Quant Mark Econ
DOI 10.1007/s11129-012-9124-9

Augmenting discrete-choice data to identify common preference scales for inter-subject analyses

Lynd Bacon · Peter Lenk

Received: 30 December 2011 / Accepted: 2 July 2012

Abstract Discrete-choice experiments are commonly used to measure subjects' preference structures and are often preferred to other measurement methods because they better align with actual choice behavior and avoid some of the well-documented biases inherent in alternative elicitation methods. A limitation of discrete-choice methods is the loss of inter-subject comparability because preference estimates are invariant to linear transformations necessitating indentifying constraints that remove a common, between-subjects utility scale. This constraint limits the application of discrete-choice results to situations where within-subject comparisons are meaningful. They enable one to sort options for each subject but not to sort subjects according to the relative intensity of their preferences. This paper uses auxiliary data to recover a common preference scale for between-subject comparisons. The model combines discrete-choice data with ratings data while adjusting for response biases due to method effects. The joint model moves the identification constraints from the sub-model for the discrete-choice data to the sub-model for the ratings data. The proposed methodology is complementary to willingness-to-pay computations when studies lack price or its economic foundation is untenable.

JEL classification C11 · C18 · M31

Keywords Failure of procedure invariance · Hierarchical Bayes · Discrete-choice conjoint · Ipsative measurements · Random utility theory · Willingness-to-pay

L. Bacon
Loma Buena Associates, 104 Swett Road, Woodside, CA 94062, USA
e-mail: lbacon@lba.com

P. Lenk (✉)
Ross School of Business, University of Michigan, 701 Tappan Street, Ann Arbor, MI 48109-1234, USA
e-mail: plenk@umich.edu

1 Introduction

Subjects' preference structures inform numerous activities, ranging from public policy to marketing (Carroll and Green 1995; Green et al. 2001; Böckenholt 2006). A wide variety of measurement methods and models have been developed to elicit and to infer preference structures. Measurement systems include ratings, ranking, and discrete choice. Ratings-based or metric conjoint uses ratings scales that typically indicate attractiveness or likelihood of purchase. Rankings-based or non-metric conjoint requires subjects to rank-order options or, in special cases, to select the most preferred option in Pick-Best conjoint or to give the most and least preferred options in Best/Worst Analysis (Finn and Louviere 1992).

Ranking methods are examples of ipsative measurements: the sum of the responses is constant a priori. Ranking K items from 1 to K in a task results in the constant sum K(K+1)/2, and Pick-Best tasks that code the best alternative as "1" and the others as "0" have a constant sum of 1. The constant sum imposes a constraint on the measurement scale and removes the common origin across subjects, thus hindering inter-subject comparisons of the intensity of preferences. After a discrete choice study, the researcher is able to sort options for each subject, thereby predicting choice shares, but is not able to sort subjects by their preferences, thus losing the opportunity to segment and to target based on strength of preference. This paper's goal is to expand discrete choice methodology to recover a common origin that allows for the latter activities without compromising the former.

Figure 1 illustrates the impact of losing a common origin. Three subjects evaluate three options. The top panel plots the utilities on an absolute scale with a common

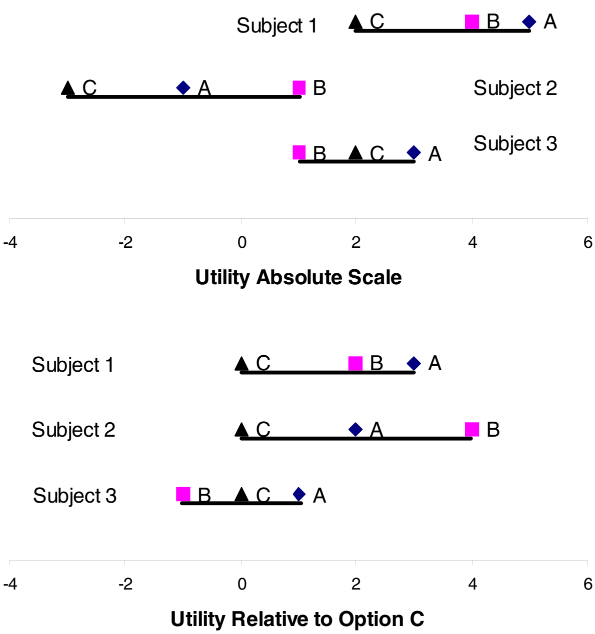


Fig. 1 Absolute versus Relative Scale. *Top panel*: three subjects' utilities for three options on the absolute utility scale. *Bottom panel*: same utilities measured relative to option C

origin. Subject 1 has the largest, average utility for the items, while subject 2 has the smallest. The bottom panel corresponds to a discrete-choice conjoint study where the utility for option C is set to zero to indentify the utilities. The ordering of the options within subjects are the same in the top and bottom panels. On the relative scale, however, it would appear that subject 2 has the largest, average utility and appears to be the best prospect. This is obviously incorrect. The arbitrary choice of the base option (the option with zero utility) can change the ordering of subjects based on their relative utilities. Discrete choice models also set a scaling parameter to one, which would have the effect of expanding or contracting the range of the relative utilities in the bottom panel.

Böckenholt (2004) proposed three methods to retain a common scale when modeling ipsative data. First, in some circumstances the researcher may be able to a priori assume a common origin in situations where preferences judgments are made relative to a common referent. Second, the researcher could force choices among bundles of options, provided that the utility for the bundle is additive in its components. An example would be to select between one luxury sports sedan and the bundle of a minivan and an economy car. By cleverly designing the bundles, Böchenholt demonstrates that the utility origins can be recovered. Third, origins can be recovered by using auxiliary data.

We pursue the third option and propose a joint model that combines the information in discrete-choice data with the information in continuous or ordinal data to estimate a scale that is common across subjects. In standard models, preference estimates from discrete-choice data are invariant to linear transformations. To uniquely estimate the preferences, common identifying constraints are to set an intercept to zero and a scale parameter to one. These identifying constraints result in measuring preferences on subject-specific scales, negating inter-subject comparability. The identifying constraints are imposed by the analyst to estimate the model, and different constraints can result in different between-subjects comparisons. We use continuous or ordinal auxiliary data to identify a common scale for the preferences of the choice data, thus eliminating the need for identification constraints on the preferences. This method thus allows the utility parameters to vary freely while allowing estimation of a common preference scale.[1] Combining different sources of information also improves estimation (Huber et al. 1993; Allenby et al. 1995; Ter Hofstede et al. 2002), which we confirm in this setting. However, our goal is to extend the applicability of discrete-choice measurements by identifying a common scale and allowing for between-subject comparisons.

Our model provides sufficient flexibility to ameliorate problems posed by method effects, in particular, the failure of procedural invariance, which occurs when estimated preference orderings depend on the type of elicitation task, such as picking an option or rating it (Lichtenstein and Slovic 1971; Grether and Plott 1979; Slovic and Lichtenstein 1983). These preference reversals are not necessarily due to violations of

[1] Note that the common scale origin we refer to is not necessarily an absolute origin. For instance, the utility scale from a cell-phone study may not be comparable to the utility scale from a career-choice study or to a medical procedure study. An absolute origin is not needed for comparing the utilities of different respondents in the same study. All that is required is that measurement be on a scale with the same origin for all respondents to be compared in the same study.

specific axioms of rational decision behavior but are due to the elicitation method (Tversky et al. 1990). Comparability effects (Tversky et al. 1987) between stimulus and measurement method may drive these method effects. Subjects pay more attention to attributes that are more strongly related to the response method than those that have weaker relations (Hsee 1996; Vriens et al. 1998; Moore 2004). Nowlis and Simonson (1997) found that easily compared attributes (e.g. price) tend to be relatively more important in choice studies, while "enriched" attributes (e.g. brand) that are hard to compare tend to receive more weight when options are evaluated without reference to competing options as with a ratings. Our model for the auxiliary data rescales the coefficients from the discrete-choice model, thus compensating for method effects, when they exist, by muting or amplifying the coefficients of attributes.

Two alternative methods for inter-subject comparisons with discrete-choice data are choice probabilities and equivalent surplus or willingness-to-pay (WTP). We support both these methods but wish to point out that they have limitations. Choice probabilities require specifying the attributes for the competing alternatives. Given the alternatives and their attributes, the analysts can identify groups of subjects who are more likely to choose different alternatives. However, the results depend critically on the input to the choice simulator, and the competitive attributes may be difficult to specify for new products or fast moving markets (Gilbride et al. 2008). Even a simple attribute such as "Price" can be difficult to specify since different customers face different price schedules, and prices can change rapidly. Other attributes may be unknown to the analysts. Moreover, in realistic applications, consumers face a large number of alternatives, and choice simulators may not produce dominate choices for many consumers. For these reasons, some analysts forego using choice probabilities and directly compare subjects' estimated coefficients. However, such comparisons can be misleading, as illustrated in Fig. 1. The methods of this paper will be useful for researchers who do wish to compare subjects' coefficients after a discrete choice study.

Willingness-to-pay scales the estimated choice coefficients by the price coefficient (Sonnier et al. 2007; Train 2003) if the outside good and price are included in the choice tasks. Clearly, WTP computations are not available when price is not in the study. Attribute screening studies where subjects choose among attributes usually do not include price because price for a single attribute may not be meaningful in isolation from a complete product. Conjoint studies for truly new products or categories in the early stages of development often do not include price because costs are not well developed or subjects do not have a suitable reference for economic value. Researchers may leave price out of a study because it compresses information into a single measure, thus losing important dimensions of the decision problem (Vatn and Bromley 1994). For example, subjects may view high price as a proxy for better quality given the attribute levels. When the researcher does not want to impose a functional form for the utility of money, conjoint designs often include price as a discrete attribute. Then WTP is no longer unique but takes a range of values, which makes inter-subject ranking on WTP indeterminate. Some conjoint studies use subjective price levels (e.g. Low, Medium, and High), which allows subjects to interpret price relative to their knowledge and experience of the category, but also degrades between-subject comparisons of WTP.

The validity of WTP estimation requires the correct specification of the random utility model for the stimuli and money and subjects who obey the laws of rational behavior. Standard WTP computation assumes that the utility of money is linear. When it is nonlinear, economically consistent WTP requires the subjects' wealth (Hanemann 1984), which is often not elicited in conjoint studies and is subject to measurement error. Method effects can also plague WTP computations (McFadden 1994; Kanninen 1995; Roe et al. 1996; Ryan and Wordsworth 2000; Koszegi and Rabin 2004). The estimated price coefficient is critical in WTP computations while probably being one of the least valid estimates when compared to market behavior (Rogers and Renken 2003; Renken et al. 2004) for at least two, contrasting reasons. Subjects are not actually making purchases and may be less price sensitive than in the marketplace, or comparability effects in choice studies may increase price sensitivity. The method proposed here for inter-subject comparisons provides an alternative to WTP that avoids these issues.

The next section presents the joint model for supplementing discrete-choice data with ratings data in order to recover a common scale origin. Section 3 applies the models to several datasets, and demonstrates the efficacy of the methodology for estimating common utility scales and making inter-subject comparisons. We analyze two commercial marketing research studies that test the bounds of ranking subjects by the intensity of their preferences. In the first study, an attitude screening study for test products, the heterogeneity in the preferences for the base attitude is comparable to that for the other attitudes, and they have low correlations. Because of this, the ranking of subjects with and without a common utility origin are substantially different. In the second study, which is a brand study that includes "None," the preferences for the base stimulus "None" is smaller than the preferences for brands, and they are highly correlated. Consequently, the ranking of subjects are similar with and without a common utility origin. However, including a common origin provides insights into the heterogeneity of preferences structures that would be useful for positioning the products. Section 4 concludes with implications of the procedure and future research.

2 Joint model for choice and ratings

Figure 2 gives a schematic of the joint model for augmenting discrete-choice data with ratings. Observed variables are given by rectangular boxes. The left side of the figure is the sub-model for the discrete-choice data, and the right side is for the ratings sub-model. We take the view that subjects' preferences or partworths are the focal parameters of the study, and we use the ratings data to identify a common scale. The heterogeneity in individual level partworths is determined by individual-level variables and random error. The preferences are combined with profile attributes and random errors to form random utilities for the choice task, from which results the observed choice data, depending on the method of measurement. The individual preferences are adjusted by modifying parameters, which account for method effects, to obtain latent variables, which then pass through a cut-point process to obtain the ratings. The identification constraints are put on the ratings parameters and not the preferences. The formal model follows.

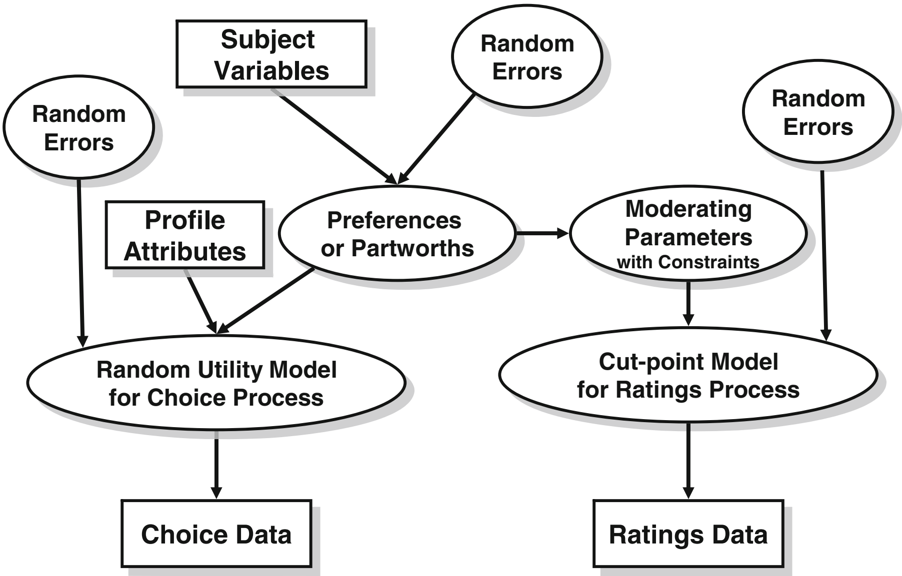


Fig. 2 Model Schema for Discrete Choice and Ratings Data

2.1 Discrete-choice data

Subject i receives a survey that includes a ratings task and a discrete-choice task. There are n subjects. For the discrete-choice part of the survey, subject i completes J_i choice tasks where each task consists of K options.[2] Let Y_{ijk} be subject i's latent, random utility for the kth option in the jth choice set. We assume the standard, linear, compensatory, random utility model:

$$Y_{ijk} = x_{ijk}'\beta_i + \varepsilon_{ijk.} \text{ for } i = 1, \ldots, n; j = 1, \ldots, J_i; \text{ and } k = 1, \ldots, K \qquad (1)$$

where $\mu_{ijk}=x_{ijk}'\beta_i$ is the deterministic component, x_{ijk} is the design vector for the option, β_i is a p-vector of coefficients, and $\varepsilon_{ijk.}$ is the normally distributed random component with mean 0, thus giving the probit model (Aitchison and Bennet 1970; Hausman and Wise 1978). The specification of the error covariance matrix depends on the application. In the empirical section that follows, we use both correlated and uncorrelated probit models in different case studies.

Non-metric methods are based on ranks, which imposes constraints on the random utilities. If a subject i ranks all K options in choice set j according to R_1 for the least preferred option to R_K for the most preferred option, then the relation between the ranks and random utilities is: $Y_{ijR1} \leq Y_{ijR2} \leq \ldots \leq Y_{ijRK}$. With normal errors, inequality constraints are easily incorporated in the estimation algorithm (Albert and Chib 1993; McCulloch and Rossi 1994) for generating the latent utilities by sequentially generating the random utilities from truncated normal distribution. In a Pick-Best study, subject i selects only the most preferred option R_K in choice set j if $Y_{ijRK} \geq Y_{ijk}$ for all

[2] We assume here for explanatory purposes that the number of options is fixed across tasks, although our method does not require it.

k. In Best/Worst (or MaxDiff of Finn and Louviere 1992; Marley and Louvier 2005), subject i selects the least and most preferred options R_1 and R_K if $Y_{ijR1} \leq Y_{ijk} \leq Y_{ijRK}$ for all k.

2.2 Auxiliary data

The auxiliary data we consider may be continuous or ordinal. We assume that the m^{th} auxiliary observation V_{im} for subject i can be related to normally distributed latent variables W_{im} through a link function f: $V_{im}=f(W_{im}|\Omega)$ where Ω is a set of parameters. This structure can take a variety of specific forms, including generalized linear models and limited variable models. If the observations are continuous, then the link function can be the identity.

We focus specifically on ratings data. The rating task uses a H point measurement scale: 1, 2, .., H with appropriate anchors where 1 indicates the least intense response, such as "Strongly Disagree," and H indicates the most intense response, such as "Strongly Agree." In the ratings task, subject i evaluates M items. A cut-point or threshold model (Aitchison and Silvery 1957; McCullagh 1980) is used to translate the latent, continuous response $W_{i,m}$ into the observed ordinal response $V_{i,m}$ according to the cut-points $c_1 < \ldots < c_{H-1}$ where c_1 and c_{H-1} are fixed constants. The relation between the observed ratings and latent variables is: $V_{i,m}=1$ if $W_{i,m} \leq c_1$; $V_{i,m}=j$ if $c_{j-1} < W_{i,m} \leq c_j$; for j=2,…m-1; and $V_{i,m}=H$ if $W_{i,m} > c_{H-1}$. Gelfand et al. (1992) developed MCMC methods for ordered data, and Rossi et al. (2001) and Johnson (2003) used heterogeneous threshold models to correct for scale-usage effects. An advantage of cut-point models is that the observed ratings can be highly, non-normal: bi-modal or skewed, yet the data are consistent with normally distributed latent variables W_{im}.

We define a latent variable for the ratings data as:

$$W_{im} = \varphi_i + \alpha_m + u'_{im}\Psi\beta_i + \xi_{im} \text{ for } i = 1, \ldots, n \text{ and } m = 1, \ldots, M \tag{2}$$

where φ_i is a subject-specific, random effect that has a normal distribution with mean 0 and standard deviation τ; α_m a item-specific main effect; u_{im} is a p-vector of observed variables for item m; Ψ is a pxp diagonal matrix with positive entries ψ_1, .., ψ_p; and $\{\xi_{im}\}$ are independent, normally distributed error terms with standard deviations $\{\sigma_{W,m}\}$. If the components of β_i are not common to both Eqs. (1) and (2), then the design vectors u_{im} or x_{ijk} zero-out components that do not belong to the corresponding equation. If components of β are specific only to Eq. (2) and not Eq. (1), then the corresponding entries in Ψ are one. The parameters φ_i, α_m, and Ψ are introduced to the ratings sub-model to make it more flexible and to accommodate the possibility of method effects, as we will discuss in the next paragraph. A researcher could more simply fit $W_{im}=u_{im}' \beta_i+\xi_{im}$ if method effects are not a concern.

Equation (2) accounts for heterogeneity in rating scale usage and method effects in several ways. First, the threshold model with random effects $\{\varphi_i\}$ accommodates yea-sayer response bias that varies across subjects. Subject who always tend to use the top end of the scale will have a positive value for φ_i, while subject who use the bottom end of the scale will have a negative value for the realized value of the random effect.[3] Essentially, including these random effects perform the same function as mean centering the ratings within subjects. Second, Ψ compensates for method effects by adjusting the impact of the discrete-choice partworths β_i in the ratings sub-model. For example in Nowlis and Simonson (1997) findings, if "enriched" attributes are more salient in ratings than choice, then the corresponding entry in Ψ will be greater than one for that attribute. Conversely, "comparable" attributes may have Ψ values less than one. It also adjusts the partworths to the implicit scale based on the cut-points. If a researcher is unconcerned about method effects, then setting Ψ to the identity matrix is an option. Third, the inclusion of item effects $\{\alpha_m\}$ accounts for additive translations between the ratings and discrete choice scales. Various modifications of this parameterization are possible based on the data structures.

2.3 Identification

Discrete-choice models are not identified because the inequality relations that define the ranking of the options are invariant to affine transformations of the preferences: $Y^*_{ijk}=a+bY_{ijk}$ where a and b are constants and $b>0$ has the same likelihood function or choice probabilities as Y_{ijk} in Eq. (1). Commonly used identification constraints set one of the alternative's intercepts to zero and one of the error variances to one. If the "None" option or outside good is included in the model, common practice is to set its partworth to zero and its error variance to one.

In our joint model, we impose the constraints on the parameters for the ratings data in Eq. (2) and not on the utilities for the discrete-choice data in Eq. (1). The identifying constraints in Eq. (2) assume that at least one of the parameters on the major diagonal of Ψ is one, and at least one of the item effects α_m is zero. In addition, the design vectors u_{im} cannot all be zero, and one of the nonzero entries of the design vector must correspond to the constraint where $\alpha=0$ and $\psi=1$. As previously mentioned, if the researcher is less concerned about methods effects, he or she could simply set all α_m to zero, and the entire diagonal of Ψ to one.

To see that these constraints identify the model, partition β_i into two vectors: $\beta_i=(\beta_{i0}', \beta_{i1}')'$ where β_{i0} is the vector of intercepts, and β_{i1} is the vector of coefficients for non-constant components of x_{ijk}. Also, partition the design vector $u_{im}=(u'_{im0}, u'_{im1})'$ and the diagonal matrix Ψ into Ψ_0 and Ψ_1 in Eq. (2), so that $u'_{im}\ \Psi\ \beta_i= u'_{im0}\ \Psi_0\ \beta_{i0}+u'_{im1}\ \Psi_1\ \beta_{i1}$. Next consider an affine transformation of the utilities: $Y^*_{ijk}=a+bY_{ijk}$ for constants a and $b>0$. Define $\beta^*_i=(a+b\beta_{i0}', b\beta_{i1}')'$, and $\varepsilon^*_{ij}=b\varepsilon_{ij}\sim N(0,b^2\ \Sigma)$. Then $Y^*_{ijk}=x'_{ijk}\beta^*_i+\varepsilon^*_{ijk}$ has the same likelihood or choice probabilities as Y_{ijk} in Eq. (1) . Next, substitute β^*_i into Eq. (2) to obtain

$$W_{im}=au'_{im0}\Psi_0+\varphi_i+\alpha_m+bu'_{im}\Psi\beta_i+\xi_{im}.$$

Because the random effects $\{\varphi_i\}$ and error terms $\{\xi_{im}\}$ have mean zero and one of the α is zero, the constant $au'_{im0}\ \Psi_0$ cannot be added to these elements without

[3] Bayesians conflate fixed and random effects. We follow traditional terminology and call $\{\varphi_i\}$ a "random effect" because we assume that it has a normal distribution across subjects. However, in estimating the model, we treat it as a fixed effect and estimate it for each subject instead of integrating it out of the likelihood function. The traditional approach to fixed effects does not assume a distribution for them.

violating these assumptions. Also, b cannot be absorbed into Ψ because one of its diagonal elements is one. So, we obtain $W*_{im} = \phi*_i + \alpha*_m + u'_{im}\Psi\beta_i + \xi*_{im}$ with $W*_{im}=(W_{im}-au'_{im0}\ \Psi_0)/b$; $\varphi*_i=\varphi_i/b$; $\alpha*_m=\alpha_m/b$; and $\xi*_{im}=\xi_{im}/b$. However, the cut-points c_1 and c_{H-1} are known constants, and $P(W*_{im}<c_1)$ is not equal to $P(W_{im}<c_1)$, and $P(W*_{im}>c_{H-1})$ is not equal to $P(W_{im}>c_{H-1})$. Thus, a linear transformation of $\{Y_{ijk}\}$ changes the joint likelihood of $\{Y_{ij},W_{im}\}$.

2.4 Heterogeneity distribution

In the example applications that follow, we will assume that the heterogeneity in the β_i is a multivariate regression model where z_i is an observed covariate for subject i:

$$\beta_i = \Theta' z_i + \delta_i \text{ and } \delta_i \sim N_p(0, \Lambda). \tag{3}$$

This particular form of heterogeneity is not critical for our thesis, and we could just as easily use other specifications, such as a mixture of normal distributions (Lenk and DeSarbo 2000) or Dirichlet processes (Ansari and Raghuram 2006).

2.5 Prior distributions and estimation

We assume the following prior distributions. For the correlated probit model of Eq. (1) the error covariance matrix has an inverted Wishart distribution, and for the independent probit model, the error variances have inverted Gamma distributions. The cut-points are uniformly distributed subject to the constraint $c_1<\ldots<c_{H-1}$ where c_1 and c_{H-1} are fixed constants. In Eq. (2), the item effects $\{\alpha_m\}$, which are not fixed to zero, have independent normal distributions; the variance τ^2 for the random effects $\{\varphi_i\}$ has an inverted Gamma distribution; the variances $\sigma_{W,m}{}^2$ have inverted Gamma distributions; and the free scaling factors of Ψ have log-normal distributions. In Eq. (3), Θ has a matrix normal distribution, and Λ has an inverted Wishart distribution. When we analyze only the discrete-choice data without the auxiliary data, we use the constrained, inverted Wishart distribution of Nobile (2000) where one of the diagonal elements is set to one.

3 Empirical examples

3.1 Simulated data

A simulation study demonstrates the ability of the joint model to estimate preferences on a common scale without imposing identifying constraints on the coefficients in the discrete-choice sub-model. The simulation mimics 500 "subjects" picking the best option from 20 choice tasks consisting of four profiles each. Three of the profiles in the choice tasks are "branded" goods, and the fourth option is the outside good. Two continuous attributes are "Quality" and "Price." Each subject also rates six profiles, two for each brand, on a five point ordinal scale. Subjects do not rate the outside good. Tables 1 and 2 display the true parameters, the posterior means, and the posterior standard deviations for the ratings sub-model and probit sub-model,

Table 1 Posterior distribution of ratings parameters for simulation study

	TRUE	Posterior Mean	Posterior STD DEV
Cut Points			
c1	0.0	–	–
c2	8.3	8.129	0.145
c3	16.7	16.453	0.155
c4	25.0	–	0.000
Random Effects STD DEV	4.0	4.003	0.146
α			
Item 1[b]	0.0	–	–
Item 2	0.0	−0.413	0.423
Item 3	5.0	5.113	0.218
Item 4	2.0	2.023	0.112
Item 5	−2.0	−2.036	0.419
Item 6	−5.0	−5.109	0.208
ψ			
Brand 1[b]	1.000	–	–
Brand 2	7.389	7.393	0.00129
Brand 3	2.718	2.719	0.00068
None[a]	1.000	–	–
Quality	0.607	0.606	0.00022
Price	0.368	0.368	1.1E-6
Error STD DEV			
Item 1	0.5	0.458	0.150
Item 2	1.0	0.430	0.165
Item 3	1.5	1.472	0.177
Item 4	0.8	0.494	0.180
Item 5	0.3	0.416	0.142
Item 6	0.1	0.381	0.130

[a]"Subjects" did not rate "None." The design vector has a zero for the "None" component of β_i. We fixed the corresponding Ψ to one for convenience[b]Identification constraints are $\alpha_1=0$ and $\psi_{1,1}=1$

respectively. Table 2 also has the posterior means and standard deviations from the standard probit model, which sets the expected utility for the outside good to zero and its error variance to one. The joint model with origins recovers the true parameters, while the parameter estimates for the probit model without origins are biased due to the identification constraints.

The correlations and root-mean-squared errors (RMSE) in Table 3 between the true latent variables or partworths and their posterior means indicate that the joint model outperforms the probit model without origins, which is not surprising because the data were generated from the joint model. The estimates of brand preferences are biased in the standard model because they are measured measures relative to the outside good, and rescaled so that the error variance of the outside good is one. The correlations are about the same for the coefficients of Quality and Price because the

Table 2 Posterior distribution for choice parameters for simulation study

			Posterior Mean		Posterior STD DEV	
		TRUE	With Origin	No Origin	With Origin	No Origin
Error Covariance						
Brand 1	Brand 1	0.50	0.350	0.357	0.158	0.082
Brand 1	Brand 2	0.00	−0.201	0.182	0.085	0.078
Brand 1	Brand 3	0.00	−0.032	0.116	0.138	0.086
Brand 1	None	0.00	0.165	0.475	0.190	0.070
Brand 2	Brand 2	1.00	0.801	0.275	0.233	0.114
Brand 2	Brand 3	0.70	0.555	0.134	0.240	0.110
Brand 2	None	−0.50	−0.612	0.345	0.164	0.085
Brand 3	Brand 3	1.50	1.367	0.186	0.302	0.115
Brand 3	None	−1.00	−1.097	0.178	0.173	0.090
None	None	2.00	1.259	1	0.267	–
Preference Mean						
	Brand 1	5.0	5.022	1.387	0.329	0.416
	Brand 2	3.0	3.038	0.567	0.283	0.429
	Brand 3	2.0	2.149	0.199	0.311	0.437
	None	1.0	1.123	0	0.563	–
	Quality	2.0	1.927	0.818	0.229	0.221
	Price	−1.5	−1.452	−0.620	0.203	0.191
Preference Covariance						
Brand 1	Brand 1	1.0	0.838	0.297	0.217	0.227
Brand 1	Brand 2	0.7	0.609	0.222	0.123	0.217
Brand 1	Brand 3	−0.7	−0.725	0.013	0.110	0.159
Brand 1	None	−0.5	−0.293	–	0.240	–
Brand 1	Quality	0.2	0.224	0.004	0.101	0.044
Brand 1	Price	−0.1	0.052	−0.035	0.055	0.030
Brand 2	Brand 2	1.0	0.967	0.317	0.095	0.226
Brand 2	Brand 3	−0.7	−0.640	0.052	0.070	0.171
Brand 2	None	−0.5	−0.549	–	0.169	–
Brand 2	Quality	0.1	0.062	−0.041	0.053	0.046
Brand 2	Price	0.0	0.041	−0.016	0.040	0.026
Brand 3	Brand 3	1.0	1.037	0.335	0.107	0.191
Brand 3	None	0.5	0.115	–	0.219	–
Brand 3	Quality	0.0	−0.023	−0.052	0.060	0.047
Brand 3	Price	0.0	0.013	−0.008	0.038	0.026
None	None	1.0	0.956	–	0.351	–
None	Quality	−0.2	−0.258	–	0.222	–
None	Price	0.0	−0.160	–	0.127	–
Quality	Quality	0.4	0.466	0.097	0.053	0.016
Quality	Price	0.2	0.197	0.020	0.029	0.006
Price	Price	0.3	0.203	0.053	0.033	0.009

Table 3 Comparison of true latent variables and Partworths to their subject-level posterior means for the simulation study

	Correlation		RMSE	
	With Origin	No Origin	With Origin	No Origin
Latent Variables				
Choice	0.953	0.947	2.826	7.228
Rating	0.985	–	1.967	–
Random Effect φ_i	0.861	–	2.052	–
Preferences				
Brand 1	0.717	0.581	0.660	3.689
Brand 2	0.922	0.723	0.375	2.516
Brand 3	0.873	0.712	0.488	2.031
None	0.444	–	0.907	–
Quality	0.880	0.852	0.321	1.262
Price	0.880	0.856	0.277	0.983

RMSE root mean squared error

probit model only rescales them. Their RMSEs are larger in the standard model without origins.

Parameter estimates from the model with ratings are significantly more accurate than those from the model without ratings, which was expected because the data were generated with the model with ratings. Since the purpose of using auxiliary data is to enable between-subject comparisons, we check the ordering of the subjects based on their partworths and willingness-to-pay (WTP). WTP for a brand is $-(\beta_{i,\text{Brand}}-\beta_{i,\text{None}})/\beta_{i,\text{Price}}$, and WTP for Quality is: $-\beta_{i,\text{Quality, Level}}/\beta_{i,\text{Price}}$. Table 4 displays the Kendall tau statistic, which measures concordance between two measurement scales based on their rank-orders for each item. A statistic of zero indicates that the two sets of measurements are independent, while a statistic of one occurs when both sets of measurements have the same ordering. The Kendall tau statistics in Table 4 indicate that concordance of subject ranks based on brand preferences with and without origins is weak, while they are much stronger for the continuous measures of Quality and Price. A similar pattern occurs for WTP and partworths with a common origin. Figure 3 plots the ranks to verify that the ranks for brands are weakly related, while the ranks for Quality and Price have stronger concordance. This finding may be

Table 4 Kendall tau statistics for simulation study

	Partworths with and without Origin	Partworths and WTP with Origin
Brand A	0.4988	0.5477
Brand B	0.5715	0.6472
Brand C	0.6197	0.4263
Quality	0.8023	0.7344
Price	0.8413	–

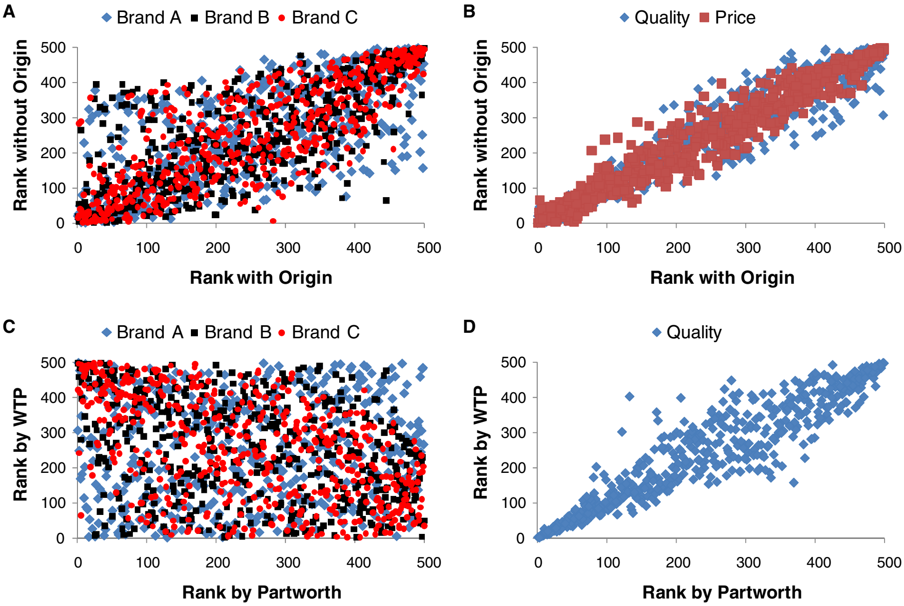


Fig. 3 Subjects' Ranks for Simulated Data. **a** Brand partworths with and without a common origin. **b** Quality and Price with and without a common origin. **c** Brand willingness-to-pay versus brand partworths with a common origin. **d** Quality willingness-to-pay versus quality partworth with a common origin

related to the finding of Lusk and Schroeder (2004), who compared conjoint WTP and actual purchases. They found WTP for a continuous quality variable matched market behavior, while WTP for purchase decisions did not. The simulation confirms that the loss of a common origin in standard choice-models leads to spurious ranking of subjects based on their brand preferences, which have been scaled relative to "None," but ranking of subjects based on preferences or WTP for continuous attributes are appropriate. Keep in mind that the simulation study assumes the correct utility model. If the utility for money is not linear, then the ranking of subjects by the standard WTP may not be appropriate.

3.2 Best/worst study of the importance of product attributes

The second application is a commercial study[4] of attribute importance in a convenience food category. A national sample of 206 subjects who shop the category was given a sensory stimulus of the test product and completed two tasks. First, the subjects rated 14 attributes on a five point Likert scale. Next, they completed 14 Best/Worst[5] choice tasks consisting of 3 attribute items each.

The goal of the study was to inform decisions about product placement, segmentation, and marketing communications. This goal requires not only ranking attitudes within subjects but also comparing attitudes across subjects. The study screens

[4] We wish to thank Knowledge Networks for the data.

[5] Best/Worst with 3 alternatives is equivalent to a full ranking.

attitudes towards the test product and does not relate attributes of potential products to the attitudes. The study does not include price and the outside good.

Figure 4 displays the mean ratings for the attitudes and the difference between the numbers of times the attitude was selected "Best" and "Worst" across all Best/Worst tasks and subjects. The two measures roughly follow the same pattern with greater dispersion in the counts than the mean ratings. The mean ratings are not terribly distinctive due to heterogeneity across subjects. Figure 4 includes 95 % confidence intervals for the mean ratings, which cover the average ratings for most of the other attitudes.

With ratings, subjects' attitudes $\{\beta_{i,j}\}$ were measured on an absolute scale with a common origin. Without ratings, we identified the probit model by setting $\beta_{i,14}$ to zero and the error variance for item 14 to one. Attitudes scores where measured relative to item 14: $\beta^*_{i,j}=(\beta_{i,j}-\beta_{i,14})/\sigma_{14}$ for j=1, .., 13. The selection of item 14 as the base item was arbitrary. Because each choice task only has three options, we used Zeithammer and Lenk (2006) to impute the absent dimensions when estimating the error covariance matrix Σ. Most of the parameter estimates are not particularly interesting to our main thesis, other than Ψ in Eq. 2, which moderates the effect of $\{\beta_{i,j}\}$ in the ratings task. The estimated diagonal of Ψ ranged between 0.855 and 1.001, and three of the items had posterior means that are more than two posterior standard deviations from one. This result indicates method bias between the importance ratings and the Best/Worst task: the expected attitudes tend to be downweighted when used in the rating task.

The focus of the example is the estimated attitude scores $\{\beta_{i,j}\}$ on the common origin scale and $\{\beta^*_{i,j}\}$ estimated without a common, or arbitrary, origin. Table 5 displays the means and standard deviations across subjects of the individual-level posterior means. The population means of the attribute scores are positive on the absolute scale and are mixed positive and negative on the relative scale. Also, parameter heterogeneity (standard deviations) is greatly reduced on the relative scale. The correlation between the two sets of scores ranged between −0.13 and 0.43, and the Kendall tau statistics ranged between −0.06 and 0.33. These results indicate that the ranks of subjects based on their attitudes differ substantially for the two measurement systems. The lack of concordance between the two measurements scales is illustrated in Fig. 5.

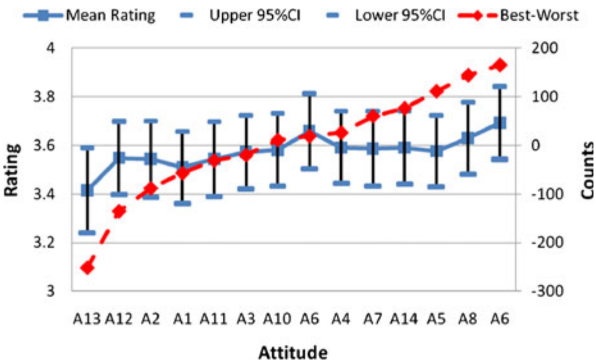


Fig. 4 Attitudes by Mean Rating and Best-Worst Counts

Table 5 Summary statistics of the posterior means of attitude scores for test product

Attitude	Mean With Origin	Mean No Origin	STD DEV With Origin	STD DEV No Origin	Correlation	Kendall's Tau
1	0.916	−0.431	3.226	1.423	0.430	0.332
2	0.868	−0.525	2.802	0.275	−0.020	0.002
3	1.027	−0.333	2.911	0.450	0.218	0.208
4	1.179	−0.133	2.854	0.508	0.123	0.072
5	1.334	0.032	2.933	0.770	0.225	0.158
6	1.618	0.367	2.928	0.779	0.211	0.146
7	1.391	0.115	2.911	0.929	0.206	0.142
8	1.484	0.201	2.885	0.924	0.165	0.132
9	1.326	0.016	2.892	0.374	0.097	0.068
10	1.241	−0.088	2.883	1.032	0.174	0.142
11	1.147	−0.189	2.817	1.066	0.161	0.138
12	0.867	−0.523	2.761	0.334	−0.132	−0.059
13	0.587	−0.831	2.744	0.764	0.044	0.076
14[a]	1.292	0.000	2.897	0.000	–	–

[a] Base item without ratings data

A driver of the different ordering of subjects is the heterogeneity of the base item, which can be seen to be comparable to that of the other 13 items in Table 5. If subjects are homogeneous in their responses about the base item, then subtracting the base item uniformly shifts the relative measures across subjects, thus preserving the ordering of subjects. However, in the presence of heterogeneity for the base item, the ordering of subjects will be altered, and the amount of re-ordering increases with heterogeneity in the base item. The attribute items are ordered in the same way with or without a common origin, within subjects. But the ordering of subjects is different when the base item is heterogeneous.

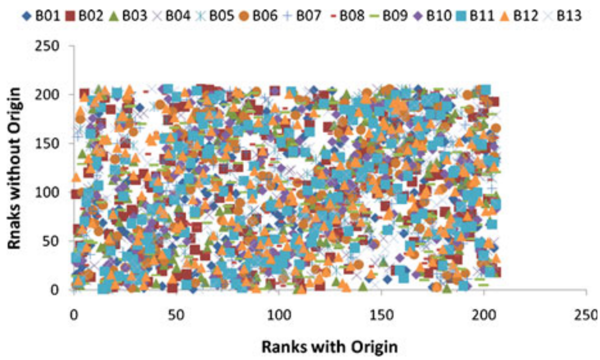
3.3 Pick-best conjoint study

The last application is a commercial conjoint study[6] of a mature, consumer electronics product. Current market prices range from \$25 to \$150, depending on form and features. The study uses a random sample of 416 subjects. Profiles in the study consist of four brands, price on a continuous scale, and six other attributes with a total of 19 levels. The study was implemented with Sawtooth Software's Adaptive Choice Based Conjoint™. The number of tasks and the attribute levels for each profile depend on the subject's previous responses and vary across subjects. The total number of choice tasks ranged from 34 to 71 per subject with an average of 48.2. Subjects also rate four profiles on a five-point scale.

In brand studies it is natural to associate the error terms with brands (Zeithammer and Lenk 2006). The random utility can be thought of as enhanced brand preferences

[6] We wish to thank Chris Chapman and Sawtooth Software for providing the data.

Fig. 5 Ranks of Subjects by Attitudes Differ with and without a Common Origin



where the error variance measures the uncertainty in the utility of the brands given the other attributes. When the same brand appears more than once in some choice task, as it does with the current data, the correlated probit model leads to degenerate choice probabilities because profiles with the same brand are perfectly correlated. To avoid degenerate likelihoods, we assume that random errors are uncorrelated and use uncorrelated probit models with error variances associated with the inside and outside good. Similarly, the profiles in the rating task varied across subjects, and many subjects did not rate all four brands. We simplified Eq. (2) by assuming that all of the rating profiles have the same error variance, and all of the grand means, α, are equal to zero instead of just one component.

To compensate somewhat for these restrictions, we added individual-level scale parameters to the error terms:

$\mathrm{Var}(Y_{i,b})=\sigma_{Y,b}^2/\zeta_{Y,i}$ for subject i and brand b in Eq. (1)
$\mathrm{Var}(W_{i,m})=\sigma_W^2/\zeta_{W,i}$ for subject i and rating m in Eq. (2).

The heterogeneity distributions for the scale factors are Gamma distributions : $\zeta_{Y,i} \sim G(\gamma_Y, \gamma_Y)$ and $\zeta_{W,i} \sim G(\gamma_W, \gamma_W)$, which have means of one and standard deviations of $1/\gamma_Y$ and $1/\gamma_W$. The mean of the heterogeneity distribution is one to identify the individual scale factors and the common variance factors. Lenk et al. (1996) proposed individual-level error variances from an inverted Gamma distribution for metric conjoint, and Fiebig et al. (2010) introduced a log-normal, scale heterogeneity in a logistic choice model. In our model, the random utility error terms are t-distributions after integrating over the individual scale factors. The prior distributions for γ_Y and γ_W are Gamma distributions, G(3,3). The full conditionals for the scale factors also have Gamma distributions; however, estimation of γ_Y and γ_W require a simple Metropolis step.

We detected considerable scale heterogeneity. The posterior means and standard deviations for γ_Y were 0.865 and 0.119 without ratings and 1.245 and 0.165 with ratings. The posterior mean and standard deviation of γ_W were 1.018 and 0.933. The large standard deviation reflects the fact that the sample information for each $\zeta_{W,i}$ is small with only four observations per subject.

The Ψ matrix in Eq. (2) rescales the coefficients from the discrete-choice model to accommodate method effects. To identify the model, we set $\psi_{Price}=1$. The estimated coefficients for the brands are greater than one (1.52, 1.50, 1.08, 1.28), and the posterior standard deviations (0.29, 0.27, 0.20, 0.22) indicate that the coefficients for Brands A and B are different from one. This result is consistent with Nowlis and Simonson (1997) finding that brands are more compatible with ratings than choice. The estimated ψ's for the other attribute levels are also larger than one, which indicate that they have more impact on ratings than in choice relative to the impact of price.

Table 6 displays summary statistics of the individual-level posterior means of the partworths with and without the ratings data. Both the correlations and Kendall's tau statistics between the two sets of partworths are large. Figure 6 plots the ranks for the brand preferences with and without origins and illustrates the concordance between the two sets of ranks. The reasons for this result are that the partworths for "None" have strong correlations, around −0.90, with brand preferences, and that the heterogeneity in "None" is much smaller than the heterogeneity in brand preferences (see "STD DEV with Origin" in Table 6). Consequently, subtracting the preference for "None" from brand preferences does not alter the ordering of subjects by very much.

The model with the common origin provides additional insights into the data. The data have two unusual features. First, brand utilities are highly correlated for each model; they are larger than 0.9 with and without ratings. Subjects do not seem to discriminate among brands. Second, the coefficient for price is negatively correlated with brand preferences, around −0.68 without ratings and −0.75 with ratings. (See Fig. 7) This result seems unusual as one of the main reasons for building strong brand preferences is that brand loyal customers tend to be less price-sensitive. These results

Table 6 Summary statistics of the posterior means of Partworths from conjoint study

Attribute	Mean With Origin	Mean No Origin	STD DEV With Origin	STD DEV No Origin	Corr	Kendall's Tau
Brand A	−0.537	3.464	0.739	7.166	0.984	0.890
Brand B	−0.642	2.726	0.764	7.263	0.984	0.888
Brand C	−0.507	3.720	0.741	7.096	0.982	0.882
Brand D	−0.523	3.634	0.770	7.316	0.982	0.881
A1	−0.008	−0.026	0.124	0.785	0.936	0.760
A2	0.089	0.613	0.122	0.789	0.943	0.782
B1	−0.124	−1.158	0.154	1.495	0.964	0.830
C1	0.114	0.775	0.197	1.710	0.971	0.856
C2	0.095	0.781	0.135	0.795	0.940	0.782
D1	0.039	0.271	0.096	0.472	0.795	0.583
D2	0.059	0.433	0.085	0.301	0.813	0.611
D3	−0.030	−0.219	0.069	0.234	0.652	0.463
E1	0.211	1.596	0.184	1.543	0.972	0.858
E2	−0.047	−0.354	0.091	0.522	0.758	0.538
F1	−0.015	−0.153	0.171	1.266	0.970	0.835
F2	0.052	0.408	0.148	0.976	0.953	0.803
F3	0.027	0.182	0.101	0.569	0.854	0.646
Price	−0.015	−0.105	0.018	0.116	0.982	0.915
None	−1.113	0.000	0.270	0.000	–	–

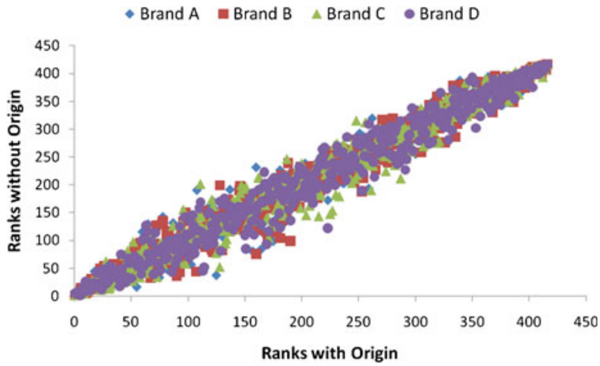


Fig. 6 Ranks of Subjects by Brand Preferences are nearly identical with and without Common Origin

indicate that competition among brands is extremely fierce and small price discounts induce brand switching.

Panel B of Fig. 7 plots the brand preferences by partworths for price and the outside good. One interpretation of the outside good is that it is the subject's utility for the product that he or she currently owns (Haaijer et al. 2001), which is reasonable in this product category due to its wide diffusion. The correlations between brand preferences and "None" are around −0.90. Subjects who have a large utility for the outside good tend to have small, brand preferences and are less likely to select any of the brands, and vice versa. Additionally, the outside good is positively correlated (0.68) with the price coefficient: if a subject likes his or her product, it will be very difficult to use price to induce purchase. If a subject likes his or her current product, he or she is out of the market, and is not interested in any of the brands, regardless of the price. However, if he or she dislikes his or her current product, then none of the

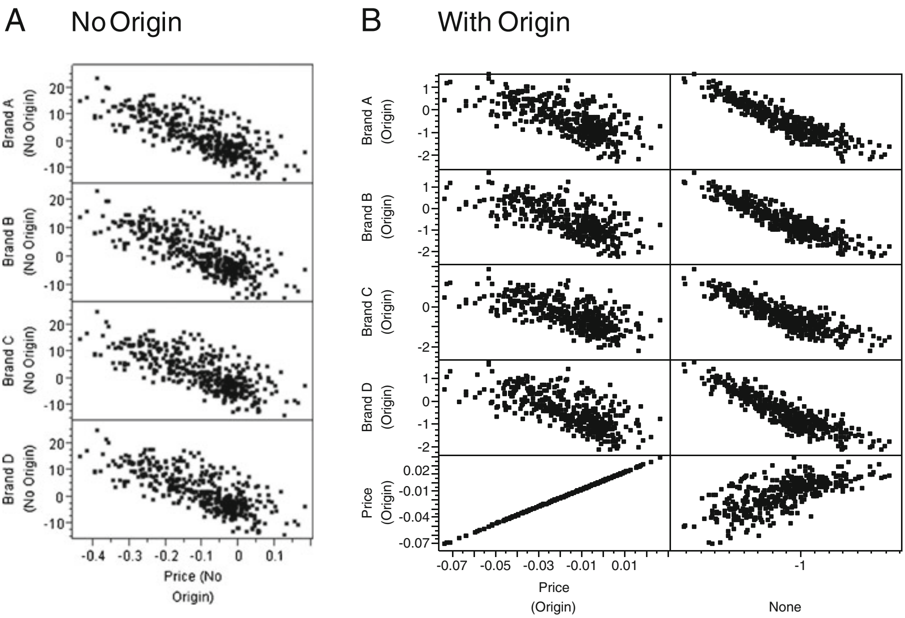


Fig. 7 Brand Preferences are negatively correlated with Price Preferences and None

brands are particularly distinctive, and price becomes the main driver of choice given the same set of product attributes. Estimating subjects' preferences for "None" allows marketing managers to refine their targeting and communication messages, for example, by appealing to customers who are dissatisfied with their current product or by creating dissatisfaction with their current product. Emphasizing brand many not be very effective.

Selected percentiles for WTP are presented in Table 7, and they are roughly similar with and without origin, as expected from the mathematics of WTP despite the different distribution for the partworths in Table 6. The median WTP is reasonable for this product category and features. However, at least 25 % of the subjects have negative WTP for the brands, partially because of positive price coefficients for some subjects (see Fig. 7), and the upper tails are unrealistically large and greatly exceed current market prices. Unreasonable WTP estimates can be discarded or recalibrated (Lenk and Orme 2009).

4 Conclusion

Ipsative measurements lose a common origin for between-subject comparisons because the data from each subject has a constant sum. Discrete-choice methods result in ipsative scales, making it possible to sort options within subjects but not to

Table 7 Willingness-to-pay percentiles from conjoint study

	5 %		25 %		Median		75 %		95 %	
	No Origin	Origin	No Origin	Origin	No Origin	Origin	No Origin	Origin	No Origin	Origin
Brand A	−269.94	−266.47	−7.66	−8.97	34.22	34.06	64.90	62.23	273.30	248.86
Brand B	−338.10	−286.09	−20.92	−14.11	29.14	31.03	60.19	60.80	316.94	262.44
Brand C	−277.11	−229.78	−4.46	−1.54	36.19	35.38	66.34	66.95	296.01	271.54
Brand D	−298.38	−282.78	−7.31	−6.48	33.72	33.43	67.12	64.18	287.63	264.94
A1	−52.38	−45.19	−3.97	−5.31	0.15	−0.08	5.65	6.75	49.94	55.75
A2	−62.27	−65.26	−1.14	−2.67	2.69	2.35	10.81	9.96	75.73	76.08
B1	−160.10	−107.07	−20.25	−12.69	−4.60	−3.40	2.19	2.29	80.16	67.61
C1	−109.96	−68.45	−5.55	−4.26	3.82	3.62	18.26	16.59	100.78	115.05
C2	−46.57	−52.29	−0.64	−2.50	4.52	2.76	11.95	10.23	109.78	80.59
D1	−34.59	−40.34	−1.81	−3.43	1.11	1.03	4.84	6.10	28.16	44.21
D2	−39.83	−38.74	0.23	−1.35	2.30	2.07	7.42	7.51	43.57	46.18
D3	−24.63	−34.55	−4.23	−5.62	−1.09	−0.83	0.33	1.82	19.07	35.46
E1	−141.07	−97.00	−1.62	−0.74	6.33	5.95	29.78	25.49	146.80	141.40
E2	−46.69	−50.13	−7.78	−7.19	−1.11	−1.19	1.91	2.30	33.81	31.75
F1	−77.70	−65.36	−9.92	−8.19	−1.80	−1.32	4.08	4.80	125.20	111.42
F2	−75.56	−72.66	−3.49	−4.17	1.96	1.29	9.01	9.39	65.74	62.43
F3	−44.69	−40.29	−2.20	−4.07	1.24	1.06	5.07	6.32	35.43	39.18

compare subjects by the intensity of their preferences. This paper describes a joint model for discrete-choice data and auxiliary data that estimates a common utility scale to enable inter-subject comparisons based on intensity of preferences. The joint model retains a common scale by shifting identification constraints from the discrete-choice sub-model to the sub-model for the auxiliary data. The joint model accommodates method effects that are inherent in ratings and choice processes by moderating the effects of the partworths on the auxiliary data.

We show that a common scale is retained by the joint model, but we did not investigate how different assumptions for the auxiliary data impact the measurement scale for the partworths. Two rating per subject should suffice to identify the common scale. The rating model includes a subject-specific random effect to accommodate scale-usage heterogeneity, and at least two ratings are needed to separate the scale-usage random effect from the error terms. However, this minimal requirement may not strongly identify the common scale in all applications. Multiple ratings per subject should improve partworth estimation and is a topic for further research. Secondly, the type content of the rating items may affect the partworth measurements. For instance, is it better to rate the same profiles as in the discrete-choice conjoint task, or should they be different? Should the rating and conjoint tasks use the same attributes and levels? Is it better to rate profiles or some other stimuli, such as attribute importance?

This paper focuses on discrete-choice experiments or surveys; however, the method and model can be used with transactional data about product choice if continuous, supplemental data, such as quantity (Arora et al. 1998; Kim et al. 2007) or inter-purchase duration (Jain and Vilcassim 1991; Vilcassim and Jain 1991), are available. Segmentation and targeting procedures frequently are based on market choice data (Grover and Srinivasan 1987; Kamakura and Russell 1989; Rossi et al. 1996) despite the fact that the measurements are ipsative. This paper demonstrates that sorting subjects based on ipsative data can be misleading. Recovering common utility origins would facilitate inter-subject comparisons.

References

Aitchison, J., & Bennet, J. A. (1970). Polychotomous quantal response by maximum indicant. *Biometrika, 57*, 253–262.

Aitchison, J., & Silvery, S. D. (1957). The generalization of probit analysis to the case of multiple responses. *Biometrika, 44*, 131–150.

Albert, J. H., & Chib, S. (1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association, 88*(422), 669–679.

Allenby, G. M., Arora, N., & Ginter, J. L. (1995). Incorporating prior knowledge into the analysis of conjoint studies. *Journal of Marketing Research, 35*, 152–162.

Ansari, A., & Raghuram, I. (2006). Semiparametric Thurstonian models for recurrent choices: a Bayesian analysis. *Psychometrika, 71*(4), 631–657.

Arora, N., Allenby, G. M., & Ginter, J. L. (1998). A hierarchical Bayes model for primary and secondary demand. *Marketing Science, 17*(1), 29–44.

Böckenholt, Ulf. (2004). Comparative judgments as an alternative to ratings: identifying the scale origin. *Psychological Methods, 9*(4), 453–465.

Böckenholt, Ulf. (2006). Thurstonian-based analyses: past, present, and future utilities. *Psychometrika, 71* (5), 615–692.

Carroll, J. D., & Green, P. E. (1995). Psychometric methods in marketing-research. 1. Conjoint-analysis. *Journal of Marketing Research, 32*(5), 385–391.
Fiebig, D. G., Keane, M. P., Louviere, J., & Wasi, N. (2010). The generalized multinomial logit model: accounting for scale and coefficient heterogeneity. *Marketing Science, 29*(3), 393–431.
Finn, A., & Louviere, J. J. (1992). Determining the appropriate response to evidence of public concern: the case of food safety. *Journal of Public Policy and Marketing, 11*(1), 12–25.
Gelfand, A. E., Smith, A. F. M., & Lee, T. M. (1992). Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampling. *Journal of the American Statistical Association, 87* (418), 523–532.
Green, P. E., Krieger, A. M., & Wind, J. (2001). Thirty years of conjoint analysis: reflections and prospects. *Interfaces, 31*(3), S56–S73.
Gilbride, T., Lenk, P., & Brazell, J. (2008). Market share constraints and the loss function in choice-based conjoint analysis. *Marketing Science, 27*(6), 995–1011.
Grether, D. M., & Plott, C. (1979). Economic theory of choice and the preference reversal phenomena. *American Economic Review, 69*, 623–638.
Grover, R., & Srinivasan, V. (1987). A simultaneous approach to market-segmentation and market structuring. *Journal of Marketing Research, 24*(2), 139–153.
Haaijer, R., Kamakura, W., & Wedel, M. (2001). The 'no-choice' alternative in conjoint choice experiments. *International Journal of Market Research, 43*(1), 93–106.
Hanemann, M. W. (1984). Welfare evaluations in contingent valuation experiments with discrete responses. *American Journal of Agricultural Economics, 66*(3), 332–341.
Hausman, J. A., & Wise, D. A. (1978). A conditional probit model for qualitative choice: discrete decisions recognizing interdependence and heterogeneous preferences. *Econometrica, 46*(2), 403–426.
Hsee, C. K. (1996). The evaluability hypothesis: an explanation for preference reversals between joint and separate evaluations of alternatives. *Organizational Behavior and Human Decision Processes, 67*(30), 247–257.
Huber, J., Wittink, D. R., Fiedler, J. A., & Miller, R. (1993). The effectiveness of alternative preference elicitation procedures in predicting choice. *Journal of Marketing Research, 30*(1), 105–114.
Jain, D. C., & Vilcassim, N. J. (1991). Investigating household purchase timing decisions: a conditional hazard function approach. *Marketing Science, 10*(1), 1–23. Winter.
Johnson, T. R. (2003). On the use of heterogeneous thresholds ordinal regression models to account for individual differences in response style. *Psychometrika, 68*(4), 563–583.
Kamakura, W., & Russell, G. (1989). A probabilistic choice model for market segmentation and elasticity structure. *Journal of Marketing Research, 26*, 379–390.
Kanninen, B. J. (1995). Bias in discrete response contingent valuation. *Journal of Environmental Economics and Management, 28*, 114–125.
Kim, J., Allenby, G. M., & Rossi, P. E. (2007). Product attributes and models of multiple discreteness. *Journal of Econometrics, 38*(1), 208–230.
Koszegi, B., & Rabin, M. (2004). A model of reference-dependent preferences. *Quarterly Journal of Economics, 121*(4), 1133–1166.
Lenk, P., & DeSarbo, W. (2000). Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika, 65*(1), 93–119.
Lenk, P., DeSarbo, W., Green, P., & Young, M. (1996). Hierarchical Bayes conjoint analysis: recovery of partworth heterogeneity from reduced experimental designs. *Marketing Science, 15*(2), 173–191.
Lenk, P., & Orme, B. (2009). The value of informative priors in Bayesian inference with sparse data. *Journal of Marketing Research, 46*(6), 832–845.
Lichtenstein, S., & Slovic, P. (1971). Reversal of preferences between bids and choices in gambling decisions. *Journal of Experimental Psychology, 89*, 46–55.
Lusk, J. L., & Schroeder, T. C. (2004). Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics, 86*(2), 467–482.
Marley, A. A. J., & Louvier, J. J. (2005). Some probabilistic models of best, worst, and best-worst choices. *Journal of Mathematical Psychology, 49*, 464–480.
McCullagh, P. (1980). Regression models for ordinal data (with Discussion). *Journal of the Royal Statistical Society, Series B, 42*, 109–142.
McCulloch, R., & Rossi, P. E. (1994). An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics, 64*(1–2), 207–240.
McFadden, D. L. (1994). Contingent valuation and social choice. *American Journal of Agricultural Economics, 76*(4), 689–708.

Moore, W. L. (2004). A cross-validity comparison of rating-based and choice-based conjoint analysis models. *International Journal of Research in Marketing, 21*(3), 299–312.
Nobile, A. (2000). Comment: Bayesian multinomial probit models with a normalization constraint. *Journal of Econometrics, 99*, 335–345.
Nowlis, S. M., & Simonson, I. (1997). Attribute-task compatibility as a determinant of consumer preference reversals. *Journal of Marketing Research, 34*(2), 205–218.
Renken, T., Rogers, G., & Huber, J. (2004). *A Comparison of Conjoint and Scanner Data-Based Price Elasticity Estimates*. Advanced Research Techniques Forum, Conference presentation not published.
Roe, B., Boyle, K. J., & Teisl, M. F. (1996). Using conjoint analysis to derive estimates of compensating variation. *Journal of Environmental Economics and Management, 31*, 145–159.
Rogers, G. & Renken, T. (2003). Validation and Calibration of Choice-Based Conjoint for Pricing Research. *Sawtooth Software Conference Proceedings*, 209–215.
Rossi, P. E., Gilula, Z., & Allenby, G. M. (2001). Overcoming scale usage heterogeneity: a Bayesian hierarchical approach. *Journal of the American Statistical Association, 96*(453), 20–31.
Rossi, P. E., McCulloch, R. E., & Allenby, G. M. (1996). The value of purchase history data in target marketing. *Marketing Science, 15*(4), 321–340.
Ryan, M., & Wordsworth, S. (2000). Sensitivity of willingness to pay estimates to the level of attributes in discrete choice experiments. *Scottish Journal of Political Economy, 47*(5), 504–523.
Sonnier, G., Ainslie, A., & Otter, T. (2007). Heterogeneity distributions of willingness-to-pay in choice models. *Quantitative Marketing and Economics, 5*(3), 313–331.
Slovic, P., & Lichtenstein, S. (1983). Preference reversals—a broader perspective. *American Economic Review, 73*(4), 596–605.
Ter Hofstede, F., Kim, Y. C., & Wedel, M. (2002). Bayesian prediction in hybrid conjoint analysis. *Journal of Marketing Research, 39*, 253–261.
Train, K. (2003). *Discrete Choice Methods with Simulation*, (pp. 43–45, 156, 312). Cambridge, England: Cambridge University Press.
Tversky, A., Sattath, S., & Slovic, P. (1987). Contingent weighting in judgment and choice. *Psychological Review, 95*(3), 371–384.
Tversky, A., Slovic, P., & Kahneman, D. (1990). The causes of preference reversal. *American Economic Review, 80*(1), 204–217.
Vatn, A., & Bromley, D. W. (1994). Choices without prices without apologies. *Journal of Environmental Economics and Management, 26*, 129–148.
Vriens, M., Oppewal, H., & Wedel, M. (1998). Ratings based versus choice-based latent class conjoint models—an empirical comparison. *Journal of the Market Research Society, 40*(3), 237–248.
Vilcassim, N. J., & Jain, D. C. (1991). Modeling purchase-timing and brand-switching behavior incorporating explanatory variables and unobserved heterogeneity. *Journal of Marketing Research, 28*(1), 29–41.
Zeithammer, R., & Lenk, P. (2006). Bayesian estimation of multivariate normal models when dimensions are absent. *Journal of Quantitative Marketing and Economics, 4*(3), 241–265.