hole, A. R. and Yoo, H. I. (2017) "The use of heuristic optimization algorithms to facilitate the maximum simulated likelihood estimation of random parameter logit models", *Journal of the Royal Statistical Society: Series C* 66, 997-1013.

Johar, M., Fiebig, D.G., Haas, M. and Viney, R. (2013), "Using repeated choice experiments to evaluate the impact of policy changes on cervical screening", *Applied Economics* 45 (14), 1845-1855.

Joyce, C., Scott, A., Jeon, S., Humphreys, J., Kalb, G., Witt, J. and Leahy, A. (2010), "The "Medicine in Australia: Balancing Employment and Life (MABEL)" longitudinal survey - Protocol and baseline data for a prospective cohort study of Australian doctors' workforce participation", *BMC Health Services Research*, 10-50.

Kamakura, W. A. and Russell, G. (1989), "A probabilistic choice model for market segmentation and elasticity structure", *Journal of Marketing Research* 26, 379–390.

Kesternich, I., Heiss, F., McFadden, D. and Winter, J. (2013), "Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D", *Journal of Health Economics*, 32, 1313-24.

King, M.T., Hall, J.P., Lancsar, E.J., Fiebig, D.G., Hossain, I., Louviere, J.J., Reddell, H.K. and Jenkins, C.R. (2007), "Patient preferences for managing asthma: Results from a discrete choice experiment", *Health Economics*, 16, 703-717.

Lancsar, E., Fiebig, D.G. and Hole, A.R. (2017), "Discrete choice experiments: A guide to model specification, estimation and software", *PharmacoEconomics*, 35(7), 697-716.

Layton, D. F. and Levine, R. A. (2003), "How much does the far future matter? A hierarchical Bayesian analysis of the public's willingness to mitigate ecological impacts of climate change", *Journal of the American Statistical Association* 98: 533-544.

List, J.A., Sinha, P. and Taylor, M.H. (2006), "Using choice experiments to value non-market goods and services: Evidence from field experiments", *BE Journal of Economic Analysis and Policy*, 6(2), Article 2.

Louviere, J.J., Flynn, T.N. and Marley A.A.J. (2015), *Best-worst Scaling: Theory, Methods and Applications*, Cambridge University Press.

Louviere, J.J., Hensher, D.A. and Swait, J.D. (2000), *Stated choice methods: Analysis and application*, Cambridge University Press.

McFadden, D. (2001), "Economic choices", *American Economic Review* 91, 351-378.

McFadden, D. and Train, K. (2000), "Mixed MNL models for discrete response", *Journal of Applied Econometrics* 15, 447-470.

Manski, C.F. (2004), "Measuring expectations", *Econometrica* 72, 1329-1376.

Magidson, J. and Vermunt, J. K. (2007) "Removing the Scale Factor Confound in Multinomial Logit Choice Models to Obtain Better Estimates of Preference", Sawtooth Software Conference Proceedings.

Marley, A. A. J. and Louviere, J. J. (2005). "Some probabilistic models of best, worst, and best–worst choices". *Journal of Mathematical Psychology* 49, 464–480.

Oviedo, J. L. and Yoo, H. I. (2017), "A latent class nested logit model for rank-ordered data with application to cork oak reforestation", *Environmental and Resource Economics*, 68, 1021-1051.

Sándor, Z. and Frances, P.H. (2009), "Consumer price evaluations through choice experiments", *Journal of Applied Econometrics* 24, 517-535.

Scarpa, R., Gilbride, T.J., Campbell, D., and Hensher, D. A. (2009) "Modelling attribute non-attendance in choice experiments for rural landscape valuation", *European Review of Agricultural Economics*, 36, 151-174.

Small, K.A., Winston, C. and Yan, J. (2005), "Uncovering the distribution of motorists' preferences for travel time and reliability", *Econometrica* 73, 1367-1382.

Street, D.J. and Burgess, L. (2007), *The construction of optimal stated choice experiments: theory and methods*. New Jersey: Wiley.

Train, K. (2001), "A comparison of hierarchical Bayes and maximum simulated likelihood for mixed logit", Working Paper, Department of Economics, University of California, Berkeley.

Train, K. (2008), "EM algorithms for nonparametric estimation of mixing distributions", *Journal of Choice Modelling* 1, 40–69.

Train, K. (2009), *Discrete choice methods with simulation*. New York: Cambridge University Press.

Train, K. and Weeks, M. (2005), "Discrete choice models in preference space and willingness-to-pay space", Ch. 1 of R. Scarpa and R. Alberini eds. *Application of Simulation Methods in Environmental and Resource Economics*, Dordrecht.

Vann Ophem, H., Stam, P., and Van Praag, B. (1999), "Multichoice Logit: Modeling Incomplete Preference Rankings of Classical Concerts", *Journal of Business & Economic Statistics,* 17, 117-128.

Vossler, C.A., Doyon, M. and Rondeau, D. (2012), "Truth in consequentiality: Theory and field evidence on discrete choice experiments", *American Economic Journal: Microeconomics* 4, 145-171.

Yoo, H.I. and Doiron, D. (2013), "The use of alternative preference elicitation methods in complex discrete choice experiments", *Journal of Health Economics*, 32, 1166-79.