Case Study 3
Overcoming Scale Usage Heterogeneity

Questions that use a discrete ratings scale are commonplace in survey research. Examples in marketing include customer satisfaction measurement and purchase intention. Survey research practitioners have long commented that respondents vary in their usage of the scale; common patterns include using only the middle of the scale or using the upper or lower end. These differences in scale usage can impart biases to correlation and regression analyses. In order to capture scale usage differences, Rossi *et al.* (2001) develop a model with individual scale and location effects and a discrete outcome variable. The joint distribution of all ratings scale responses is modeled, rather than specific univariate conditional distributions as in the ordinal probit model. The model is applied to a customer satisfaction survey where it is shown that the correlation inferences are much different once proper adjustments are made for the discreteness of the data and scale usage. The adjusted or latent ratings scale is also more closely related to actual purchase behavior.

BACKGROUND

Customer satisfaction surveys, and survey research in general, often collect data on discrete rating scales. Figure CS3.1 shows a sample questionnaire of this type from Maritz Marketing Research Inc., a leading customer satisfaction measurement firm. In this sample questionnaire a five-point scale (excellent to poor) is used, while in other cases seven- and ten-point scales are popular. Survey research practitioners have found that respondents vary in their usage of the scale. In addition, it has been observed that there are large cultural or cross-country differences in scale usage, making it difficult to combine data across cultural or international boundaries. These different usage patterns, which we term 'scale usage heterogeneity', impart biases to many of the standard analyses

Bayesian Statistics and Marketing P. E. Rossi, G. M. Allenby and R. McCulloch
 ISBN: 0-470-86367-6

MARITZ
Service Quality Review

Please mark the appropriate circle for each question. Compare OUR PERFORMANCE during the PAST 12 MONTHS to YOUR EXPECTATIONS of what QUALITY SHOULD BE.

	Much Better Than	Better Than	Equal to	Less Than	Much Less Than	Not Applicable
Overall Performance	○	●	○	○	○	○
Service						
1. Efficiency of service call handling.	○	○	○	○	○	○
2. Professionalism of our service personnel.	○	○	○	○	○	○
3. Response time to service calls.	○	○	○	○	○	○
Contract Administration						
4. Timeliness of contract administration.	○	○	○	○	○	○
5. Accuracy of contract administration.	○	○	○	○	○	○

Please share your comments and suggestions for improvements:

Figure CS3.1 Example of customer satisfaction survey questionnaire

conducted with ratings data, including regression and clustering methods as well as the identification of individuals with extreme views.

The standard procedure for coping with scale usage heterogeneity is to center each respondent's data by subtracting the mean over all questions asked of the respondent and dividing by the standard deviation. The use of respondent means and standard deviations assumes that the response data is continuously distributed and from an elliptically symmetric distribution. Furthermore, the estimates of the individual location and scale parameters obtained by computing the mean and standard deviation over a small number of questions are often imprecise.

In order to choose an appropriate modeling strategy for ratings data, we must consider the types of analyses that will be conducted with this data as well as the basic issues of what sort of scale information (ratio, interval or ordinal) is available in this data. To facilitate this discussion, assume that the data is generated by a customer satisfaction survey; however, the methods developed here apply equally well to any data in which a ratings scale is used (examples include purchase intentions and psychological attitude measurement). In the typical customer satisfaction survey, respondents are asked to give their overall satisfaction with a product as well as assessments of satisfactions with various dimensions of the product or service. Ratings are made on five-, seven- or ten-point scales. We will focus on two major uses of customer satisfaction ratings data: measurement of the relationship between overall satisfaction and satisfaction with specific product attributes; and identification of customers with extreme views. Scale usage heterogeneity can substantially bias analyses aimed at either use.

For example, if some respondents tend to use either the low or high end of the scale, this will tend to bias upward any measure of correlation between two response

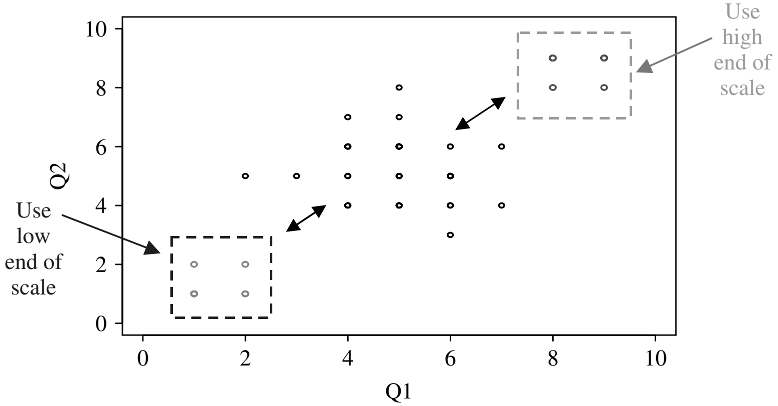

Figure CS3.2 Scale usage heterogeneity and upward correlation bias

items. The middle group of points in Figure CS3.2 represents a hypothetical situation in which all respondents have the same scale usage. If some respondents use the upper or lower end of the scale, this will move points outward from the middle grouping, creating a higher but spurious correlation. Thus, any covariance-based analysis of rating scale data such as regression or factor analysis can be substantially biased by scale usage heterogeneity, aside from the problems associated with using discrete data in methods based on the assumption of continuous elliptically symmetric distributions. In addition, any cluster analysis or filtering of the data for respondents with extreme views will tend to identify a group of nay or yea sayers whose true preferences may not be extreme or even similar.

Practitioners have long been aware of scale usage patterns and often center the data. We observe N respondents answering M questions on a discrete rating scale consisting of the integers 1 to K; the data array is denoted $X = \{x_{ij}\}$, an $N \times M$ array of discrete responses $x_{ij} = \{k\}$, $k = 1, \ldots, K$. Centering would transform the X array by subtracting row means and dividing by the row standard deviation:

$$X^* = \left[\left(x_{ij} - \overline{x}_i\right)/s_i\right].$$

After the data are centered, standard correlation and regression methods are used to examine the relationship between various questions. To identify extreme respondents or to cluster respondents, it is more typical to use the raw response data.

To select the appropriate analysis method, it is important to reflect on the nature of the scale information available in ratings data. Our perspective is that the discrete response data provides information on underlying continuous and latent preference/satisfaction. Clearly, the ratings provide ordinal information in the sense that a higher discrete rating value means higher true preference/satisfaction. It is our view that ratings data can also provide interval information, once the scale usage heterogeneity has been properly accounted for. However, we do not believe that even properly adjusted ratings data can provide ratio-level information. For example, if a respondent gives only ratings at the top end of the scale, we cannot infer that he/she is extremely satisfied. We can only infer that the level of relative satisfaction is the same across all items for this respondent.

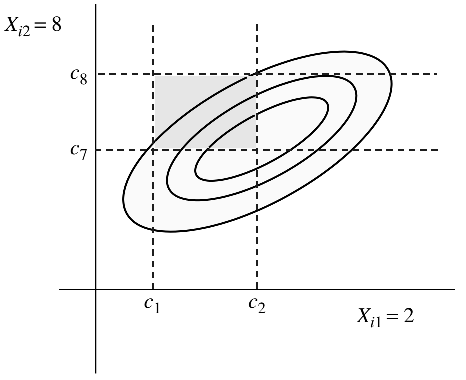

Figure CS3.3 Computing the multinomial probabilities

Centering acknowledges this fundamental identification problem, but results in imprecise row mean and row standard deviation estimates which introduce considerable noise into the data. In most cases, fewer than 20 questions are used to form respondent means and standard deviation estimates. Furthermore, the use of centered data in correlation, regression or clustering analyses ignores the discrete aspect of this data (some transform the data prior to centering, but no standard transformation can change the discreteness of the data). In the next section, we develop a model which incorporates both the discrete aspects of the data and scale usage heterogeneity.

MODEL

Our model is motivated by the basic view that the data in the X response array is a discrete version of underlying continuous data. For $i = 1, \ldots, N$ and $j = 1, \ldots, M$, let y_{ij} denote the latent response of individual i to question j. Let $y_i' = [y_{i1}, \ldots, y_{iM}]$ denote the latent response of respondent i to the entire set of M questions. Assume there are $K + 1$ common and ordered cut-off points $\{c_k : c_{k-1} \leq c_k, k = 1, \ldots, K\}$, where $c_0 = -\infty$, $c_K = +\infty$, such that for all i, j and k,

$$x_{ij} = k \text{ if } c_{k-1} \leq y_{ij} \leq c_k \tag{CS3.1}$$

and

$$y_i \sim N(\mu_i^*, \Sigma_i^*). \tag{CS3.2}$$

The interpretation of the model in (CS3.1)–(CS3.2) is that the observed responses are iid multinomial random variables, where the multinomial probabilities are derived from an underlying continuous multivariate normal distribution. The set of cut-offs $[c_0, \ldots, c_K]$ discretizes the normal variable y_{ij}.

The probability that $x_i' = [x_{i1}, \ldots, x_{iM}]$ takes on any given realization (a vector of M integers between 1 and K) is given by the integral of the joint normal distribution of y over the appropriately defined region. For example, if $M = 2$, and $x_i' = [2, 8]$, then Figure CS3.3 depicts this integral of a bivariate normal distribution over a rectangle defined by the appropriate cut-offs.

The above-proposed model is different in nature from some latent variable models used in Bayesian analyses of discrete data. Most such models deal with grouped data in the form of contingency tables, where the prior distribution of multinomial probabilities is usually taken as Dirichlet over multidimensional arrays. There, the problem of interest is usually of modeling probabilities forming certain patterns of statistical dependence. Here we are interested in modeling individual responses to make individual measurements comparable for the sake of correlation and regression analysis.

It should be noted that the model in (CS3.1) is not a standard ordinal probit model. We have postulated a model of the *joint* discrete distribution of the responses to all M questions in the survey. Standard ordinal probit models would focus on the conditional distribution of one discrete variable given the values of another set of variables. Obviously, since our model is of the joint distribution, we can make inferences regarding various conditional distributions so that our model encompasses the standard conditional approach. In analysis of ratings survey data, we are required to have the capability of making inferences about the marginal distribution of specific variables as well as conditional distributions so that a joint approach would seem natural.

The model in (CS3.1)–(CS3.2) is overparameterized since we have simply allowed for an entirely different mean vector and covariance matrix for each respondent. In order to allow for differences in respondent scale usage without overparameterization, the y vector is written as a location and scale shift of a common multivariate normal variable:

$$y_i = \mu + \tau_i \iota + \sigma_i z_i, \qquad z_i \sim N(0, \Sigma). \tag{CS3.3}$$

We have allowed for a respondent-specific location and scale shift to generate the mean and covariance structure in (CS3.2) with $\mu_i^* = \mu + \tau_i \iota$ and $\Sigma_i^* = \sigma_i^2 \Sigma$. Both the analysis of the joint distribution of questions and the identification of customers with extreme views are based on the set of model parameters $\{z_i\}$, μ, Σ.

The model (CS3.3) accommodates scale usage via the (τ_i, σ_i) parameters. For example, a respondent who uses the top end of the scale would have a large value of τ and a small value of σ. It is important to note that this model can easily accommodate lumps of probability at specific discrete response values. For example, in many customer satisfaction surveys, a significant fraction of respondents will give only the top value of the scale for all questions in the survey. In the model outlined in (CS3.1)–(CS3.3), we would simply have a normal distribution centered far out (via a large value of τ) so that there is a high probability that y will lie in the region corresponding to the top rating. We can also set τ to zero and make σ_i large to create 'piling up' at both extremes. However, our model cannot create lumps of probability at two different non-extreme values (to accommodate a respondent who uses mostly 2 s and 9 s on a ten-point scale).

As a modeling strategy, we have chosen to keep the cut-offs c common across respondents and shift the distribution of the latent variable. Another strategy would be to keep a common latent variable distribution for all respondents and shift the cut-offs to induce different patterns of scale usage. The problem here would be choosing a flexible but not too overparameterized distribution for the cut-offs. Given that we often have a small number of questions per respondent (M), a premium should be placed on parsimony. There is also a sense in which the model in (CS3.3) can be interpreted as a model with respondent-specific cut-offs. If we define

$$c_i^* = \tau_i + \sigma_i c$$

where c is the vector of cut-offs, then we have the same model with respondent-specific cut-offs and a common invariant distribution of latent preferences.

The model specified in (CS3.3) is not identified. The entire collection of τ_i parameters can be shifted by a constant, and a compensating shift can be made to μ without changing the distribution of the latent variables. Similarly, we can scale all σ_i and make a reciprocal change to Σ. As discussed below, we solve these identification problems by imposing restrictions on the hierarchical model. $(\tau_i, \ln\sigma_i)$ are assumed to be bivariate normal:

$$\begin{bmatrix} \tau_i \\ \ln\sigma_i \end{bmatrix} \sim N(\phi, \Lambda). \tag{CS3.4}$$

The model in (CS3.4) allows for a correlation between the location and scale parameters. For example, if there is a subpopulation which uses the high end of the scale, then we would expect an inverse relationship between τ and σ. In most applications of hierarchical models, the location and scale parameters are assumed to be independent.

We achieve identification of τ_i by imposing the restriction $\mathrm{E}[\tau_i] = 0$. Since the distribution of τ_i is symmetric and unimodal, we are also setting the median and mode of this distribution to zero. Greater care must be exercised in choosing the identification restriction for σ_i, as this distribution is skewed. One logical choice might be to set $\mathrm{E}[\sigma_i^2] = 1$. This imposes the identification restriction, $\phi_2 = -\lambda_{22}$. However, as the dispersion parameter (λ_{22}) is increased, the distribution of σ_i becomes concentrated around a mode smaller than 1.0 with a fat right tail. Our view is that this is an undesirable family of prior distributions. The right panel of Figure CS3.4 illustrates the prior on σ_i for small and large values of λ_{22}.

A more reasonable approach is to restrict the mode of the prior on σ_i to be 1. This imposes the restriction $\phi_2 = \lambda_{22}$. The left panel of Figure CS3.4 shows this family of distributions. As λ_{22} increases, these distributions retain the bulk of their mass around 1, but achieve greater dispersion by thickening the right tail.[1] Thus, we employ two identification restrictions:

$$\phi_1 = 0 \quad \text{and} \quad \phi_2 = \lambda_{22}. \tag{CS3.5}$$

Even with the cut-offs $\{c_k\}$ assumed fixed and known, the model in (CS3.1) and (CS3.3) is a very flexible model which allows for many possible discrete outcome distributions. In particular, the model allows for 'piling up' of probability mass at either or both endpoints of the ratings scale – a phenomenon frequently noted in CSM data. For further flexibility, we could introduce the cut-offs as free parameters to be estimated. However, we would have to recognize that identification restrictions must be imposed on the cut-off parameters since shifting all cut-offs by a constant or scaling all cut-offs is redundant with appropriate changes in μ and Σ. For this reason, we impose the following identification restrictions:

$$\begin{aligned} \sum_k c_k &= m_1, \\ \sum_k c_k^2 &= m_2. \end{aligned} \tag{CS3.6}$$

[1] In Rossi *et al.* (2001) the strategy of setting the mean to 1 is used for identification.

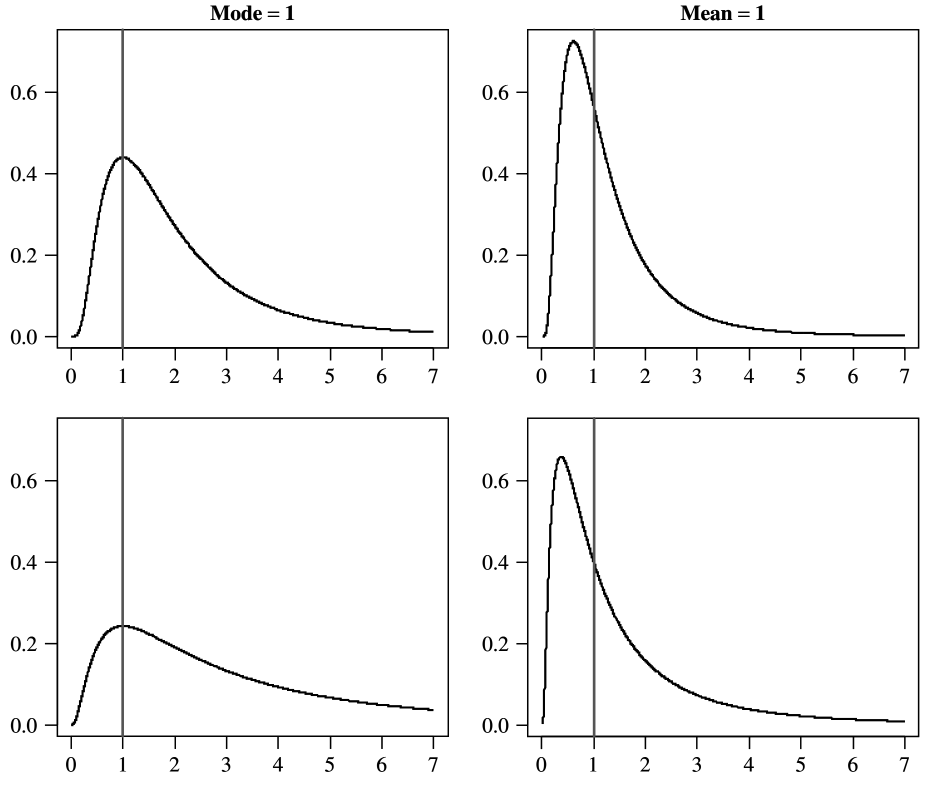

Figure CS3.4 Alternative prior distributions for σ_i

The model in (CS3.1)–(CS3.6) is fully identified but introduces $K-2$ free cut-off parameters. In order to make the model more parsimonious, we will consider further restrictions on the c_k parameters. If we impose equal spacing of the cut-offs, then by (CS3.6) there would be no free c_k parameters. Once the identification restrictions in (CS3.6) are imposed, the only sort of flexibility left in the cut-offs is to introduce skewness or nonlinear spread in the values. In order to allow for nonlinear spread while keeping the number of parameters to a minimum, we impose the further restriction that the cut-off values satisfy a quadratic equation:

$$c_k = a + bk + ek^2, \qquad k = 1, \dots, K-1. \tag{CS3.7}$$

For example, consider the case of a ten-point scale with $a = 5.5$, $b = 1$ and $e = 0$; then $c_1 = 1.5, c_2 = 2.5, \dots, c_9 = 9.5$. Johnson and Albert (1999) review univariate ordinal probit models in which the cut-offs are not parameterized and are estimated with a diffuse prior subject to different identification conditions.

Given the identification restrictions in (CS3.6) and the parameterization in (CS3.7), e is the only free parameter; that is, given m_1, m_2, and e, we can solve for a and b by substituting for c_k in (CS3.6) using (CS3.7). In our implementation, m_1 and m_2 are selected so that when $e = 0$, we obtain a standard equal spacing of cut-off values, with each centered around the corresponding scale value. This means that for a ten-point scale, $m_1 = \sum_{k=1}^{K-1}(k + 0.5) = 49.5$ and $m_2 = \sum_{k=1}^{K-1}(k + 0.5)^2 = 332.25$. The mapping from e to the quadratic coefficients in (CS3.7) is only defined for a bounded range of e values.

The role of e is to allow for a skewed spreading of the cut-off values as shown in Figure CS3.5. A positive value of e spreads out the intervals associated with high scale ratings and compresses the intervals on the low end. This will result in massing of probability at the upper end.

PRIORS AND MCMC ALGORITHM

To complete the model, we introduce priors on the common parameters:

$$\pi(\mu, \Sigma, \phi, \Lambda, e) = \pi(\mu)\pi(e)\pi(\Sigma)\pi(\phi)\pi(\Lambda), \tag{CS3.8}$$

with

$$\begin{aligned} &\pi(\mu) \propto \text{constant}, \\ &\pi(e) \propto \ \text{Unif}\,[-0.2, 0.2], \\ &\Sigma \sim \text{IW}(\nu_\Sigma, V_\Sigma), \\ &\Lambda \sim \text{IW}(\nu_\Lambda, V_\Lambda). \end{aligned} \tag{CS3.9}$$

That is, we are using flat priors on the means and the cut-off parameter and standard Wishart priors for the inverse of the two covariance matrices. We note that the identification restriction in (CS3.5) means that the prior on Λ induces a prior on ϕ. Figure CS3.5 shows that the range of e in our uniform prior is more than sufficient to induce wide variation in the patterns of skewness in the cut-offs. We use diffuse but proper settings

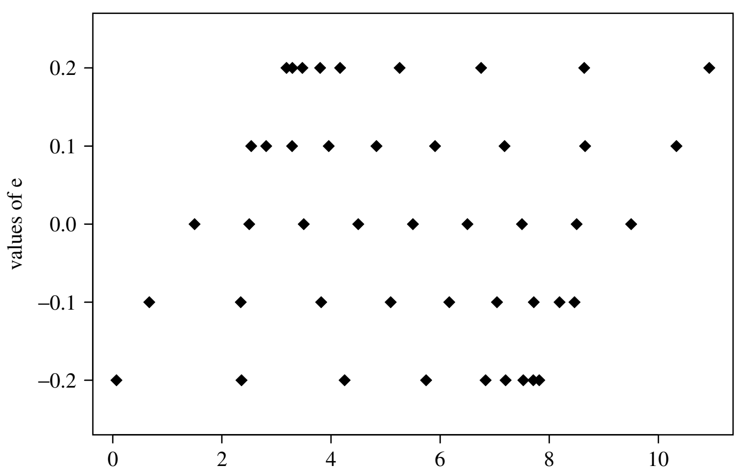

Figure CS3.5 The role of the cut-off parameter, e

for the priors on Σ^{-1} and Λ^{-1}. These parameter values center the prior on Σ over the identity matrix

$$\nu_\Sigma = \dim(\Sigma) + 3 = K + 3, \qquad V_\Sigma = \nu_\Sigma I. \tag{CS3.10}$$

The prior on Λ influences the degree of shrinkage in the τ_i, σ_i estimates. Our hierarchical model adapts to the information in the data regarding the distribution of τ_i, σ_i, subject to the influence of the prior on the hyperparameter Λ. There will rarely be more than a small number of questions on which to base estimates of τ_i, σ_i. This means that the prior on Λ may be quite influential. In most hierarchical applications, there is a subset of units for which a good deal of information is available. This subset allows for determination of Λ via adaptive shrinkage. However, in our situation, this subset is not available and the prior on Λ has the potential to exercise more influence than normal. For these reasons, we will exercise some care in the choice of the prior on Λ. We will also consider and recommend prior settings somewhat tighter than typically used in hierarchical contexts.

The role of the prior on Λ is to induce a prior distribution on τ_i, σ_i. To examine the implications for choice of the prior hyperparameters, we will compute the marginal prior on τ_i, σ_i via simulation. The marginal prior is defined by

$$\pi(\tau, \sigma) = \int p(\tau, \sigma|\Lambda)\pi(\Lambda|\nu_\Lambda, V_\Lambda)\mathrm{d}\Lambda. \tag{CS3.11}$$

To assess ν_Λ, V_Λ, we consider a generous range of possible values for τ_i, σ_i, but restrict prior variation to not much greater than this permissible range. For τ, we consider the range ± 5 to be quite generous in the sense that this encompasses much of a ten-point scale. For σ, we must consider the role of this parameter in restricting the range of possible values for the latent variable. Small values of σ correspond to respondents who only use a small portion of the scale, while large values would correspond to respondents who use the entire scale. Consider the ratio of the standard deviation of a 'small' scale range user (e.g. someone who only uses the bottom or top three scale numbers) to the standard deviation of a 'mid-range' user who employs a range of five points on the ten-point scale. This might correspond to a 'small' value of σ. The ratio for a respondent who uses endpoints and the 'middle' of the scale (e.g. 1, 5, 10) – a 'large range' user – to the 'mid-range' user could define a 'large value' of σ. These computations suggest that a generous range of σ values would be (0.5, 2). We employ the prior settings corresponding to a relatively informative prior on τ_i, σ_i:

$$\nu_\Lambda = 20, \qquad V_\Lambda = (\nu_\Lambda - 2 - 1)\bar{\Lambda}, \qquad \bar{\Lambda} = \begin{bmatrix} 4 & 0 \\ 0 & 0.5 \end{bmatrix}. \tag{CS3.12}$$

The settings in (CS3.12) ensure that $E[\Lambda] \doteq \bar{\Lambda}$. If ν_Λ is set to 20, then the resulting marginal prior on τ_i, σ_i provides coverage of the relevant range without admitting absurdly large values, as illustrated in Figure CS3.6.

The model defined in equations (CS3.1)–(CS3.4) with priors given by (CS3.9) and identification restrictions (CS3.5)–(CS3.7) is a hierarchical model that can be analyzed with some modifications to the standard Gibbs sampler. Our interest centers not only on common parameters but also on making inferences about the respondent-specific scale usage and latent preference parameters, ruling out the use of classical statistical methods. Four problems must be overcome to construct the sampler.

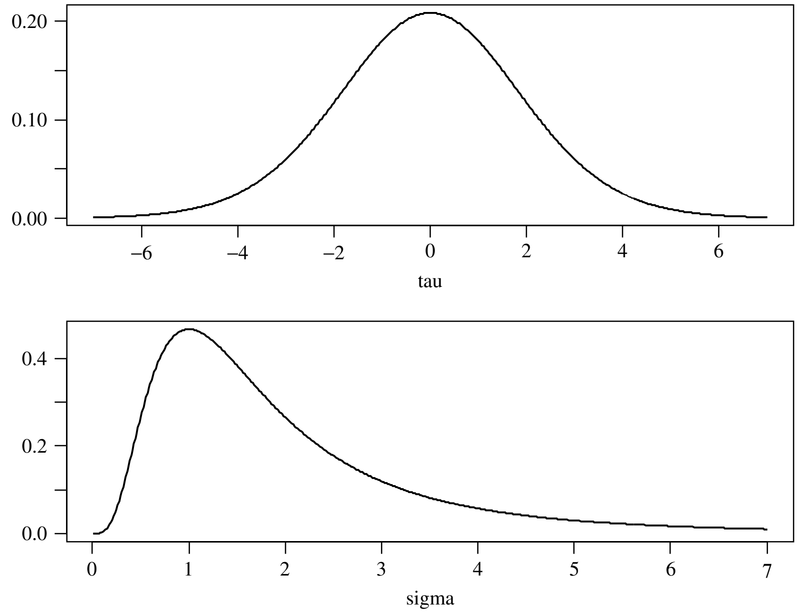

Figure CS3.6 Marginal priors on τ and σ

1. Data augmentation requires a method for handling truncated multivariate normal random variables as in the multinomial probit model (see Chapter 4).
2. One of the conditional distributions in the Gibbs sampler requires the evaluation of the integral of multivariate normal random variables over a rectangle. We use the GHK simulation method (see Chapter 2).
3. The random effects model and priors are not always conditionally conjugate.
4. We accelerate the Gibbs sampler by integrating out some of the latent variables to block e and $\{y_i\}$.

We refer the reader to the appendix of Rossi *et al.* (2001) for more details.

DATA

To illustrate our method, we examine a customer satisfaction survey done in a business-to-business context with an advertising product. This dataset can be loaded once the package
R *bayesm* package has been installed using the R command `data(customerSat)`. A total of 1811 customers were surveyed as to their satisfaction with overall product performance, aspects of price (three questions) and various dimensions of effectiveness (six questions). Figure CS3.7 lists the specific questions asked. All responses are on a 10 point ratings scale.

On a scale from 1 to 10 where 10 means an "Excellent" performance and 1 means a "Poor" performance, please rate BRAND on the following items:

Q1. Overall value

Price:

Q2. Setting competitive prices.

Q3. Holding price increases to a reasonable minimum for the same ad as last year.

Q4. Being appropriately priced for the amount of customers attracted to your business.

Effectiveness:

Q5. Demonstrating to you the potential effectiveness of your advertising purchase.

Q6. Attracting customers to your business through your advertising.

Q7. Reaching a large number of customers.

Q8. Providing long-term exposure to customers throughout the year.

Q9. Providing distribution to the number of households and/or business your business needs to reach.

Q10. Proving distribution to the geographic areas your business needs to reach.

Figure CS3.7 List of questions

Scale Usage Heterogeneity

Figure CS3.8 plots the median over the ten questions versus the range of responses for each of the 1811 respondents. Since all responses are integer, the points are 'jittered' slightly so that the number of respondents at any given combination of range (0–9) and median (1–10) can be gauged. Figure CS3.8 shows considerable evidence of scale usage heterogeneity. A number of respondents are using only the top end of the scale which is represented by points in the lower right-hand corner of the figure. In fact, a reasonably large number give only the top rating response (10) to all questions. On the other hand, there are very few customers who use the lower end of the scale (lower left-hand corner) and a large number who use a much of the scale.

Our hierarchical model should capture the scale usage heterogeneity in the distribution of (τ_i, σ_i). Figure CS3.9 provides histograms of the N posterior means of (τ_i, σ_i). Both τ_i and σ_i display a great deal of variation from respondent to respondent, indicating a high degree of scale usage heterogeneity. Figure CS3.9 shows results for three prior settings, ranging from highly diffuse (top) to tight (bottom). For $\nu = 5$ and $\nu = 20$, the posteriors of both τ_i and σ_i are very similar, indicating little sensitivity to the prior. The hierarchical prior adapts to the information in the data, centering on values of τ_i and σ_i that are shrunk quite a bit. When $\nu = 100$, the prior is reasonably tight about the mean value. This prior has a greater influence and reduces the extent of shrinkage. For example, when $\nu = 100$, there appears to be a small mode in the τ_i distribution around 3. This mode accommodates those respondents who always give the highest rating (10) in answering all questions. More diffuse priors induce more shrinkage from the information in the data, and the τ_i distributions for those respondents that always

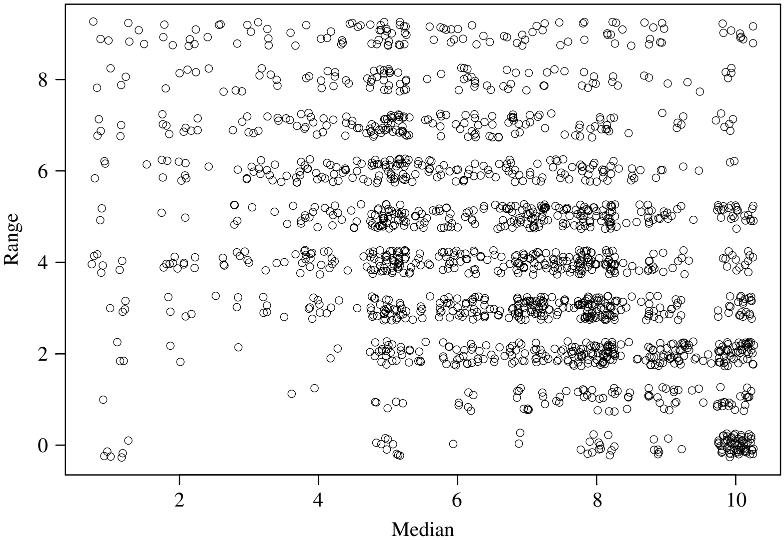

Figure CS3.8 Respondent range versus median (jittered values)

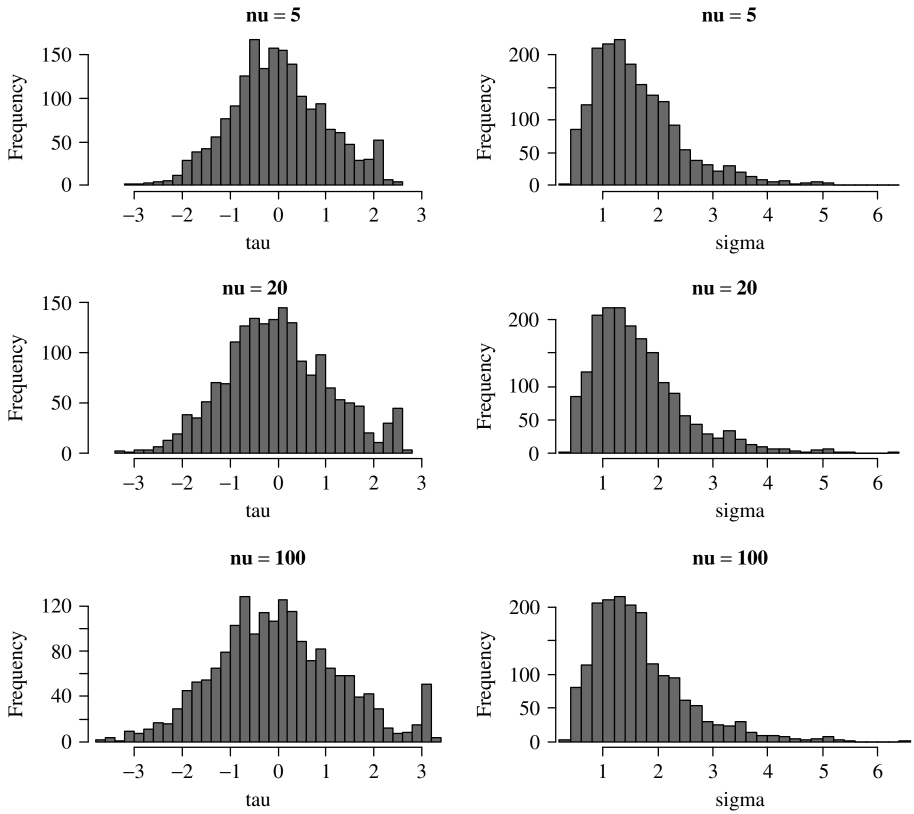

Figure CS3.9 Posterior means of τ_i and σ_i

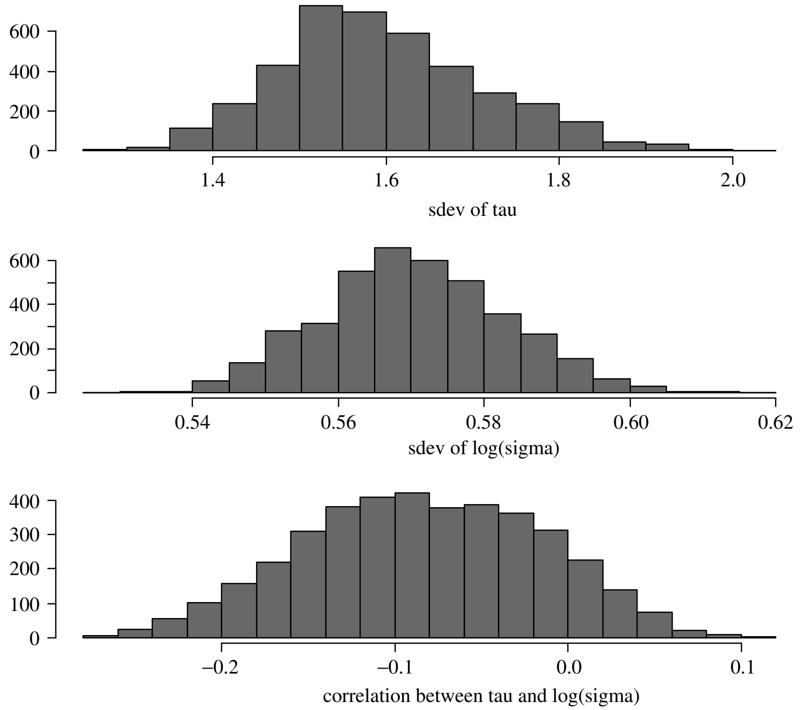

Figure CS3.10 Posterior distribution of Λ

respond with the highest rating are shrunk toward zero. With the higher prior, there is less of this shrinkage.

Figure CS3.10 displays the posterior distribution of Λ. The top panel shows the posterior distribution of $\sqrt{\lambda_{11}}$, the standard deviation of τ_i. This distribution is reasonably tight around 1.55 or so and puts virtually no mass near 2 which corresponds to the location of the prior. Even more striking is the posterior distribution of $\sqrt{\lambda_{22}}$, the standard deviation of $\log(\sigma_i)$. This distribution is centered tightly about 0.57, indicating that the more than 1000 respondents are quite informative about this scale usage parameter. Finally, the posterior distribution of the correlation between τ_i and $\log(\sigma_i)$ puts most of its mass on small negative values, indicating that there is a tendency for those who give high ratings to use less of the scale.

Our model also differs from the standard centering and the standard normal approaches in that we explicitly recognize the discrete nature of the outcome data. Moreover, the discrete outcomes do not appear to be merely a simple filtering of the underlying latent data in which latent data is simply rounded to integers using equal size intervals. The posterior of the quadratic cut-off parameter, e, has a mean of 0.014 with a posterior standard deviation of 0.0039. This implies a skewed set of cut-offs.

Correlation Analysis

One of the major purposes of these surveys is to learn about the relationship between overall satisfaction and various dimensions of product performance. The presence of scale

Table CS3.1 Raw Data means, together with Covariance (lower triangle) and Correlation (upper triangle) matrix

Q.	Mean										
1	6.06	6.50	0.65	0.62	0.78	0.65	0.74	0.59	0.56	0.44	0.45
2	5.88	4.38	7.00	0.77	0.76	0.55	0.49	0.42	0.43	0.35	0.35
3	6.27	4.16	5.45	7.06	0.72	0.52	0.46	0.43	0.46	0.38	0.40
4	5.55	5.36	5.43	5.16	7.37	0.64	0.67	0.52	0.52	0.41	0.40
5	6.13	4.35	3.83	3.62	4.53	6.84	0.69	0.58	0.59	0.49	0.46
6	6.05	4.82	3.29	3.15	4.61	4.61	6.49	0.59	0.59	0.45	0.44
7	7.25	3.64	2.70	2.73	3.42	3.68	3.66	5.85	0.65	0.62	0.60
8	7.46	3.28	2.61	2.79	3.23	3.51	3.41	3.61	5.21	0.62	0.62
9	7.89	2.41	1.99	2.18	2.39	2.72	2.47	3.20	3.02	4.57	0.75
10	7.77	2.55	2.06	2.33	2.42	2.67	2.51	3.21	2.95	3.54	4.89

Table CS3.2 Posterior mean of latent responses (z), together with covariance (lower triangle) and correlation (upper triangle) matrix (Σ); standard deviations in parentheses

Q.	Mean (μ)	Covariance/Correlation (Σ)									
1	6.50	2.46	0.59	0.53	0.73	0.57	0.66	0.45	0.40	0.24	0.26
	(0.08)	(0.29)									
2	6.23		2.98	0.77	0.76	0.47	0.36	0.25	0.26	0.14	0.15
	(0.08)		(0.31)								
3	6.55			3.33	0.69	0.43	0.34	0.27	0.30	0.20	0.23
	(0.08)			(0.32)							
4	6.08				3.28	0.56	0.57	0.37	0.36	0.20	0.21
	(0.08)				(0.35)						
5	6.53					2.82	0.63	0.47	0.47	0.32	0.32
	(0.08)					(0.30)					
6	6.55						2.39	0.49	0.45	0.26	0.26
	(0.08)						(0.30)				
7	7.46							2.93	0.63	0.62	0.60
	(0.08)							(0.28)			
8	7.56								2.43	0.61	0.58
	(0.08)								(0.23)		
9	7.90									2.64	0.78
	(0.08)									(0.25)	
10	7.82										2.76
	(0.08)										(0.26)

usage heterogeneity prevents meaningful use of the raw data for correlation purposes. Scale usage heterogeneity will bias the correlations upward. Table CS3.1 provides the means and correlations for the survey data. The correlations are uniformly positive and in the (0.4, 0.7) range. Table CS3.2 provides posterior means (standard deviations) and correlations of the standardized latent variable, z. Table CS3.2 provides a dramatically

Table CS3.3 Means of standard responses $(x - \bar{x})/s_x$, together with covariance (lower triangle) and correlation (upper triangle) matrix

Q.	Mean										
1	−0.29	0.66	−0.07	−0.13	0.03	−0.14	0.06	−0.11	−0.16	−0.24	−0.21
2	−0.42	−0.05	0.82	0.35	0.20	−0.19	−0.36	−0.32	−0.25	−0.26	−0.27
3	−0.18	−0.10	0.31	0.93	0.14	−0.21	−0.33	−0.33	−0.24	−0.24	−0.22
4	−0.60	0.02	0.14	0.11	0.62	−0.23	−0.17	−0.24	−0.20	−0.26	−0.28
5	−0.28	−0.09	−0.15	−0.18	−0.16	0.76	0.04	−0.07	−0.01	−0.10	−0.11
6	−0.32	0.04	−0.28	−0.27	−0.12	0.03	0.74	0.03	0.03	−0.12	−0.14
7	0.33	−0.08	−0.23	−0.26	−0.16	−0.05	0.02	0.67	0.01	0.06	0.05
8	0.46	−0.09	−0.16	−0.17	−0.12	−0.01	0.02	0.01	0.56	0.01	−0.04
9	0.68	−0.14	−0.17	−0.18	−0.16	−0.07	−0.08	0.04	0.00	0.58	0.31
10	0.61	−0.14	−0.20	−0.18	−0.18	−0.08	−0.10	0.03	−0.02	0.19	0.67

different view of the correlation structure behind the data. In general, the correlations adjusted for scale usage are smaller than those computed on the raw data. This accords with the intuition developed in the discussion of Figure CS3.2. In particular, the correlations between question 2 (price) and questions 9 and 10 (effectiveness in reach) are estimated to be one-half the size of the raw data after adjusting for scale usage heterogeneity. Intuitively, it seems reasonable that those who are very satisfied with (low) price should not also think that the advertising reach is also high. Table CS3.3 provides the correlation matrix of the centered data. Centering the data changes the uniformly positive correlations to mostly negative correlations. It does not seem intuitively reasonable that questions probing very similar aspects of the product should have zero to small negative correlations.

DISCUSSION

There are two challenges facing an analyst of ratings scale data: dealing with the discrete/ordinal nature of the ratings scale; and overcoming differences in scale usage across respondents. In the psychometrics literature, various optimal scaling methods have been proposed for transforming discrete/ordinal data into data that can be treated in a more continuous fashion. However, it is difficult to adapt these methods to allow for respondent-specific scale usage. We adopt a model-based approach in which an underlying continuous preference variable is subjected to respondent-specific location and scale shifts. This model can easily produce data of the sort encountered in practice where some respondents are observed to only use certain portions of the scale (in the extreme, only one scale value). Standard centering methods are shown to be inferior to our proposed procedure in terms of the estimation of relationships between variables.

Our analysis demonstrates that scale usage heterogeneity can impart substantial biases to estimation of the covariance structure of the data. In particular, scale usage heterogeneity causes upward bias in correlations and can be at the source of the collinearity problems observed by many in the analysis of survey data. The covariance structure is the key input to many different forms of analysis, including identification of segments via clustering and the identification of relationships through covariance-based structural modeling. Our procedures provide unbiased estimates of the covariance structure which can then be used as input to subsequent analysis.

While it is well documented that scale usage heterogeneity is prevalent in ratings scale data, not much is known about the determinants of this behavior. Important questions for future research include what sorts of respondents tend to exhibit a high degree of scale usage heterogeneity and how questionnaires and items can be designed to minimize scale usage heterogeneity. Our view is that it is desirable to reduce the magnitude of this phenomenon so as to increase the information content of the data. Whether it is possible to design survey instruments for customer satisfaction that are largely free of this problem is open to question. Ultimately, the answer must come from empirical applications of our model that will detect the extent of this problem.

R IMPLEMENTATION

R The R implementation of this method is embodied in the function, `rscaleUsage`, which is included in *bayesm*. We made a few changes to the basic algorithm in Rossi *et al.* (2001). First, we did not use an independence Metropolis step for drawing σ_i, instead we used a relatively fine grid over a relatively large interval. Second, we used a uniform prior for e on a grid on the interval $(-0.1, 0.1)$ and we used a random walk on this grid (moving to left and right grid points with equal probability except at the boundary where we propose the next innermost point with probability one). We use 100 replications to compute the GHK estimates of normal probabilities. We use 500 as the default number of grid points for the griddy Gibbs parameter draws (σ_i, e,Λ). We use the default prior settings given in (CS3.9), (CS3.10), and (CS3.12).