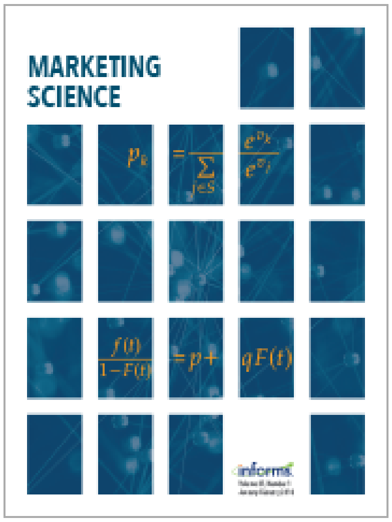


Marketing Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

The Dimensionality of Customer Satisfaction Survey Responses and Implications for Driver Analysis

Joachim Büschken, Thomas Otter, Greg M. Allenby,

MARKETING SCIENCE
Vol. 32, No. 4, July–August 2013, pp. 533–553
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.2013.0779

The Dimensionality of Customer Satisfaction Survey Responses and Implications for Driver Analysis

Joachim Büschken
Catholic University of Eichstätt-Ingolstadt, 84059 Ingolstadt, Germany, joachim.bueschken@ku.de

Thomas Otter
Goethe University Frankfurt, 60629 Frankfurt am Main, Germany, otter@marketing.uni-frankfurt.de

Greg M. Allenby
Fisher College of Business, Ohio State University, Columbus, Ohio 43210, allenby.1@osu.edu

The canonical design of customer satisfaction surveys asks for global satisfaction with a product or service and for evaluations of its distinct attributes. Users of these surveys are often interested in the relationship between global satisfaction and attributes; regression analysis is commonly used to measure the conditional associations. Regression analysis is only appropriate when the global satisfaction measure results from the attribute evaluations and is not appropriate when the covariance of the items lie in a low-dimensional subspace, such as in a factor model. Potential reasons for low-dimensional responses are that responses may be haloed from overall satisfaction and there may be an unintended lack of item specificity. In this paper we develop a Bayesian mixture model that facilitates the empirical distinction between regression models and relatively much lower-dimensional factor models. The model uses the dimensionality of the covariance among items in a survey as the primary classification criterion while accounting for the heterogeneous usage of rating scales. We apply the model to four different customer satisfaction surveys that evaluate hospitals, an academic program, smartphones, and theme parks, respectively. We show that correctly assessing the heterogeneous dimensionality of responses is critical for meaningful inferences by comparing our results to those from regression models.

Key words: Bayesian estimation; surveys; information processing
History: Received: June 4, 2010; accepted: January 2, 2013; Eric Bradlow and then Preyas Desai served as the editor-in-chief and Fred Feinberg served as associate editor for this article. Published online in *Articles in Advance* April 8, 2013.

1. Introduction

Managers use customer satisfaction surveys to measure levels of satisfaction and study drivers of satisfaction. Measures of levels of satisfaction distinguish between more and less satisfied customers, serve as benchmarks against the competition, and are useful for tracking customer satisfaction through time (Fornell et al. 1996). Satisfaction driver research relates temporal or cross-sectional variation in overall satisfaction to variation in the evaluation of conceptually independent components of a product or service. Surveys are designed to elicit evaluations of the components of a product or service and an overall measure of satisfaction. Driver analysis uses the covariance among evaluations to estimate the conditional relationship of the overall measure to its components. The standard approach relies on regression analysis and, more recently, finite mixture regression analysis (e.g., Wedel and DeSarbo 2002).[1] The conceptual advantage of regression analysis is that the estimated coefficients measure the conditional effects of changing one component score while holding all other component scores constant. In practice, however, it is an open question whether regression analysis is meaningful for a specific data set.

Despite a firm's best efforts and known constraints on respondent time, a specific survey may accidentally fail at measuring conceptually independent component evaluations and effectively ask about total satisfaction over and over again. In practice, regression analysis and other techniques for identifying drivers of satisfaction applied to such data will result in rapidly decreasing estimates of driver importance as the number of "predictor" items increases. This is due to the property of a regression coefficient as a conditional effect given the other covariates and due to

[1] When these surveys contain explicit multi-item measurement models, driver analysis uses the covariance among latent variables to infer conditional relationships between latent component evaluations and latent overall satisfaction; latent variable regression models, e.g., LISREL, are used. In this paper we focus on intended single-item measurements of component evaluations and overall satisfaction. However, our conceptual discussion applies straightforwardly to the analysis of latent covariance structures.

the item scores each reflecting the *same* information (up to error) about total satisfaction. We provide a numerical illustration in the Web appendix (available at http://dx.doi.org/10.1287/mksc.2013.0779).

Some respondents may fail to think clearly about the component evaluations, whereas others are able and motivated to provide separate evaluations for each component. In this case the joint distribution of evaluations is a mixture of low- and high-dimensional responses, and regression or other types of driver analysis should be informed by the high-dimensional responses only. However, low-dimensional responses can still be informative about levels of satisfaction. Alternatively, some respondents may provide random responses that indiscriminately fluctuate around some anchor point on the scale. Such responses are collectively outliers and do not contribute any substantively interesting information. In fact, the covariance among these responses is neither low nor high dimensional but zero. Technically, these responses are still informative about item-specific response levels but generally not about levels of satisfaction.[2]

The vast majority of customer satisfaction surveys rely on rating scales. These scales are assumed to result in ordinal data at the individual-respondent level, up to measurement error. The basic axioms of measurement (e.g., Krantz et al. 1971) rule out a direct comparison of ordinal relationships among respondents. For example, a two-scale point difference between items reported by one respondent may equate to a one-scale point difference reported by another respondent. Both responses only indicate that the two respondents' preferences are in the same order. However, when each respondent uses the same ordinal scale across multiple items, it is possible to characterize his preference for ordinal categories and separate out scale use effects that can mask the dimensionality of substantive preferences (Rossi et al. 2001, Van Rosmalen et al. 2010, Ying et al. 2006).

In this paper we develop a Bayesian mixture model for customer satisfaction surveys that uses the dimensionality of covariances among responses as the primary classification criterion while accounting for heterogeneous scale use.[3] The model probabilistically sorts responses according to the dimensionality of their covariance, distinguishing responses with a full covariance structure from those with a lower dimension. This way, driver analysis can be based on the covariance of full-dimensional responses where regression is a priori meaningful. It may still turn out to be managerially uninformative a posteriori, depending on the estimated regression coefficients, and the degree of prior management knowledge. However, the model identifies and thus controls for low-dimensional covariance structures where the application of regression lacks meaning a priori. Similarly, the model identifies and controls for responses with zero covariance.

We apply our model to four different customer satisfaction surveys that evaluate hospitals, an academic program, smartphones, and theme parks, respectively. We find that only one data set, the smartphone survey, essentially conforms to a regression model. For the theme park survey, the model determines that almost all responses are best accounted for by a one-factor model. The ability to obtain positive support for a much lower-dimensional model than the unstructured, high-dimensional covariance underlying the regression model is unique to the Bayesian approach. The two remaining data sets result in a mixture of a regression model and low-dimensional factor models. In all cases the proposed classification according to the dimensionality of responses results in intuitively meaningful inferences and fits the data better than ordinal regression models and mixtures of ordinal regression models.

The organization of the rest of this paper is as follows. We discuss prior research in §2. We develop the model in §3. In §4, we report results from four empirical applications. We provide a discussion of our results and future research topics in §5.

2. Prior Research

Methods to determine the dimensionality of a set of questionnaire items, or more generally, a set of measurements, originated in psychology and psychometrics and have found applications across a variety of disciplines, including marketing (e.g., Beckwith and Lehmann 1975, Iacobucci 1994, Sonnier and Ainslie 2011) and finance (e.g., Jones 2006). A fundamental assumption of these methods is that some fixed, possibly unknown, number of latent dimensions or factors—smaller than the number of measurements taken—generates the covariance among the observed measurements. From the earliest work in exploratory factor analysis (Spearman 1904) to sophisticated, detailed confirmatory factor models (e.g., Gibbons and Hedeker 1992, Bradlow et al. 1999, Wang et al. 2002), the hypothesis of a smaller number of latent dimensions is linked to principles of constructing the measurements. The individual items on the questionnaire are formulated with the goal of tapping into

[2] Response patterns with zero substantive (i.e., net of scale usage) covariance likely result from a lack of motivation or ability to process the items on the survey.

[3] When surveys use multi-item scales where individual items are reflective indicators of latent constructs, the classification criterion is the dimensionality of the covariance among latent constructs. However, we leave the technical development of our conceptual argument in a survey context with explicit measurement models for future research.

hypothesized latent dimensions, and each dimension is represented by multiple items (Thurstone 1937). As a consequence, a general assumption is that the latent dimensionality of a particular set of measurements is homogeneous in a population.

The design of many commercial customer satisfaction surveys proceeds differently. Designers often face severe constraints on respondent time and motivation—they need to keep surveys short and simple while trying to maximize information about a multitude of individual service or product components and their relationship to overall satisfaction. In contrast to psychological tests, the implicit intent of many commercial customer satisfaction surveys is therefore to elicit evaluations of dimension equal to the number of items in the questionnaire so that driver analysis can be conducted.[4] In Rossiter's C-OAR-SE terminology, applied customer satisfaction surveys measure how customers feel about a multitude of concrete attributes (Rossiter 2002); the goal is to collect *qualitatively* different information with each additional question on the survey.[5]

From the point of view of psychological test development, or more generally, psychometric measurement, this practice may appear strange. However, whereas psychometric measurement eventually asks for reliable estimates of individuals' scores along well-defined prior dimensions such as "analytical ability," customer satisfaction surveys aimed at driver analysis emphasize the discovery of relationships between component evaluations and overall satisfaction. And in many of these customer satisfaction applications, the goal of highly reliable posterior estimates of latent true scores has to be traded off against the measurement of additional putative drivers of satisfaction. In a sense, psychometrics has an inherent preference for items with low unique variance relative to other items measuring the same latent dimension, whereas customer satisfaction research aimed at driver analysis implicitly maximizes the uniqueness of items in an attempt to cover as many conceptually independent drivers as possible.[6]

A basic goal of marketing research is to identify drivers of some desired overall outcome (e.g., overall satisfaction using regression models). In contrast to psychometric models that are built around the idea of theory-guided dimensionality reduction, driver analysis in marketing seeks to understand how hypothetical changes in individual component evaluations translate into changes in overall satisfaction. More recently, this approach has been generalized to account for heterogeneity in the coefficients that convert changes in component evaluations into changes in total satisfaction (Wedel and DeSarbo 2002, Wedel and Kamakura 2000).[7]

Valid answers to the question of how hypothetical changes in individual component evaluations translate into changes in overall satisfaction require that responses to component items on a customer satisfaction survey are "unique" in the sense that their covariance *cannot* be explained by a relatively small number of underlying latent dimensions.[8] Here, psychometrics and psychology, guided by strong theories about the maximum dimensionality behind the covariance of a set of measurements, rely on hypothesis testing. The null hypothesis corresponds to the theoretical number of dimensions, and empirical results that *fail to reject* this null hypothesis are taken as support for the theory (Jöreskog 1967).

In contrast, driver analysis in customer satisfaction research seeks positive evidence that responses are of dimension equal to the number of questions asked. This is different from rejecting hypotheses about a lower-dimensional representation (Kass and Raftery 1995). The difference becomes critical in the context of comparing the saturated model with the number of dimensions equal to the number of items to models that posit a lower-dimensional representation. Driver analysis should not "wait for" the rejection of low-dimensional representations but build on positive evidence for unique items, i.e., a covariance structure of dimensionality equal to the number of items on the survey. Similarly, because it is impossible to reject the saturated model in favor of a more constrained model without imposing a penalty for the number of parameters, tools are needed that build on positive evidence in favor of a lower-dimensional representation before

[4] However, a model that motivates observed responses in customer satisfaction surveys from underlying latent variables of lower dimensionality similar to psychometric testing models was presented by Bradlow and Zaslavsky (1999).

[5] We are aware of the fact that (mostly) academic studies often use multiple items for individual, more abstract attributes with the intent of reducing measurement error (e.g., Smith et al. 1999). These studies are based on strong prior knowledge about relationships among limited numbers of latent variables. In contrast, standard driver analysis aims at the "discovery" of the relative importance of a larger number of putative drivers.

[6] From a measurement perspective, driver analysis attempts to establish a purely formative measurement model. The reliability of such a model increases in the number of "causal" indicators with significant (conditional) effects on the latent variable, i.e., overall satisfaction in this context (e.g., Bollen 1989).

[7] When the goal is to understand heterogeneity in conditional relationships among latent variables—a case not covered in this paper—the analysis relies on finite mixtures of structural equation models.

[8] If this condition is not met, the analyst will run into multicolinearity problems, as demonstrated in the Web appendix. Practitioners and academics alike very often find signs of multicolinearity among putative drivers in these analyses (e.g., Finn 2011).

we can learn about the potentially heterogeneous dimensionality of responses in a particular data set.

The possibility that a given set of items on a questionnaire or survey is of different inherent dimensionality in different "populations" has received limited empirical attention both in marketing and psychology. Although discrete mixtures of factor models have found applications in psychology (e.g., Lubke and Muthén 2005), the focus has been on relaxing the distributional assumptions of factor scores. The idea is that the joint distribution of responses to a measurement invariant test is caused by an unobserved mixture of discrete "types" that each consist of continuously heterogeneous respondents. Measurement invariance requires that the way a latent variable expresses itself in observable measures is population invariant and therefore, by definition, requires a fixed dimensionality of the measurements (Meredith 1993).

However, the theory of survey design (e.g., Groves et al. 2004, Hayes 2008) suggests various causes for survey responses that are mixtures of distributions with different covariance dimensionality. Potential reasons for low-dimensional responses to a set of, in principle, specific, unique items include halo effects from overall satisfaction, a lack of motivation to process the individual items in detail, or some combination of these.

The concept of a halo goes back to Thorndike (1920), who described it as the tendency to think of a person in general as rather "good" or "inferior" and to color the judgment of specific personal qualities by this general feeling. Thus, reporting a haloed response to a specific question is an instance of the more general concept of response substitution (Gal and Rucker 2011). Both the theory of halo and that of response substitution, as well as theories about the motivation to process individual items in detail, are formulated at the level of individual respondents. Therefore it is a valid empirical question if a sample of responses to a particular survey is of heterogeneous dimensionality. And, in a sense, an answer to the well-established question about the inherent dimensionality of a set of measurements is incomplete without accounting for the possibility that different respondents may respond to the same set of items with differing underlying dimensionalities.

Technically, we build on the literatures on finite mixtures (Frühwirth-Schnatter 2006), factor analysis (Mulaik 1972), and models for rating scales (Johnson and Albert 1999). We develop Bayesian exploratory factor analysis (Press and Shigemasu 1989) as our model of reduced dimensionality and use finite mixtures to capture the heterogeneity in the dimensionality of responses. We need to account for heterogeneous scale use (Cunningham et al. 1977, Rossi et al. 2001, van Rosmalen et al. 2010, Ying et al. 2006) when investigating potentially heterogeneous dimensionalities of survey responses. A scale usage model is essentially a device to distinguish between substantive and nonsubstantive covariance in a particular data set, and we are interested in the structure of the substantive covariances. Joint identification of substantive and nonsubstantive covariance comes from the definition of a scale use model in any model of scale usage.

In these models, scale use is a respondent trait (in a given survey) that applies indiscriminately to every item on that survey. Therefore, models of scale use per se cannot explain variability across items within a respondent. The variance and covariance of responses within respondents then identifies the model of substantive interest. The scale use model picks up variation on a collection of (practically) identical scales across respondents that is orthogonal to any differences between items. Therefore, joint identification of substantive and nonsubstantive covariances hinges on survey items that are consistently perceived as different from each other.

An intuitive example for consistent differences between items is the combination of two "positive" items and one "negative" item. Depending on the substantive model, one might expect positive covariance between the positive items and negative covariance between each of the positive items and the negative item. Models of scale use can motivate neither the negative covariances in this example nor the differences in the covariances between different items because they do not distinguish among items by definition. More generally, any difference in the response distribution across items cannot be attributed to heterogeneous scale usage and thus identifies the substantive model. We develop our model formally in the next section.

3. Mixture Model for Customer Evaluations

We first derive the model from continuous latent variables that generate the observed ordinal discrete responses, e.g., five-point or seven-point rating scales, through a link in the form of a step function. We then discuss how to incorporate heterogeneous scale use effects (Rossi et al. 2001, Ying et al. 2006).

3.1. Mixture Model

We assume that observed customer evaluations provided on a fixed-point scale arise from unobserved continuous evaluations. The latent evaluations are denoted as

$$\mathbf{z}'_i = (z_{y,i}, \mathbf{z}'_{x,i}) \tag{1}$$

consisting of an overall score (y) and a component score (x) assumed to be distributed multivariate normal with the covariance matrix dependent on the dimensionality of responses. The index i is for respondents. The latent continuous evaluations are not observed. We assume that observed censored realizations are generated from latent continuous evaluations according to a cut-point model. Let y_i and x_{ij}, $j = 1, \ldots, q$, denote the observed integer responses for respondent i, where

$$y_i = k \quad \text{if } c_{i,k-1} \le z_{y,i} < c_{i,k}, \tag{2}$$

$$x_{i,j} = k \quad \text{if } c_{i,k-1} \le z_{x,i,j} < c_{i,k}, \tag{3}$$

for $K+1$ respondent-specific ordered cut points $\{c_{i,k}: c_{i,k-1} \le c_{i,k} k = 1, \ldots, K\}$, where $c_0 = -\infty$ and $c_K = \infty$. Thus, the observed responses (y_i, $\mathbf{x}_i'$) follow an ordered probit model as in Rossi et al. (2001).

We use unconstrained, "exploratory" linear factor models to account for responses with low-dimensional covariance. For such low-dimensional responses, the evaluations $\mathbf{z}_i$ arise from a lower-dimensional vector of latent overarching evaluations ξ_i. We indicate the length of this vector by h with $h < q+1$, i.e., smaller than the number of items on the questionnaire. The relationship between evaluations and latent scores is

$$\mathbf{z}_i = \mathbf{\Gamma}^h \xi_i + \mathbf{c}^h + \boldsymbol{\varepsilon}_i^h. \tag{4}$$

Here, $\mathbf{\Gamma}$ is a coefficient matrix of factor loadings of dimension $(q+1) \times h$, $\mathbf{c}^h$ is a vector intercept, and $\boldsymbol{\varepsilon}_i^h$ are unique and independently distributed normal errors with possibly different variances. Without loss of generality, we assume that the latent overarching evaluations ξ_i are distributed independently, multivariate normal; $\xi_i \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_h)$. The joint distribution of $\mathbf{z}$ is then

$$\mathbf{z}_i \sim \mathrm{N}(\mathbf{c}^h, \mathbf{\Gamma}^h \mathbf{\Gamma}^{h'} + \mathbf{\Psi}) = \mathrm{N}(\boldsymbol{\mu}^h, \mathbf{\Sigma}^h), \tag{5}$$

where $\mathbf{\Psi}$ is a diagonal matrix of error variances.[9]

In contrast, responses that are a priori consistent with a regressions model exhibit a "full" covariance structure that cannot be projected into a space of smaller dimensionality than the number of items on the questionnaire. One way to capture these responses is to simply estimate a full, unrestricted covariance matrix. However, in our application we are eventually interested in the regression coefficients for potentially meaningful driver analysis, a posteriori, i.e., given that fitting a regression model is meaningful a priori as per the dimensionality of the observed responses. Therefore, we derive the full, unrestricted covariance matrix from a regression model:

$$z_{y,i} = \mathbf{z}_{x,i}' \boldsymbol{\beta} + c_y^f + \varepsilon_i^f. \tag{6}$$

Here, c_y^f is a scalar intercept. The vector of predictors $\mathbf{z}_{x,i}$ does not include a constant and the intercept c_y allows for a location shift in the expected value of the overall evaluation (y) relative to the components (x). The joint distribution of $\mathbf{z}_i$ implied by the regression model is then

$$\mathbf{z}_i \sim \mathrm{N}\left(\begin{pmatrix} (\mathbf{c}_\mathbf{x}^f)' \boldsymbol{\beta} + c_y^f \\ \mathbf{c}_\mathbf{x}^f \end{pmatrix}, \begin{pmatrix} \boldsymbol{\beta}' \mathbf{\Sigma}_\mathbf{x} \boldsymbol{\beta} + \sigma_{\varepsilon_y}^2 & \boldsymbol{\beta}' \mathbf{\Sigma}_\mathbf{x} \\ \mathbf{\Sigma}_\mathbf{x} \boldsymbol{\beta} & \mathbf{\Sigma}_\mathbf{x} \end{pmatrix} \right) = \mathrm{N}(\boldsymbol{\mu}^f, \mathbf{\Sigma}^f), \tag{7}$$

where $\mathbf{c}_\mathbf{x}^f$ and $\mathbf{\Sigma}_\mathbf{x}$ are the marginal means and variance–covariance of $\mathbf{z}_x$, respectively, i.e., of the putative drivers of the overall evaluation. The parameter $\sigma_{\varepsilon_y}^2$ denotes the conditional variance of z_y given $\mathbf{z}_x$.

Equation (7) is equivalent to estimating an unrestricted, full covariance matrix. Therefore the classification of responses is based on the inherent dimensionality of response vectors where (5) posits covariances of lower dimensionality, h, than the number of items on the questionnaire, $q+1$. Another implication is that all factor models in (5) are special, restricted cases of Equation (7). When h in (5) approaches $q+1$ (i.e., when the number of overarching latent evaluations approaches the number of items on the questionnaire)—specifically, when $h \ge \lceil \frac{1}{2}(2(q+1) + 1 - \sqrt{8(q+1)+1}) \rceil$ (Ledermann 1937)—Equation (5) no longer constrains the dimensionality of the covariance of questionnaire items.[10] In this case, the mixture is defined by different covariances (and means) of equal dimensionality. The popular finite mixture of regression approach (Wedel and Kamakura 2000), for example, assumes equally full dimensionality of covariances in different mixture components a priori.[11]

Finally, we account for respondents who provide random responses y and $\mathbf{x}$ that indiscriminately fluctuate around some anchor point on the scale, after accounting for heterogeneous scale usage discussed in §3.2.[12] Such responses are collectively outliers and do not contribute any substantively interesting information. In fact, the covariance among these responses is neither low nor high dimensional but zero. These observations obfuscate any covariance-based inference such as regression analysis, factor analysis, or covariance-based mixture models.

[9] We do not impose prior identification constraints on the factor loadings and discuss the identification of factor loadings $\mathbf{\Gamma}$ in the appendix.

[10] Here, $\lceil\,\rceil$ denotes the "ceiling," i.e., the smallest integer that is greater than or equal to the argument.

[11] Obviously, a regression model or a mixture of regressions can be fit to data with a less than full-dimensional covariance structure. The point is that doing so lacks meaning a priori because a low-dimensional covariance is inconsistent with individual component evaluations driving a global evaluation.

[12] When combined with our model of heterogeneous scale use, our independence model can capture any observed systematic response such as indiscriminate alternations between "high" and "low" responses, which are independent of systematic differences between response items, as perceived by the mixture population of more informative, discriminate responses.

Independent evaluations, characterized by a diagonal covariance matrix, are special cases of both models in (5) and (7). The factor model in (5) produces independent responses when the matrix of factor loadings $\boldsymbol{\Gamma}$ contains only zeros. The regression model in (7) produces independent responses when all regression coefficients are equal to zero and $\boldsymbol{\Sigma}_{\mathbf{x}}$ is diagonal. However, when the data are a mixture of independent responses and dependent responses of different dimensionality, the former constitute outliers with respect to both models (5) and (7). We model the joint distribution of independent evaluations as

$$\mathbf{z}_i \sim \mathrm{N}(\boldsymbol{\mu}^u, \Psi^u), \tag{8}$$

where Ψ^u is a diagonal matrix of unique variances.

We assume that the latent continuous evaluations $\mathbf{z}_i$ are marginally distributed according to a component mixture model:

$$p(\mathbf{z}_i) = \sum_{h=1}^{H} \eta^h p^h(\mathbf{z}_i) + \eta^f p^f(\mathbf{z}_i) + \eta^u p^u(\mathbf{z}_i), \tag{9}$$

where η is the mixture probability denoting segment size; and the superscripts correspond to the factor models in (5), the regression model in (7), and the independence model in (8). H refers to the factor model with the largest number of factors included in the mixture. Equation (9) implies that a random observation is drawn with probability η^h from the factor model with h factors, probability η^f from the regression model, and probability η^u from the independence model.

When a survey elicits overall and partial evaluations for multiple objects, as in two of our empirical examples, we draw latent component membership indicators conditionally independently for each object evaluated by a respondent. Therefore, the structure by which responses occur is a priori the result of an interaction between an object and the respondent rather than a respondent trait. In contrast, scale usage, to be discussed next, is by definition a respondent trait.

3.2. Scale Usage

Respondents are often observed to use different portions of the scale, with yea-sayers employing the upper end of the scale and naysayers using the bottom portion. Some respondents are observed to use a narrow range while others use the entire range of the scale. Heterogeneous scale use causes responses to covary across respondents, similar to any other "trait" that manifests itself across multiple observations.[13] This nonsubstantive covariance masks the model that generated the responses and thus interferes with classification based on the inherent dimensionality of responses.

[13] Therefore, the discrete *observed* response vectors in our independence component are not independent, but the latent evaluations after correcting for heterogeneous scale use are.

The goal of any form of scale adjustment is to standardize the data so that inference can proceed among respondents based on within-respondent variance (Rossi et al. 2001). Alternatively, one can think of scale adjustments as changing the cutoff's $\{c_k\}$ among respondents (Ying et al. 2006). To ensure that our structural inferences are not driven by heterogeneous scale usage, we employ a variant of the general model suggested by Ying et al. (2006). The authors propose the following model of respondent-specific cut points for the ordinal probit model for a rating scale with K categories:

$$\text{(YFW)}\quad \begin{aligned} \mathbf{c}_i &= (c_{1,i}, c_{2,i}, \ldots, c_{K-1,i}) \\ &= \bigg(c_{1,i}, c_{2,i} = c_{1,i} + \Delta_{1,i}, c_{3,i} = c_{1,i} \\ &\qquad + \sum_{k=1}^{2} \Delta_{k,i}, \ldots, c_{K-1,i} = c_{1,i} + \sum_{k=1}^{K-2} \Delta_{k,i}\bigg), \\ &\log(\boldsymbol{\Delta}_i) \sim \mathrm{N}(\boldsymbol{\mu}_\Delta, \boldsymbol{\Sigma}_\Delta), \end{aligned} \tag{10}$$

and $c_0 = -\infty$ and $c_K = \infty$ for all i. For notational simplicity, we write $(c_{1,i}, \Delta_{1,i}, \ldots, \Delta_{K-2,i})$, the vector of cut-point vector baseline and increments, as $\boldsymbol{\Delta}_i$. The model without scale use heterogeneity is obtained as the limit where $\boldsymbol{\Sigma}_\Delta \to 0$. Equation (10) is extremely flexible but requires estimating $K-1$ cut-point parameters per respondent. With sparse likelihood information, the log-normal hierarchical prior on the cut-point increments is likely to be influential. Therefore, we extend the YFW model by specifying a mixture of log-normal prior distributions:

$$\begin{aligned} \log(\boldsymbol{\Delta}_i) &\sim \mathrm{N}(\boldsymbol{\mu}_{\Delta,\, ind_i}, \boldsymbol{\Sigma}_{\Delta,\, ind_i}), \\ ind_i &\sim \mathrm{Multinomial}_L(\boldsymbol{\eta}_{ind}). \end{aligned} \tag{11}$$

The indicator variable ind takes on values $1, \ldots, L$, and $\boldsymbol{\eta}_{ind}$ is a vector of mixture probabilities. The number of mixture components L is determined by comparing models with different specifications.

Given the parameters

$$(\{c_{i,k}\}, \eta, (\boldsymbol{\mu}^{h=1}, \boldsymbol{\Sigma}^{h=1}), (\boldsymbol{\mu}^{h=2}, \boldsymbol{\Sigma}^{h=2}), \ldots, (\boldsymbol{\mu}^{h=H}, \boldsymbol{\Sigma}^{h=H}), (\boldsymbol{\mu}^f, \boldsymbol{\Sigma}^f), (\boldsymbol{\mu}^u, \boldsymbol{\Sigma}^u)),$$

the likelihood of the observed data is a mixture of normal integrals over the rectangular region defined by the cutoff points. To facilitate estimation, we employ data augmentation (Tanner and Wong 1987). The appendix summarizes our Markov chain Monte Carlo (MCMC) sampler and provides detailed information about subjective prior choices.

3.3. Determining the Number of Mixture Components

Mixture models formalize the trade-off between decreasing the amount of data to inform parameters in a particular component and uncovering additional structure in the data that reflect respondent heterogeneity. However, the maximum of the likelihood is (at least) weakly increasing in the number of components. Therefore, the number of relevant components is determined using penalized likelihood measures or information criteria and cross validation in frequentist settings (Frühwirth-Schnatter 2006).

In a Bayesian mixture model, subjective priors structure the trade-off between increasing the in sample likelihood and decreasing the amount of data to inform parameters in individual components, i.e., losing efficiency. The subjective prior for the component sizes η plays a special role in this context (see the appendix for a complete list of subjective prior settings). We use a (standard) weakly informative Dirichlet prior, Dir$(3, \ldots, 3)$. This prior expresses our lack of prior knowledge about component sizes, including the possibility that one or more components of our mixture model are not needed, i.e., have a true size equal to zero.

Our MCMC sampler explores the posterior and, by construction, visits regions with higher posterior support more often. For example, allocating homogeneous observations, equally split, between two different components instead of one decreases the *posterior* support of the data under the model. This is because the posterior of the component-specific parameters is less concentrated than in that state where (homogeneous) observations are allocated to one component, leaving the second component empty. As explained next, the empty component is essentially irrelevant for both the *posterior* and the *prior* support of the data under the model.

Once all observations are allocated to one component, the penalty for carrying the empty component forward exacted by the (weakly informative) Dirichlet prior is by definition, minimal.[14] Therefore, MCMC with a weakly informative subjective prior on component sizes can be used to implicitly select the number of empirically relevant components (Frühwirth-Schnatter 2006). This property of Bayesian mixture models is related to the observation that marginal likelihoods for models with a posteriori empirically irrelevant (i.e., empty) components are essentially identical to those from the correspondingly reduced model, conditional on the posterior distribution of group sizes.[15]

This approach toward component selection works well as long as the maximum number of (potentially empty) components under consideration is reasonably small, which is the case in our applications. When the maximum number of components becomes very large, any Dirichlet prior of fixed dimensionality is informative, and transdimensional MCMC algorithms are called for (e.g., Green 1995). We report the results of various tests of our MCMC sampler using simulated data in the Web appendix. We confirm the ability of our MCMC to identify the correct (smaller) number of components when we prespecify too many components in the model.

4. Empirical Application

4.1. Data Sets

We investigate the empirical performance of our mixture model using four customer satisfaction data sets. Incomplete responses were deleted from all data sets prior to any analysis. Incomplete responses contain at least one missing satisfaction evaluation. The first data set contains the evaluation of hospitals in a major metropolitan area of the United States. A sample of 383 randomly chosen respondents from three neighboring zip codes provided complete evaluations of hospitals in their area. Of those, 203 evaluated two hospitals, and 180 respondents evaluated only a single hospital; 3.8% of respondents in this data set provided incomplete responses. The hospitals were evaluated on the basis of whether the hospital would be the respondent's first choice for treatment. Potential drivers included various service attributes of the respondent's hospitalization. The specific wording of questions is provided in Table 1. All evaluations were measured on a nine-point ordinal rating scale ranging from 1 ("strongly disagree") to 9 ("strongly agree"). Respondents only evaluated hospitals with which they were familiar.

The second data set was obtained from students enrolled in a business bachelor's program at a German university. Global and partial satisfaction with the program was measured using the criteria described in

[14] If the penalty from the subjective prior were large, empty components could not be observed.

[15] Note that the dimensionality, or, more generally, the subjective prior on parameters other than the component size in empty components, does not affect the marginal likelihood. For example, in the case of two components with identical component-specific priors,

$$\iint \prod_{i=1}^{N} (1 \cdot p(y_i \mid \theta^1) p(\theta^1) + 0 \cdot p(y_i \mid \theta^2) p(\theta^2))\, d\theta^1\, d\theta^2 = \int \prod_{i=1}^{N} p(y_i \mid \theta^1) p(\theta^1)\, d\theta^1.$$

Moreover,

$$\iint \prod_{i=1}^{N} (0.5 p(y_i \mid \theta^1) p(\theta^1) + 0.5 p(y_i \mid \theta^2) p(\theta^2))\, d\theta^1\, d\theta^2 \le \int \prod_{i=1}^{N} p(y_i \mid \theta^1) p(\theta^1)\, d\theta^1$$

when the data are from one component, i.e., homogeneous. The inequality is strict for finite samples.

Table 1 Attributes and Descriptive Statistics for the Hospital Data Set

Attribute	Format in survey
Overall evaluation	Question
First choice	"If I needed medical tests or treatment, this hospital would be my first choice"
Partial attributes	Question
Language	"Has doctors and employees who speak my language"
Puts patients first	"Has doctors and employees that put the needs of patients first"
Safe place	"It is a safe place to be"
Conveniently located	"It is conveniently located"
Skilled doctors	"Has doctors who are highly skilled in their areas of specialty"
Modern	"It is modern looking and up-to-date"
Clean	"Has a clean environment"
Concerned	"This organization is very concerned about the welfare of people like me"

	Descriptive statistics for hospital data[a]			
Attribute	*N*	Mode	Mean	SD
Global evaluation: First choice	586	8	6.65	2.76
Partial evaluations				
Language	586	8	7.58	2.01
Puts patients first	586	8	7.11	2.22
Safe place	586	8	7.40	2.16
Conveniently located	586	8	7.62	2.04
Skilled doctors	586	8	7.56	2.04
Modern	586	8	7.65	1.88
Clean	586	8	7.72	1.88
Concerned	586	8	6.88	2.34

[a]On a scale of 1 (strongly disagree) to 9 (strongly agree).

Table 2. The program started in 2005, and a total of 545 students were enrolled in it at the time the data were collected in early 2008. Out of this population, 258 students participated in the survey. For all satisfaction measures, a five-point scale was used. None of the respondents in this data set provided incomplete responses.

The third data set comprises satisfaction evaluation of U.S. smartphone users. Evaluations were obtained online through the use of an online smartphone user panel in 2011. Six hundred sixty-two users participated in the study and provided complete responses, with each respondent evaluating his or her smartphone brand and model. As global scores, we used participants' likelihoods to recommend their smartphone to friends or colleagues. As partial scores, we used the evaluation of smartphone attributes used in a recent J.D. Power and Associates smartphone user satisfaction survey (see Table 3).[16]

Finally, we use a data set comprising evaluations of Florida family vacation destinations. In this data set, 542 complete evaluations were obtained from 311 respondents who had visited at least one of the following destinations within the past five years: Disney Parks (Magic Kingdom, Epcot, Animal Kingdom, and Hollywood Studios), Universal Studios Florida and Universal's Islands of Adventure, and Busch Gardens. The respondents evaluated a maximum of two destinations, given that a visit to both had occurred. The evaluations comprise 17 brand belief dimensions that were evaluated on a seven-point scale ranging from "does not describe it at all" to "describes it very well" (see Table 4). Two percent of the responses in the vacation data set were incomplete and deleted.

Figure 1 provides a plot of the data. The left upper panel plots the hospital data, the left lower panel the student data, the upper right panel the Florida vacation data, and the lower right panel the smartphone data. The median response is on the horizontal axis, and the range of responses on the vertical axis. Each point represents one response vector. The points on the discrete median–range grid are slightly jittered to illustrate their distribution.

The hospital data and the smartphone data exhibit the stylized features of many customer satisfaction data sets. Many response vectors are concentrated at the upper, positive end of the scale with only small response variation across items. Students also tend toward a positive median response, but very few median responses are in the top box. Moreover, students show more response variation across items after taking into account that the hospital and the smartphone data were collected using a nine-point scale and the student data were collected using a five-point scale. In the hospital data, for example, 33% of the

[16] See J.D. Power and Associates (2012) for a summary of the results from this survey.

Table 2 Attributes and Descriptive Statistics for the Student Data Set

Attribute	Format in survey
Overall evaluation	Question
Overall satisfaction	"How satisfied are you with this program in general?"
Attributes	Question
Partner companies	"How satisfied are you with the partner companies integrated into the program?"
Grades	"How satisfied are you with the grades which you have achieved in this program?"
Representation	"How satisfied are you with the work of the students' representation?"
Partner universities	"How satisfied are you with the current choice of foreign partner universities to study abroad?"
Administration	"How satisfied are you with the support provided by the administration?"
Contact to instructors	"How satisfied are you with the accessibility of instructors in this program?"
Organization	"How satisfied are you with the organization of this program?"
Quality of instruction	"How satisfied are you with the quality of instruction in this program?"
Fellow students	"How satisfied are you with your fellow students and their input into this program?"

	Descriptive statistics for student data[a]			
Attribute	*N*	Mode	Mean	SD
Global evaluation: Overall satisfaction	256	4	3.72	0.71
Partial evaluations				
Partner companies	256	4	3.82	1.05
Grades	256	4	3.77	0.78
Representation	256	3	3.02	1.10
Partner universities	256	4	3.96	0.83
Administration	256	4	3.92	0.87
Contact to instructors	256	3	3.30	1.09
Organization	256	4	3.52	0.96
Quality of instruction	256	3	3.02	1.08
Fellow students	256	4	3.72	0.95

[a]On a scale of 1 (not at all satisfied) to 5 (very satisfied).

response vectors have a range of 1 or less and 8% exhibit the maximum range on the nine-point scale. In the student data set, 16% responded with a range equal to 1 and 19% with the maximum range on the response scale.

In the Florida vacation data set, one brand belief dimension ("good for one-time visit only") is of opposite polarity. Thus, disagreement with this brand belief indicates that respondents think that the destination is worthwhile to visit again. The distribution

Table 3 Attributes and Descriptive Statistics for the Smartphone Data Set

Overall evaluation	Question
Likelihood of recommendation	"Based on your ownership experience, how likely is it that you would recommend your smartphone model to a friend or colleague?"

	Descriptive statistics for smartphone data[a]			
Attribute	*N*	Mode	Mean	SD
Global evaluation: Likelihood of recommendation	662	9	7.55	1.68
Partial evaluations				
Usability of basic features	662	9	7.93	1.37
Ease of navigation menu screens	662	9	7.93	1.28
Speed of operating system	662	9	7.30	1.64
Navigating through the Internet	662	8	7.37	1.56
View and send emails	662	9	7.71	1.43
Installing new applications	662	9	7.66	1.57
Visual appeal	662	9	7.96	1.40
Weight	662	9	7.74	1.46
Clarity of display	662	9	8.06	1.24
Display's resistance to scratches	662	9	7.18	1.81
Toggle or navigation wheel or buttons	662	9	7.36	1.60
Handling the touch-screen	662	9	7.56	1.61
Multimedia capabilities	662	9	7.56	1.51
Taking pictures with camera	662	9	7.62	1.55
Wi-Fi connectivity	662	9	7.58	1.58
Battery stamina	662	7	6.17	2.14

[a]On a scale of 1 (very bad) to 9 (very good).

Table 4 Attributes and Descriptive Statistics for the Florida Vacation Data Set

Overall evaluation	Question
Global satisfaction	"Of the theme parks that you have visited in the past five years, how satisfied are you with that theme park brand?"

	Descriptive statistics for Florida vacation data[a]			
Attribute	*N*	Mode	Mean	SD
Global evaluation: Global satisfaction	542	7	5.88	1.30
Partial evaluations				
Relaxing	542	5	4.66	1.67
Wholesome	542	7	5.68	1.34
Fun	542	7	5.79	1.33
Exciting	542	7	5.59	1.41
Premium	542	6	5.27	1.50
Memorable	542	7	5.70	1.42
Good for one-time visit only	542	1	3.36	2.00
Good enough to go back	542	7	5.44	1.65
Interesting	542	7	5.73	1.34
More fun than most	542	6	5.23	1.47
Offers wide variety	542	6	5.54	1.38
Enjoyable	542	7	5.76	1.33
Best value for price	542	5	4.88	1.46
Authentic	542	7	5.44	1.44
Safe	542	7	5.89	1.21
Great for whole family	542	7	5.86	1.32
Great for adults	542	7	5.60	1.46

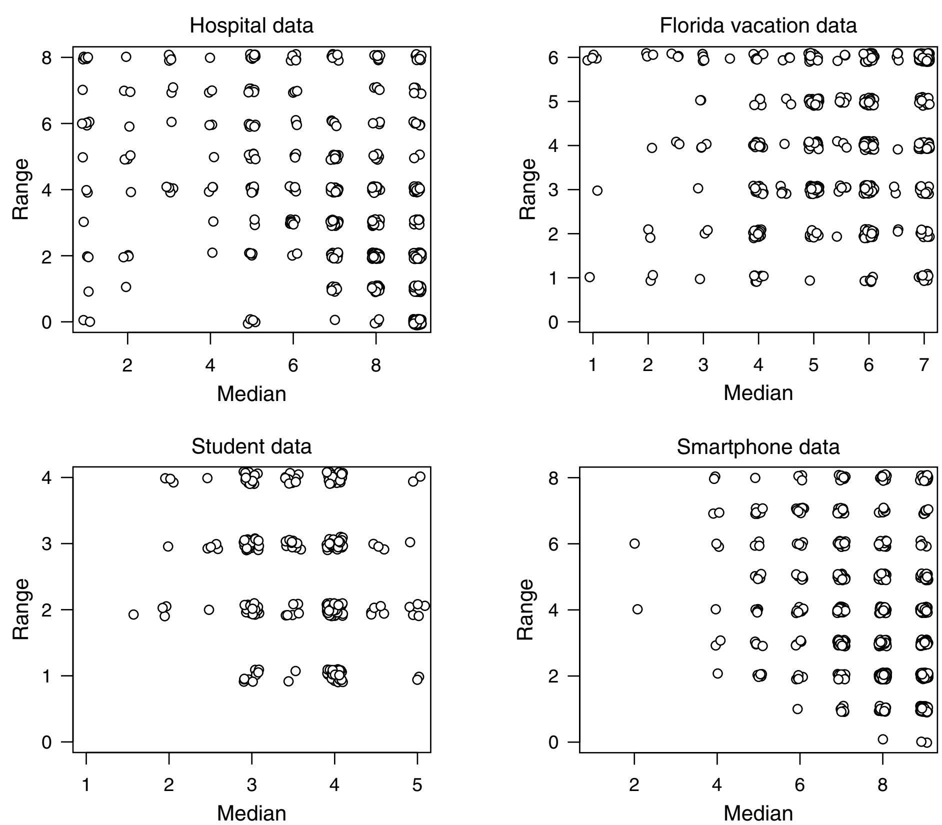
[a]On a scale of 1 (does not describe it at all) to 7 (describes it very well).

of responses to this variable has its mode at the lowest scale point. This leads to a much larger range of the raw data (see Figure 1).

4.2. Model Comparison

We compare the proposed mixture of dimensionalities to (1) an ordinal regression model in which all overall

Figure 1 Plot of Range vs. Median (Jittered Values)



satisfaction scores are assumed to be formed by aggregating partial scores and (2) a mixture of regression models. The simple ordinal regression model can be viewed as the standard approach to conducting driver analysis for customer satisfaction data. A more sophisticated approach to driver analysis is a mixture of regression models that allows for heterogeneity in driver effects across discrete mixture components (Wedel and DeSarbo 2002). We extend the standard mixture of regression model by a component that captures responses without systematic covariance (Equation (8)) for an even comparison to the mixture of dimensionalities model:

$$p(\mathbf{z}_i) = \sum_{f=1}^{F} \eta^f p^f(\mathbf{z}_i) + \eta^u p^u(\mathbf{z}_i). \tag{12}$$

Regression components (Equation (7)) are indexed by f, and u indicates the independence model from Equation (8). Equation (12) implies that a random observation is drawn with probability η^f from the regression mixture component f and probability η^u from the independence model. We note that the mixture of regression components corresponds to a mixture of covariance matrices with a priori full-dimensional covariance structures $\Sigma_z^1, \ldots, \Sigma_z^F$.

In the mixture of dimensionalities, we set the maximum dimensionality of a factor component to 5 and find throughout that only factor components of dimensionality 2 or smaller are supported by the data. In the mixture of regressions model, we set the maximum number of regression components F to 3 and find throughout that less than 3 regression components' are required to fit the data.

We estimate all models with random cut points as in Equation (10) and determine the number of components in the hierarchical mixture of Equation (11) from the data. Overall, there are three basic models (see Table 5 for an overview). We use the log-marginal density (LMD) of the data given the model as a measure of comparison and employ the approximation suggested by Newton and Raftery (1994) to compute the log-marginal density from the MCMC output. We run the MCMC sampler for 60,000 iterations and discard the first 40,000 iterations as burn-in. Convergence is assessed by graphically monitoring the convergence of key quantities such as component sizes and other component-specific parameters.

As additional measures of fit, we compute the mean absolute deviation (MAD) and root mean squared error (RMSE) of the predicted ratings, given estimates of the model parameters that form the hierarchical prior for individual response vectors. We predict latent evaluations by simulating unconstrained $\mathbf{z}_i$ from hierarchical prior distributions in Equations (9) and (12). Individual-level cut points and component allocation are simulated from their hierarchical priors. We use the simulated individual-level cutoff points $\mathbf{c}_i$ to transform continuous latent predictions in ordinal predictions. The ordinal predictions are then compared to the observed ratings. We compute the MAD and RMSE at each iteration and average these quantities over the MCMC chain. Thus, the computed MAD and RMSE do not condition on posterior component membership or on posterior information about individual-level cutoff points.

Results are summarized in Tables 6–8. Table 6 presents model fits. Tables 7 and 8 summarize the allocation of response vectors to mixture components in the mixture of regression model (Table 7) and the mixture of dimensionalities model (Table 8). Comparing the fit of the models, the mixture of dimensionalities fits all data sets best according to MAD and RMSE. The same holds for LMD for the hospital, student, and Florida vacation data sets. However, LMD suggests that the mixture of regressions fits the smartphone data best but that it is only barely better than the simple ordinal regression model.

In the mixture of regressions, three out of four data sets result in a single regression component, after accounting for independent responses (see Table 7). Given our weakly informative, subjective prior on component sizes, the posterior indicates that the data

Table 5 Model Overview

Model	Heterogeneity of response styles	Factor model specification	Component for independent evaluations	Cut-point specification
Ordinal regression	None	None	No	Heterogeneous cut points
Mixture of regression model	Multiple regression components to account for parameter heterogeneity	None	Yes	Heterogeneous cut points
Mixture of regression and factor model	Mixture of regression and low-dimensional response styles to account for structural heterogeneity	Multiple-factor models to account for dimensional heterogeneity	Yes	Heterogeneous cut points

Table 6 Model Fit

	Model 1: Ordinal regression model			Model 2: Mixture of ordinal regression			Model 3: Mixture of dimensionalities		
Data set	LMD	MAD	RMSE	LMD	MAD	RMSE	LMD	MAD	RMSE
Hospital	−5,644	1.648	2.176	−5,675	1.648	2.170	−5,300	1.648	2.161
Student	−2,797	0.710	0.951	−2,804	0.697	0.953	−2,676	0.695	0.949
Smartphone	−12,518	1.369	1.570	−12,517	1.361	1.570	−12,528	1.375	1.576
Florida vacation	−9,413	1.396	1.464	−9,439	1.392	1.468	−9,350	1.328	1.480

Notes. LMD measured as suggested by Newton and Raftery (1994). The MAD and RMSE are of predicted ordinal ratings based on observed ratings.

Table 7 Allocation to Components in the Mixture of Regression Model

	Share of regression components		Share of independent evaluations
Data set	Component 1	Component 2	
Hospital	0.829 (0.807, 0.862)	0.000 (0.000, 0.000)	0.171 (0.137, 0.193)
Student	0.686 (0.473, 0.865)	0.000 (0.000, 0.000)	0.314 (0.135, 0.527)
Smartphone	0.807 (0.758, 0.867)	0.000 (0.000, 0.000)	0.193 (0.133, 0.242)
Florida vacation	0.778 (0.745, 0.804)	0.061 (0.055, 0.068)	0.161 (0.135, 0.192)

Note. Reported are posterior means and in parentheses the 2.5% and 97.5% bounds of the posterior distribution.

Table 8 Allocation to Components in the Mixture of Dimensionalities Model

	Share of factor components			Share of regression	Share of independent evaluations
Data set	$h=1$	$h=2$	$h=3$		
Hospital	0.341 (0.294, 0.403)	0.12 (0.060, 0.172)	0.050[a] (0.036, 0.071)	0.395 (0.340, 0.459)	0.095 (0.054, 0.174)
Student	0.334 (0.202, 0.481)	0.007 (0.004, 0.012)	0.007[a] (0.006, 0.008)	0.364 (0.236, 0.508)	0.280 (0.182, 0.368)
Smartphone	0.006 (0.003, 0.009)	0.002 (0.001, 0.004)	0.008[a] (0.006, 0.009)	0.748 (0.698, 0.799)	0.236 (0.182, 0.289)
Florida vacation	0.765 (0.747, 0.784)	0.004 (0.003, 0.006)	0.006[a] (0.003, 0.01)	0.071 (0.066, 0.077)	0.154 (0.133, 0.174)

Note. Reported are posterior means and in parentheses the 2.5% and 97.5% bounds of the posterior distribution.
[a]Includes allocation to $h > 3$.

are not informative about additional regression components. A second regression component is only supported in the Florida vacation data set. However, the share of responses in this component is small (6%). The share of independent evaluations ranges from 16% (Florida vacation data) to 31% (student data).

In the mixture of dimensionalities, the shares of responses allocated to the factor models range from essentially 0% (smartphone data) to 77% (Florida vacation data) (see Table 8). No data set supports factor components of dimensionality larger than 2, and out of the three data sets where factor components receive substantial allocation (hospital, student, and Florida vacation), the student data and Florida vacation data only support one-dimensional factor components. Only the hospital data support a two-dimensional factor component in addition to a relatively much larger one-dimensional component. The share of independent evaluations decreases relative to the mixture of regression model, with the exception of the smartphone data. Finally, the share of the regression component ranges from 1% (Florida vacation data) to 75% (smartphone data).

4.3. Hospital Data

Table 9 summarizes the results from three models for the hospital data. We only report coefficients that are a posteriori informed by the data. That is, we do not report coefficients associated with empty components. The mixture of dimensionalities reveals that about 34% of responses exhibit a one-dimensional covariance structure, 12% of responses exhibit a two-dimensional structure, and 10% of responses are classified as random. Only about 40% of the responses are a priori consistent with a regression model based on the dimensionality of their covariance.

The factor loadings in the single-factor component of the mixture are all credibly positive but vary in size. This result is consistent with respondents who process individual items but color all responses by an overarching sentiment as in a factor model. Processing of items that are not identical causes variation of loadings across items. The influence of one and the same overarching sentiment when responding to individually different items causes the one-dimensional covariance structure. In the two-factor component of the mixture, the items “first choice,” "puts patients first," and "safe place" load credibly positively on the first factor. The items "first choice," "language," "skilled doctors," and "clean" load credibly positively on the second factor, and the item "conveniently located" loads credibly negatively. Thus, hospital evaluations exist in two a priori independent dimensions in this component of the mixture where the second dimension is more about "functional" characteristics of the hospital.

In the regression component of our mixture of dimensionalities that accounts for about 40% of responses, six items emerge as statistically credible drivers of overall hospital evaluations. "Skilled doctors" is most important a posteriori, followed by "safe place," "puts patients first," "conveniently located," "clean," and "language."

In comparing items across components of different dimensionalities, we see that "modern" and "concerned" are reflective of an overall hospital evaluation in the single-factor component but do not drive this overall evaluation. A qualitative difference between the two-factor and one-factor components is that "modern" is independent of both factors in the former; "concerned" misses the mark for statistical credibility in the two-factor component only barely. A qualitative difference between the two-factor component and the regression component is the pronounced judgmental trade-off between "conveniently located" and "skilled doctors" in the former. This trade-off may be caused by a strong perceived ecological correlation between these items, but it nevertheless impedes driver analysis.

Comparing inference from the regression component in the mixture of dimensionalities to that from established technologies (i.e., ordinal regression and mixtures of ordinal regressions), we see that both methodologies fail to recognize "language" and "clean" as credible drivers of hospital overall evaluations. Moreover, simple ordinal regression spuriously points to "modern" as a driver of overall evaluations. Inference regarding "language" stands out especially. Both established methodologies suggest that "language" has a borderline credibly negative impact on overall hospital evaluations. After controlling for "skilled doctors" and other variables, a negative impact of language seems hard to reconcile with any reasonable set of prior expectations about drivers.

Overall, we obtain qualitatively different and more face-valid inference about drivers of overall hospital evaluations from the proposed mixture of dimensionalities, and the comparison to the mixture of regression corroborates the empirical relevance of mixture of dimensionalities vis-à-vis heterogeneity in full-dimensional covariance structures.

Table 9 Parameter Estimates for the Hospital Data Set

	Ordinal regression			Mixture of regression			Mixture of dimensionalities											
				Regression component 1			Regression component			$h=1$			$h=2$ (factor 1)			$h=2$ (factor 2)		
Coefficients	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%
Global (factor)	—	—	—	—	—	—	—	—	—	**2.35**	**1.83**	**3.03**	**2.27**	**0.9**	**4.52**	**1.76**	**0.47**	**3.96**
Intercept (regression)	−0.24	−0.41	−0.06	−0.27	−0.43	−0.12	0.58	−0.23	1.37	—	—	—	—	—	—	—	—	—
Language	−0.12	−0.27	0.02	−0.11	−0.24	0.02	**0.29**	**0.02**	**0.58**	**0.95**	**0.67**	**1.32**	0.6	−0.18	1.81	1.52	0.5	3.2
Puts patients first	0.12	−0.14	0.37	0.28	0	0.55	**0.68**	**0.29**	**1.13**	**1.27**	**1.02**	**1.59**	**0.75**	**0.24**	**1.85**	0.21	−0.2	0.89
Safe place	0.42	0.19	0.67	**0.3**	**0.05**	**0.55**	**0.69**	**0.37**	**1.05**	**1.23**	**0.97**	**1.54**	**2.23**	**0.7**	**5.89**	0	0	0
Conveniently located	**0.19**	**0.04**	**0.33**	**0.19**	**0.07**	**0.32**	**0.65**	**0.31**	**0.93**	**1.12**	**0.75**	**1.56**	1.04	−0.13	3.4	**−1.5**	**−3.92**	**−0.22**
Skilled doctors	**0.59**	**0.37**	**0.81**	**0.62**	**0.41**	**0.83**	**0.93**	**0.53**	**1.27**	**1.22**	**0.95**	**1.52**	−0.28	−1.43	0.66	**1.66**	**0.79**	**3.81**
Modern	**0.23**	**0.05**	**0.42**	**0.15**	**−0.05**	**0.36**	0.14	−0.14	0.42	**1.5**	**1.15**	**1.92**	0.14	−0.54	1.18	0.7	−0.1	1.74
Clean	**−0.13**	**−0.37**	**0.11**	−0.11	−0.35	0.14	**0.65**	**0.16**	**0.97**	**1.36**	**1.06**	**1.72**	−0.24	−0.9	0.18	**0.41**	**0.01**	**1.01**
Concerned	0.14	−0.1	0.4	0.09	−0.16	0.34	0.27	−0.17	0.63	**1.33**	**1.04**	**1.69**	0.44	−0.06	1.3	0.42	−0.02	1.2

Note. All estimates credibly different from 0 are in bold.

4.4. Student Data

Table 10 summarizes the results for the student data. The mixture of dimensionalities reveals that about one-third of the responses exhibit a one-dimensional covariance structure; 28% of the responses are classified as random. Mixture components featuring factor models of dimensionality greater than 1 were not empirically relevant for this data set. About 36% of the responses are a priori consistent with a regression model based on the dimensionality of their covariance.

"Overall satisfaction" and satisfaction with "partner companies," the student "representation," and "fellow students" are credibly positively associated with a one-dimensional latent variable in the single-factor component of the mixture, and satisfaction with the "organization" of this program is borderline credibly positive. Satisfaction with the "administration" is credibly negatively associated with the same latent variable in this group, and satisfaction with "partner universities" is borderline credibly negative.

This pattern is consistent with respondents who process individual items; i.e., heterogeneous preferences for rating scale categories cannot explain the mix of positive and negative loadings. However, it is inconsistent with a simple factor process. The stronger the overarching latent sentiment toward the program, the higher the overall satisfaction, and satisfaction with partner companies, the student representation, and fellow students but the lower the satisfaction with the administration.

In comparing inference from the regression component in the mixture of dimensionalities to that from established technologies, we see that both ordinal regression and the mixture of regression uncover fewer and different credible drivers of overall student satisfaction. These methodologies fail to recognize the importance of satisfaction with partner universities and the student representation as drivers of overall student satisfaction. Moreover, all posterior means of regression coefficients estimated in the mixture of dimensionalities are clearly positive with substantial posterior mass in the positive domain. In contrast, the borderline credibly negative impact of satisfaction with the organization of the program inferred by the ordinal regression model as well as the mixture of ordinal regressions lacks face validity.

4.5. Florida Vacation Data

Table 11 summarizes the results for Florida vacation data. The mixture of dimensionalities reveals that about three-quarters of responses exhibit a one-dimensional covariance structure; 15% of responses are classified as random. Mixture components featuring factor models of dimensionality greater than 1 were not empirically relevant for this data set. Only about 7% of the responses are a priori consistent with a regression model based on the dimensionality of their covariance.

In the single-factor component of the mixture of dimensionalities, all loadings except that of "good for a one-time visit only" are credibly positive. That item loads credibly negative on the latent factor in this component of the mixture. The result is consistent with respondents who process individual items but color all responses by an overarching sentiment as in a factor model. A likely reason for the overwhelming share of one-dimensional responses in this case is that the designers of the survey, perhaps unintentionally, chose items that are different but not concrete enough to engage respondents beyond their overall evaluation.

The percentage of respondents in the regression component of the mixture of dimensionalities is too small to transform into a posteriori meaningful information about drivers of overall satisfaction with the

Table 10 Parameter Estimates for the Student Evaluation Data Set

	Ordinal regression			Mixture of regression			Mixture of dimensionalities					
				Regression component 1			Regression component			$h=1$		
Coefficients	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%
Global (factor)	—	—	—	—	—	—	—	—	—	**−0.34**	**−0.67**	**−0.01**
Intercept (regression)	−0.1	−0.32	0.1	0.06	−0.19	0.28	2.68	1.31	4.85	—	—	—
Partner companies	0.05	−0.05	0.15	**0.18**	**0**	**0.39**	0.24	−0.09	0.7	**−0.96**	**−1.74**	**−0.38**
Grades	**0.15**	**0.03**	**0.29**	0.12	−0.08	0.3	**0.35**	**0.03**	**0.64**	−0.16	−0.51	0.2
Representation	0.05	−0.04	0.14	0	−0.13	0.13	**0.31**	**0.01**	**0.61**	**−0.75**	**−1.43**	**−0.17**
Partner universities	0.08	−0.04	0.21	0.12	−0.08	0.33	**0.49**	**0.16**	**0.89**	0.44	−0.05	0.97
Administration	−0.04	−0.19	0.11	0.01	−0.17	0.22	0.25	−0.15	0.57	**0.76**	**0.06**	**1.41**
Contact to instructors	−0.01	−0.1	0.08	0.07	−0.04	0.19	0.16	−0.1	0.45	−0.07	−0.5	0.38
Organization	−0.09	−0.21	0.02	−0.1	−0.24	0.05	0.23	−0.08	0.53	−0.3	−0.78	0.06
Quality of instruction	0.1	−0.02	0.24	**0.27**	**0.11**	**0.49**	**0.47**	**0.19**	**0.8**	−0.31	−1.17	0.35
Fellow students	**0.19**	**0.09**	**0.28**	**0.18**	**0.05**	**0.32**	**0.44**	**0.11**	**0.73**	**−0.51**	**−1.1**	**−0.03**

Note. All estimates credibly different from 0 are in bold.

Table 11 Parameter Estimates for the Florida Vacation Data Set

	Ordinal regression			Mixture of regression						Mixture of dimensionalities					
				Regression component 1			Regression component 2			Regression component			$h=1$		
Coefficients	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%
Global (factor)	—	—	—	—	—	—	—	—	—	—	—	—	**0.33**	**0.22**	**0.46**
Intercept (regression)	0.98	0.66	1.31	0.99	0.62	1.34	2.45	−6.96	12.32	−1.79	−14.35	10.24	—	—	—
Relaxing	0.14	−0.01	0.31	0.12	−0.05	0.29	−0.29	−4.58	3.74	1.73	−7.02	10.29	**0.24**	**0.16**	**0.33**
Wholesome	−0.07	−0.33	0.22	−0.16	−0.46	0.12	0.05	−4.38	4.93	−0.5	−8.37	7.47	**0.32**	**0.22**	**0.42**
Fun	0.21	−0.17	0.58	0.27	−0.19	0.78	0.77	−3.26	5.38	−1.21	−12.01	8.3	**0.38**	**0.27**	**0.5**
Exciting	−0.08	−0.42	0.26	−0.1	−0.41	0.22	0.6	−3.71	6.38	3.79	−7	15.29	**0.35**	**0.25**	**0.48**
Premium	0.06	−0.24	0.32	0.24	−0.02	0.52	−0.82	−5.05	3.17	−3.56	−12.01	6.54	**0.35**	**0.24**	**0.47**
Memorable	−0.06	−0.36	0.23	−0.22	−0.58	0.15	−1.54	−5.72	2.19	−0.46	−9.6	9.38	**0.41**	**0.29**	**0.56**
Good for one-time visit only	−0.37	−0.53	−0.21	−0.29	−0.47	−0.14	0.51	−3.53	5.79	−4.52	−13.77	3.82	**−0.65**	**−0.88**	**−0.45**
Good enough to go back	0	−0.28	0.26	0.12	−0.17	0.38	−0.06	−5.3	4.4	0.56	−11.19	11.88	**0.62**	**0.44**	**0.83**
Interesting	−0.03	−0.33	0.29	−0.1	−0.49	0.35	0.96	−3.3	5.38	2.57	−5.26	14.27	**0.35**	**0.24**	**0.48**
More fun than most	−0.01	−0.26	0.26	−0.14	−0.44	0.14	−0.64	−5.03	4.1	−0.97	−11.29	7.89	**0.37**	**0.26**	**0.5**
Offers wide variety	−0.14	−0.44	0.17	−0.16	−0.44	0.11	0.39	−4.18	4.42	0.15	−8.82	7.66	**0.37**	**0.26**	**0.5**
Enjoyable	0.03	−0.39	0.43	0.45	−0.03	0.97	0.54	−3.72	6.33	3.4	−5.88	13.25	**0.39**	**0.27**	**0.52**
Best value for price	0.12	−0.11	0.36	0.19	−0.07	0.48	−0.19	−4.86	4.51	−2.11	−15.18	10.6	**0.31**	**0.22**	**0.41**
Authentic	0.18	−0.07	0.41	0.22	−0.05	0.46	0.31	−3.55	4.3	−2.73	−12.94	5.96	**0.33**	**0.24**	**0.45**
Safe	−0.13	−0.35	0.12	−0.12	−0.32	0.11	−0.29	−4.66	5	0.41	−8.55	9.03	**0.25**	**0.17**	**0.34**
Great for whole family	−0.07	−0.31	0.16	−0.24	−0.51	0.04	0.18	−4.36	4.64	2.49	−7.12	10.56	**0.43**	**0.3**	**0.59**
Great for adults	0.11	−0.11	0.33	0.05	−0.19	0.28	0.26	−4.27	5.11	0.43	−8.61	10.02	**0.42**	**0.28**	**0.57**

Note. All estimates credibly different from 0 are in bold.

theme park. Both ordinal regression and the mixture of ordinal regressions point to "good for a one-time visit only" as the only credible driver of overall satisfaction. The result is consistent with our simulated application of a regression model to data in which a unidimensional covariance structure is dominant (see the Web appendix). Managerially speaking, the results from the mixture of dimensionalities imply that the recommendation to improve ratings along the item "good for a one-time visit only" is not any more operational, concrete, or actionable than the recommendation to do "something" to improve overall satisfaction.

4.6. Smartphone Data

The smartphone data set is the only one among the four investigated in this paper where the mixture of dimensionalities fits the data clearly worse than either of the two established methodologies (see Table 6). This is consistent with the minimal posterior weight of factor components in the mixture of dimensionalities (see Table 12). Moreover, the mixture of regressions barely improves on the simple ordinal regression model in terms of fit and the data do not support more than one regression component. The joint distribution of responses is therefore consistent with a single multivariate normal distribution as a data-generating mechanism, after accounting for heterogeneous scale use.

A posteriori, the smartphone survey therefore is composed of concrete, well-differentiated items in the eyes of respondents who are able and motivated to process these items and give specific responses. The ordinal regression identifies 9 statistically credible drivers of global smartphone evaluations among 16 items included in the survey. The relatively most important performance dimensions are "battery stamina," "clarity of display," and "handling the touch-screen." Additional statistically credible drivers are the "ease of navigation menu screens," "view and send emails," "installing new applications," the "display's resistance to scratches," "taking pictures with camera," and "Wi-Fi connectivity." Furthermore, the phone's "visual appeal," the "speed of operating system" and (the satisfaction with) the phone's "weight" have substantial posterior mass in the positive domain, and no coefficients are estimated credibly negative.

4.7. Scale Use

We summarize marginal posterior means and variances of respondent-specific scale use parameters in

Table 12 Parameter Estimates for the Smartphone Data Set

	Ordinal regression			Mixture of regression			Mixture of dimensionalities		
				Regression component 1			Regression component		
Coefficients	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%	Post mean	2.5%	97.5%
Intercept (regression)	0.74	0.26	1.27	1.17	0.29	2.51	3.36	1.25	5.38
Usability of basic features	0	−0.15	0.16	−0.01	−0.28	0.2	0.05	−0.25	0.34
Ease of navigation menu screens	**0.19**	**0.04**	**0.35**	**0.36**	**0.11**	**0.7**	**0.42**	**0.05**	**0.87**
Speed of operating system	0.09	−0.04	0.21	0.16	−0.06	0.37	**0.48**	**0.13**	**1.02**
Navigating through the Internet	−0.05	−0.21	0.12	0.1	−0.17	0.44	0.17	−0.19	0.62
View and send emails	**0.19**	**0.03**	**0.36**	**0.21**	**0**	**0.42**	**0.39**	**0.04**	**0.72**
Installing new applications	**0.2**	**0.06**	**0.34**	0.1	−0.1	0.31	**0.3**	**0.11**	**0.53**
Visual appeal	0.12	−0.02	0.26	0.23	−0.04	0.52	0.1	−0.23	0.4
Weight	0.06	−0.08	0.2	0.03	−0.18	0.25	**0.44**	**0.09**	**0.89**
Clarity of display	**0.33**	**0.15**	**0.52**	**0.39**	**0.11**	**0.67**	**0.52**	**0.19**	**0.89**
Display's resistance to scratches	**0.19**	**0.04**	**0.35**	**0.2**	**0.02**	**0.42**	**0.39**	**0.14**	**0.75**
Toggle or navigation wheel or buttons	0.04	−0.12	0.19	0.02	−0.22	0.34	0.31	−0.04	0.68
Handling the touch-screen	**0.32**	**0.15**	**0.47**	**0.31**	**0.1**	**0.53**	**0.36**	**0**	**0.73**
Multimedia capabilities	0	−0.16	0.18	0.06	−0.17	0.32	0.02	−0.36	0.33
Taking pictures with camera	**0.18**	**0.05**	**0.31**	0.09	−0.07	0.27	0.01	−0.18	0.19
Wi-Fi connectivity	**0.2**	**0.03**	**0.35**	0.14	−0.01	0.32	0.21	−0.02	0.53
Battery stamina	**0.39**	**0.23**	**0.58**	**0.5**	**0.23**	**0.92**	**1.13**	**0.6**	**1.61**

Note. All estimates credibly different from 0 are in bold.

Table 13.[17] We also report first differences of posterior cut-point means that measure the length of latent continuous intervals that correspond to observed ratings of 2, 3, 4, etc. Across all four data sets, we see a tendency for the interval width of ratings to increase for better, i.e., higher ratings. Another commonality across all four data sets is that the cut points separating top box ratings from the next lower ratings have the highest posterior variances among all cut points. Together, these results suggest that better ratings are somewhat more ambiguous in terms of across-respondent comparisons than ratings at the lower end of the satisfaction scales.

5. Discussion

We developed a mixture model that uses the dimensionality of the covariance among items in a survey as the primary classification criterion while accounting for heterogeneity in rating scale usage. This way, our model distinguishes between responses that are a priori inconsistent with regression-based driver analysis because their covariances can be projected onto a low-dimensional space and responses where regression analysis makes sense a priori as per their high-dimensional covariance structure. Among the possible reasons for failing consistency with a regression model are an (unintended) lack of specificity of individual items on the survey, halo effects, or a lack of motivation to process individual items.

When survey items are not specific enough to engage respondents beyond an overall sentiment toward the object of evaluation, as in the Florida vacation data, our model produces positive evidence in favor of a low-dimensional factor model. In turn, this evidence alerts the researcher to the inadequacy of regression-based driver analysis based on these data. When items are, in principle, specific enough to engage respondents beyond some low-dimensional overarching sentiment structure, as intended by all commercial customer satisfaction research, our model (probabilistically) separates respondents who nevertheless respond based on low-dimensional overarching sentiments from respondents who provide specific, unique responses to individual items. We encountered this situation in the hospital and student data sets.

Finally, our model controls for respondents who seem to provide responses based on idiosyncratic preferences for categories of the rating scale only, i.e., responses that collectively fail to consistently discriminate among individual items. The covariance among observed responses of this type is driven by heterogeneous scale use only. We find evidence for the empirical relevance of such responses in three out of four data sets.

Our model differs from prior research in marketing and psychology by making a heterogeneous mixture

[17] All four data sets empirically support a discrete mixture of continuous priors for the scale random effects (see Equation (11)). In the case of the student and the smartphone data, two components fit best. In both data sets, a relatively much smaller component captures outliers. The vacation data again are fit best with two components; however, both are of substantial size (about 60% and about 40%, respectively). Finally, the hospital data support three components with component sizes equal to about 25%, 25%, and 50%, respectively.

Table 13 Posterior Means and Variances of Cut Points

	Hospital data		Student data		FL vacation data		Smartphone data	
	Post mean	First diff.	Post mean	First diff.	Post mean	First diff.	Post mean	First diff.
Mean of cut points								
Cut1	−3.582	—	−3.416	—	−1.184	—	−3.784	—
Cut2	−3.212	0.370	−2.410	1.006	−0.888	0.295	−3.492	0.292
Cut3	−2.877	0.336	−1.129	1.281	−0.530	0.358	−3.123	0.369
Cut4	−2.581	0.296	0.836	1.967	−0.174	0.355	−2.720	0.403
Cut5	−1.922	0.659	—		0.175	0.350	−2.252	0.468
Cut6	−1.548	0.374	—		0.825	0.650	−1.652	0.599
Cut7	−0.844	0.702	—		—		−0.800	0.852
Cut8	0.235	1.080	—		—		0.325	1.125
Variance of cut points								
Cut1	1.730		0.226		0.192		0.506	
Cut2	1.732		0.367		0.239		0.595	
Cut3	1.767		0.413		0.156		0.672	
Cut4	1.958		0.832		0.158		0.764	
Cut5	1.952		—		0.192		1.034	
Cut6	1.901		—		1.298		1.072	
Cut7	1.898		—		—		1.216	
Cut8	3.053		—		—		2.244	

of dimensionalities operational. Traditionally, marketing is relatively more interested in driver analysis, whereas psychometric tools tend to emphasize theory-driven dimensionality reduction. The differential emphasis is a natural consequence of the typical decision problems in the two fields. Marketing often needs to understand how to influence the experience by potentially heterogeneous but nevertheless larger groups of consumers. Psychometric tools emphasize precise estimation of latent variables at the individual level for testing purposes.[18] However, research in both disciplines assumes that items on a questionnaire or test are measurement invariant across individuals in a survey. Psychological testing and the construction of psychological tests obviously requires (strong) measurement invariance in the relevant population. Driver analysis in marketing implicitly assumes that respondents unequivocally interpret the meaning of individual items on the survey. In contrast, our model accommodates a particularly striking failure of measurement invariance: the possibility that responses to one and the same survey by different respondents are of inherently different dimensionality.

We see two immediate opportunities for future research based on this paper. First, our data are the result of mental processes occurring when one responds to the survey but do not contain indicators directly linked to information processing such as response time (e.g., Otter et al. 2008) or eye-tracking traces (e.g., van der Lans et al. 2008). We believe that such process measures will be helpful in further refining models of survey response. Second, based on the four data sets investigated here, we can only offer limited operational guidance of how to best design and administer surveys for driver analysis. The specific items on the smartphone survey provide a positive example that is in obvious contrast to the collection of items on the Florida vacation survey that render driver analysis meaningless. However, the mixture of dimensionalities model will help assess the effect of manipulations aimed at making surveys for driver analysis more informative.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/mksc.2013.0779.

Acknowledgments

The authors thank The Modellers and especially Jeff Brazell and Kevin van Horne for providing the hospital, the vacation, and the smartphone data sets used in this paper and for pointing out an error in an earlier version of their MCMC code.

Appendix: MCMC Sampling, Prior Specification, and Posterior Identification

Our MCMC recursively samples from the following conditional distributions:

Step 1. $S_i \mid \mathbf{z}_i, (\boldsymbol{\mu}^{h=1}, \boldsymbol{\Sigma}^{h=1}), (\boldsymbol{\mu}^{h=2}, \boldsymbol{\Sigma}^{h=2}), \ldots, (\boldsymbol{\mu}^{h=H}, \boldsymbol{\Sigma}^{h=H}), (\boldsymbol{\mu}^{f}, \boldsymbol{\Sigma}^{f}), (\boldsymbol{\mu}^{u}, \boldsymbol{\Sigma}^{u}), \boldsymbol{\eta}$.

Step 2. $\boldsymbol{\eta} \mid \{S_i\}$.

Step 3. $(\boldsymbol{\mu}^{h=1}, \boldsymbol{\Sigma}^{h=1}), (\boldsymbol{\mu}^{h=2}, \boldsymbol{\Sigma}^{h=2}), \ldots, (\boldsymbol{\mu}^{h=H}, \boldsymbol{\Sigma}^{h=H}), (\boldsymbol{\mu}^{f}, \boldsymbol{\Sigma}^{f}), (\boldsymbol{\mu}^{u}, \boldsymbol{\Sigma}^{u}) \mid \{S_i, \mathbf{z}_i\}$.

Step 4. $\mathbf{z}_i \mid S_i, \boldsymbol{\mu}^{S_i}, \boldsymbol{\Sigma}^{S_i}, \{c_{i,k}\}$.

Step 5. $\{c_{i,k}\} \mid \mathbf{z}_i, ind_i, \boldsymbol{\mu}_{\Delta}^{ind_i}, \boldsymbol{\Sigma}_{\Delta}^{ind_i}$.

[18] Obviously, precise individual-level estimates of latent variables are helpful when the primary goal is to understand causal relationships. However, in practical market research, the trade-off between controlling for measurement error and obtaining information about additional dimensions empirically favors the latter. This practice makes sense in the context of specific items tapping into concrete experiences with the product or service.

Step 6. $\boldsymbol{\mu}_\Delta^{ind_i}, \boldsymbol{\Sigma}_\Delta^{ind_i} \mid \{ind_i, \{c_{i,k}\}\}$.
Step 7. $ind_i \mid \{c_{i,k}\}, \{\boldsymbol{\mu}_\Delta^{ind_i}, \boldsymbol{\Sigma}_\Delta^{ind_i}\}, \boldsymbol{\eta}_{ind}$.
Step 8. $\boldsymbol{\eta}_{ind} \mid \{ind_i\}$.

Step 1 is the standard augmentation of unobserved group membership in mixture models, denoted S_i. Step 2 is a standard conjugate update of the hierarchical prior for the component membership indicators S. Step 3 updates component-specific parameters that are independent across mixture components conditional on $\{S_i\}$. When we update parameters of a factor model, we *first* augment latent factor scores that are integrated out in Step 1 and then update loadings and error variances. Step 4 cycles through the full conditional distributions of individual elements of $\mathbf{z}_i$'s generating from truncated normal distributions. Step 5 employs a random walk Metropolis step where the random walk is on the underlying log-normally distributed increments. Steps 6–8 are again standard updates in a mixture model. All individual steps are well documented in the literature.

We employ the following weakly informative subjective priors in our analysis:

$$\mathbf{c}^h \sim \mathrm{N}(\mathbf{0}, 10) \quad \forall h = 1, \ldots, H,$$
$$\boldsymbol{\Gamma}^h \sim \mathrm{N}(\mathbf{0}, 10 \cdot \mathbf{I}_{(q+1)h}) \quad \forall h = 1, \ldots, H,$$
$$\xi_h \sim \mathrm{N}(\mathbf{0}, \mathbf{I}_h) \quad \forall h = 1, \ldots, H,$$
$$\boldsymbol{\beta} \sim \mathrm{N}(\mathbf{0}, 10 \cdot \mathbf{I}_q),$$
$$c_y^f \sim \mathrm{N}(0, 10),$$
$$\boldsymbol{\Sigma}_{z_x}^f \sim \mathrm{IW}(q+3, (q+3) \cdot \mathbf{I}_q),$$
$$\boldsymbol{\mu}_{z_x}^f \sim \mathrm{N}(\mathbf{0}, 10 \cdot \mathbf{I}_q),$$
$$\sigma_{\varepsilon_y}^2, \psi_{m,m}^u, \psi_{m,m}^h \sim \mathrm{IG}(3,3) \quad \forall m = 1, \ldots, q+1, \forall h = 1, \ldots, H,$$
$$\mathbf{c}_z^u \sim \mathrm{N}(\mathbf{0}, 10 \cdot \mathbf{I}_{q+1}),$$
$$\boldsymbol{\Sigma}_z^u \sim \mathrm{IW}(q+3, (q+3) \cdot \mathbf{I}_{q+1}),$$
$$\boldsymbol{\eta}_S \sim \mathrm{Dir}(3, \ldots, 3).$$

Subjective priors for the YFW scale use model are

$$\boldsymbol{\mu}_\Delta \sim \mathrm{N}(\mathbf{0}, 10),$$
$$\boldsymbol{\Sigma}_\Delta \sim \mathrm{IW}(k+2, (k+2) \cdot \mathbf{I}_{k-1}),$$
$$\boldsymbol{\eta}_{ind} \sim \mathrm{Dir}(3, \ldots, 3).$$

The model we estimate is not likelihood identified. We "margin down" to the space of likelihood-identified parameters by postprocessing the MCMC output. We first identify the trade-off between cut points and latent variables by standardizing one latent variable. We then apply orthonormal rotations to identify the factor loadings uniquely.

Joint Identification of Cut-Points and Latent Variables. Neither the YFW model nor the extension proposed herein is identified. More specifically, the mean and the variance of the latent evaluations $\mathbf{z}_i$ and the cut points $\mathbf{c}_i$ are not jointly identified. One can shift all cut points $\mathbf{c}_i$ by a constant and shift the latent evaluations $\mathbf{z}_i$ by the same constant and obtain the same ordinal observations. The same holds for rescaling cut points and latent evaluations by an arbitrary factor. One way to solve the identification problem would be to fix two cut points. This approach is not desirable because it fixes the scale of latent evaluations across individuals. Instead, we achieve identification by "postprocessing" the MCMC output. Consider the regressions component of the mixture of dimensionalities model:

$$z_{y,i} = c_y^f + \boldsymbol{\beta}^T \mathbf{z}_{x,i} + \varepsilon_i.$$

Let

$$\tilde{\mathbf{z}}_x = \mathbf{z}_x - \mu_{z_{x,1}} \quad \text{and} \quad \tilde{\mathbf{c}} = \mathbf{c} - \mu_{z_{x,1}},$$

where $\mu_{z_{x,1}}$ is the mean of $\mathbf{z}_{x,1}$. This implies

$$\begin{aligned}\tilde{z}_{y,i} &= z_{y,i} - \mu_{z_{x,1}} = c_y^f - \mu_{z_{x,1}} + \boldsymbol{\beta}'(\tilde{\mathbf{z}}_{x,i} + \mu_{z_{x,1}}) + \varepsilon_i \\ &= c_y^f + \mu_{z_{x,1}}(\boldsymbol{\beta}'\iota - 1) + \boldsymbol{\beta}'\tilde{\mathbf{z}}_{x,i} + \varepsilon_i,\end{aligned}$$

where ι is a vector of ones. Furthermore, let

$$\tilde{\mathbf{z}}_x^* = \tilde{\mathbf{z}}_x / \sigma_{z_{x,1}} \quad \text{and} \quad \tilde{\mathbf{c}}^* = \tilde{\mathbf{c}} / \sigma_{z_{x,1}},$$

where $\sigma_{z_{x,1}}^2$ is the variance of $\mathbf{z}_{x,1}$.

It follows that

$$\begin{aligned}\tilde{\mathbf{z}}_y^* &= \tilde{\mathbf{z}}_y / \sigma_{z_{x,1}} = \sigma_{z_{x,1}}^{-1}(c_y^f + \mu_{z_{x,1}}(\boldsymbol{\beta}'\iota - 1) + \boldsymbol{\beta}'\tilde{\mathbf{z}}_{x,i} + \varepsilon_i) \\ &= \sigma_{z_{x,1}}^{-1}(c_y^f + \mu_{z_{x,1}}(\boldsymbol{\beta}'\iota - 1)) + \boldsymbol{\beta}'\tilde{\mathbf{z}}_{x,i}^* + \sigma_{z_{x,1}}^{-1}\varepsilon_i.\end{aligned}$$

This implies that we can use the estimated mean and variance of $\mathbf{z}_{x,1}$ to perform the following postprocessing steps:
Step 1. $\tilde{\mathbf{c}}_i^* = \sigma_{z_{x,1}}^{-1}(\mathbf{c}_i - \mu_{z_{x,1}})$.
Step 2. $c_y^{f*} = \sigma_{z_{x,1}}^{-1}(c_y^f + \mu_{z_{x,1}}(\boldsymbol{\beta}^T\iota - 1))$.
Step 3. $\sigma_{\tilde{z}_y^*}^2 = \sigma_{z_y}^2 / \sigma_{z_{x,1}}^2$.
Step 1 adjusts and rescales the cut points, which are then transformed to the log-delta scale to recover the hyperparameters. Step 2 adjusts and rescales the baseline of the regression model. Step 3 rescales the error variance given the variance of $\mathbf{z}_{x1}$. Postprocessing the output from the regression component implies the following postprocessing steps for the factor models:

$$\begin{pmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_{q+1} \end{pmatrix} = \begin{pmatrix} c_1^h \\ \vdots \\ c_{q+1}^h \end{pmatrix} + \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{q+1} \end{pmatrix} \boldsymbol{\xi} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \vdots \\ \boldsymbol{\varepsilon}_{q+1} \end{pmatrix}.$$

Normalization and standardization,

$$\begin{pmatrix} \tilde{\mathbf{z}}_1^* \\ \vdots \\ \tilde{\mathbf{z}}_{q+1}^* \end{pmatrix} = \sigma_{z_{x,1}}^{-1} \begin{pmatrix} c_1^h - \boldsymbol{\mu}_{z_{x,1}} \\ \vdots \\ c_{q+1}^h - \boldsymbol{\mu}_{z_{x,1}} \end{pmatrix} + \sigma_{z_{x,1}}^{-1} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_{q+1} \end{pmatrix} \boldsymbol{\xi} + \sigma_{z_{x,1}}^{-1} \begin{pmatrix} \boldsymbol{\varepsilon}_1^h \\ \vdots \\ \boldsymbol{\varepsilon}_{q+1}^h \end{pmatrix},$$

yields the following identified quantities:
1. $\boldsymbol{\gamma}^* = \sigma_{z_{x,1}}^{-1}(\boldsymbol{\gamma})$.
2. $\mathbf{c}^{h*} = \sigma_{z_{x,1}}^{-1}(\mathbf{c}^h - \mu_{z_{x,1}})$.
3. $\boldsymbol{\Psi}^* = \sigma_{z_{x,1}}^{-1}\boldsymbol{\Psi}$.

McCulloch et al. (2000) provide a detailed discussion of the effective subjective priors on identified quantities that result from normalizing a variance to 1 by postprocessing. In addition, our transformations involve (i) linear combinations of a priori normally distributed variables, (ii) products of a priori normally distributed variables, and (iii) shifting and rescaling of sums of a priori log-normally distributed quantities. We note that identification of all models compared in our paper relies on the same postprocessing strategy. Therefore, implied subjective priors on likelihood-identified parameters are identical across models.

Priors implied by (i) can be easily derived in closed form. Priors implied by (ii) and (iii) are not available in closed form but are amenable to investigation by simulation. A detailed characterization of the implied priors is beyond the scope of this paper. However, we note that at the cost of forgoing conjugacy when updating $\mathbf{\Sigma}^f_{z_x}$, an identified sampler results from fixing the mean and variance of an active variable in the regression component. Alternatively, one could fix the mean and variance of a variable in the independence component at the same cost or fix the error variance and constant of a variable in one of the factor components.

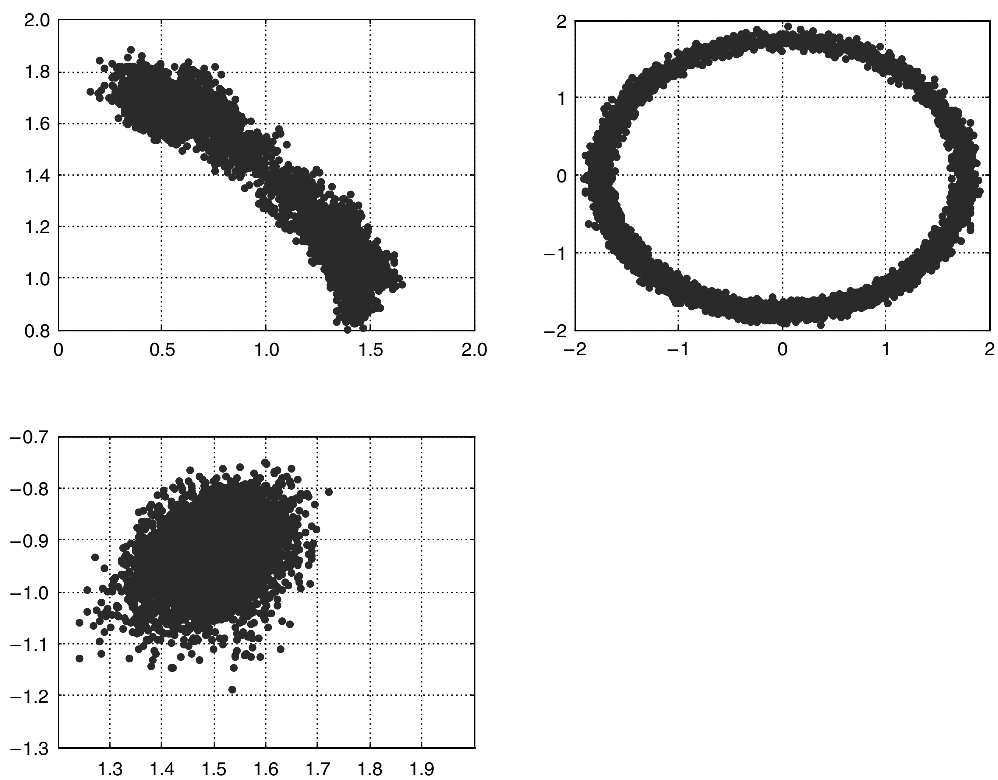
Identification of Factor Loadings. In the model (or mixture component) with one factor, the assumption of a standard normal distribution for the factor scores identifies all factor loadings up to their signs. However, in practice, sign flips are unlikely when at least one loading is credibly different from zero. In this case the differently signed, symmetric posterior modes are well separated. Careful inspection of MCMC trace plots of loadings empirically verified that sign flips did not occur in any of our analyses.

In models (or mixture components) with more than one factor, MCMC estimation without prior constraints on loadings is free to explore an infinite number of observationally equivalent orthonormally rotated loading matrices at each iteration of the sampler. The usual a priori constraint used to obtain one unique rotation that is meaningfully comparable across iterations is to set all factor loadings above the main diagonal equal to 0. A drawback of this a priori approach is that identification hinges on the assumption that all elements on the main diagonal of the loadings are different from zero a posteriori. Consider a two-factor model for six observed items, for example. MCMC estimation under the constrained loadings matrix,

$$\begin{pmatrix} \gamma_{11} & 0 \\ \gamma_{21} & \gamma_{22} \\ \gamma_{31} & \gamma_{32} \\ \gamma_{41} & \gamma_{42} \\ \gamma_{51} & \gamma_{52} \\ \gamma_{61} & \gamma_{62} \end{pmatrix},$$

only uniquely and comparably identifies factor loadings across MCMC iterations if indeed $\gamma_{11} \neq 0$ a posteriori. If the (arbitrarily chosen) first item is unique relative to the other items such that $\gamma_{11} = 0$ in the data-generating mechanism, or if the data are not informative enough to reliably detect $\gamma_{11} \neq 0$, identification based on the first item is impossible or at least tenuous. Figure A.1 shows scatterplots of loading draws for one item in a two-factor model. The x axis plots loadings on the first factor and the y axis loadings on the second factor. The top right plot clarifies two points: The first point is that identification based on a unique item is impossible. In some sense an attempt to identify a particular (orthonormal) rotation based on a unique item has the effect

Figure A.1 Scatterplots of Loading Draws



Note. Top left: unidentified sampler; top right: "identification" based on a unique item; bottom left: identification based on a common item.

to maximize mixing over different rotations. The second point is that a badly chosen identification constraint will be immediately recognized from the posterior.

The posterior from the unidentified sampler in the top panel left panel of Figure A.1 only visits a subset of the possible rotations. Finally, the bottom left plot shows the joint posterior distribution of loadings when identification is based on a (reliably) nonunique item.

One may argue that a bad *prior* identification constraint will always be recognized and can be switched a posteriori by orthonormal rotation anyway, without the need to rerun the MCMC. Despite this argument, we prefer identification a posteriori because the influence of subjective priors on loading coefficients may depend on the choice of the (arbitrary) prior identification constraint.

A posteriori identification proceeds by orthonormally rotating the loadings matrix such that elements above the main diagonal are zero. Such rotations involve simple matrix algebra, even when the number of factors is large. Note that the analyst is free to choose any item's loadings to form the first, second, etc., rows of the loadings matrix a posteriori.

References

Beckwith NE, Lehmann DR (1975) The importance of halo effects in multi-attribute attitude models. *J. Marketing Res.* 12(3):265–275.

Bollen KA (1989) *Structural Equations with Latent Variables* (John Wiley & Sons, New York).

Bradlow ET, Zaslavsky AM (1999) A hierarchical latent-variable model for ordinal data from a customer satisfaction survey with "no answer" responses. *J. Amer. Statist. Assoc.* 94(445): 43–52.

Bradlow ET, Wainer H, Wang X (1999) A Bayesian random effects model for testlets. *Psychometrika* 64(2):153–164.

Cunningham WH, Cunningham ICM, Green RT (1977) The ipsative process to reduce response set bias. *Public Opinion Quart.* 41(3):379–384.

Finn A (2011) Investigating the non-linear effects of e-service quality dimensions on customer satisfaction. *J. Retailing Consumer Services* 18(1):27–37.

Fornell C, Johnson MD, Anderson EW, Cha J, Bryant BE (1996) The American customer satisfaction index: Nature, purpose, and findings. *J. Marketing* 60(4):7–18.

Frühwirth-Schnatter S (2006) *Finite Mixture and Markov-Switching Models* (Springer, New York).

Gal D, Rucker DD (2011) Answering the unasked question: Response substitution in consumer surveys. *J. Marketing Res.* 48(1):185–195.

Gibbons RD, Hedeker DR (1992) Full-information item bi-factor analysis. *Psychometrika* 57(3):423–436.

Green PJ (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika* 82(4):711–732.

Groves RM, Fowler FJ, Couper MP, Lepkowski JM, Singer E, Tourangeau R (2004) *Survey Methodology* (John Wiley & Sons, Hoboken, NJ).

Hayes BE (2008) *Measuring Customer Satisfaction and Loyalty: Survey Design, Use, and Statistical Analysis Methods* (ASQ Quality Press, Milwaukee).

Iacobucci D (1994) Classic factor analysis. Bagozzi RP, ed. *Principles of Marketing Research* (Blackwell, Cambridge, MA), 279–316.

J.D. Power and Associates (2012) 2012 U.S. wireless smartphone and traditional mobile phone customer satisfaction study—Vol. 2 results. Press release (September 6), J.D. Power and Associates, Westlake Village, CA. http://www.jdpower.com/content/study/s1o19it/2012-u-s-wireless-smartphone-and-traditional-mobile-phone-customer-satisfaction-study-vol-2-results.htm.

Johnson VE, Albert JH (1999) *Ordinal Data Modeling* (Springer, New York).

Jones CS (2006) A nonlinear factor analysis of S&P 500 Index option returns. *J. Finance* 61(5):2325–2363.

Jöreskog KG (1967) Some contributions to maximum likelihood factor analysis. *Psychometrika* 32(4):443–482.

Kass RE, Raftery AE (1995) Bayes factors. *J. Amer. Statist. Assoc.* 90(430):779–795.

Krantz DH, Luce RD, Suppes P, Tversky A (1971) *Foundations of Measurement* (Academic Press, New York).

Ledermann W (1937) On the rank of the reduced correlational matrix in multiple-factor analysis. *Psychometrika* 2(2): 85–93.

Lubke GH, Muthén B (2005) Investigating population heterogeneity with factor mixture models. *Psych. Methods* 10(1):21–39.

McCulloch RE, Polson NG, Rossi PE (2000) A Bayesian analysis of the multinomial probit model with fully identified parameters. *J. Econometrics* 99(1):173–193.

Meredith W (1993) Measurement invariance, factor analysis and factorial invariance. *Psychometrika* 58(4):525–543.

Mulaik SA (1972) *The Foundations of Factor Analysis* (McGraw-Hill, New York).

Newton MA, Raftery AE (1994) Approximate Bayesian inference with the weighted likelihood bootstrap. *J. Roy. Statist. Soc. Ser. B* 56(1):3–48.

Otter T, Allenby G, van Zandt T (2008) An integrated model of discrete choice and response time. *J. Marketing Res.* 45(5): 593–607.

Press SJ, Shigemasu K (1989) Bayesian inference in factor analysis. Olkin I, Gleser LJ, Perlman MD, Press S, Sampson A, eds. *Contributions to Probability and Statistics* (Springer, New York), 271–287.

Rossi P, Gilula Z, Allenby G (2001) Overcoming scale usage heterogeneity: A Bayesian hierarchical approach. *J. Amer. Statist. Assoc.* 96(453):20–31.

Rossiter JR (2002) The C-OAR-SE procedure for scale development in marketing. *Internat. J. Res. Marketing* 19(4):305–335.

Smith AK, Bolton RN, Wagner J (1999) A model of customer satisfaction with service encounters involving failure and recovery. *J. Marketing Res.* 36(3):356–372.

Sonnier G, Ainslie A (2011) Estimating the value of brand-image associations: The role of general and specific brand-image. *J. Marketing Res.* 48(3):518–531.

Spearman C (1904) "General intelligence," objectively determined and measured. *Amer. J. Psych.* 15(2):201–292.

Tanner MA, Wong WH (1987) The calculation of posterior distributions by data augmentation. *J. Amer. Statist. Assoc.* 82(398): 528–549.

Thorndike EL (1920) A constant error in psychological ratings. *J. Appl. Psych.* 4(1):25–29.

Thurstone LL (1937) Current misuse of the factorial methods. *Psychometrika* 2(2):73–76.

van der Lans R, Pieters R, Wedel M (2008) Eye-movement analysis of search effectiveness. *J. Amer. Statist. Assoc.* 103(482): 452–461.

van Rosmalen J, van Herk H, Groenen PJF (2010) Identifying response styles: A latent-class bilinear multinomial logit model. *J. Marketing Res.* 47(1):157–172.

Wang X, Bradlow ET, Wainer H (2002) A general Bayesian model for testlets: Theory and applications. *Appl. Psych. Measurement* 26(1):109–128.

Wedel M, DeSarbo WS (2002) Market segment derivation and profiling via a finite mixture model framework. *Marketing Lett.* 13(1):17–25.

Wedel M, Kamakura WA (2000) *Market Segmentation: Conceptual and Methodological Foundations* (Kluwer Academic Publishers, Norwell, MA).

Ying Y, Feinberg F, Wedel M (2006) Leveraging missing ratings to improve online recommendation systems. *J. Marketing Res.* 43(3):355–365.