Modeling the Effect of Images on Product Choices

Jeffrey P. Dotson
Brigham Young University
jeff.dotson@byu.edu

Mark A. Beltramo
General Motors
mark.beltramo@gm.com

Elea McDonnell Feit
Drexel University
efeit@drexel.edu

Randall C. Smith
1randall.c.smith@gmail.com*

July 3, 2016

*We thank Vanderbilt eLab for their assistance in collecting the data we present and The Modellers, LLC for use of their software for the design of choice experiments.

Modeling the Effect of Images on Product Choices

Abstract

Conjoint is one of the most popular methods in marketing research, widely used to understand how customers trade-off features of a product. Since product images have a strong influence on customer choice, it is natural to want to include images in conjoint studies, yet this has proven to be difficult, since images are difficult to parsimoniously characterize in the utility function. This paper proposes a novel approach to account for the effect of images on respondents' choices, in which consumer heterogeneity in the appeal of the images is modeled through the covariance structure in a probit model. The covariance structure is informed by a separate task where respondents rate the images included in the study. In our application to midsize crossover vehicles, we show that our approach readily scales to a large number of images, fits better than several alternatives commonly used in practice, and makes more reasonable predictions about product substitution when a new product enters the market. We discuss how this approach could be used predict the effect of other difficult-to-characterize product attribute such as sound quality or taste on product choice.

Keywords: Choice Models, Conjoint, Images, Visual Design, Styling, Bayesian Methods, Automotive Industry, Multinomial Probit Model

1 Introduction

Choice-based conjoint is arguably the most widely-used tool in marketing research (Wittink and Cattin, 1989; Green and Srinivasan, 1990; Green et al., 2001), making it easy for product designers to quantify the relative importance of product attributes on consumer preferences. In most conjoint studies, the attributes of a product are described in words or numbers (cf. Louviere et al., 2000; Orme, 2006), yet we have long known that the visual design of a product can have a large influence on consumer choice (Kotler and Rath, 1984; Yamamoto and Lambert, 1994; Bloch, 1995; Page and Herr, 2002; Cox and Cox, 2002; Bloch et al., 2003; Creusen and Schoormans, 2005; Landwehr et al., 2011) and that consumers process images differently than they process text (Holbrook and Moore, 1981). So, it isn't surprising that practitioners want to understand how product design will influence product choice relative to other attributes. Practitioners frequently field choice surveys like that in Figure 1, where each alternative is described with both a picture and a set of verbal attributes (as described by Green and Srinivasan, 1990).[1] This image-plus-text format is very natural for respondents, mimicking the typical product page at an online retailer, and customer responses to questions like that in Figure 1 have the potential to provide marketers with information about how customers value the product design shown in the image relative to other features, an important consideration for product designers (Krishnan and Ulrich, 2001).

While conjoint tasks like the one depicted in Figure 1 are quite popular in practice, incorporating the effect of these images into a choice model is not straightforward. Since the images that are displayed convey information about the alternative not captured in the text descriptions, the effect of the images themselves should be accounted for in the choice model used to analyze the data. The most common approach in practice is to estimate the effect of each image (using a dummy variable in the utility specification), but since this requires an additional parameter for each image, it rapidly becomes infeasible as the number of images included in the conjoint design increases. For example, in our application, we include 32 distinct images that can be paired with any combination of the other attributes in the choice survey (see Figure 2). Conjoint respondents answered 20 choice tasks each with 4 alternatives, which means that each image appeared in the choice tasks only 2-3 times for each respondent. This survey design, which is typical of commercial conjoint surveys,

[1] Note that these images do not illustrate the text-based attributes, as in Vriens et al. (1998), but provide additional information about the product.

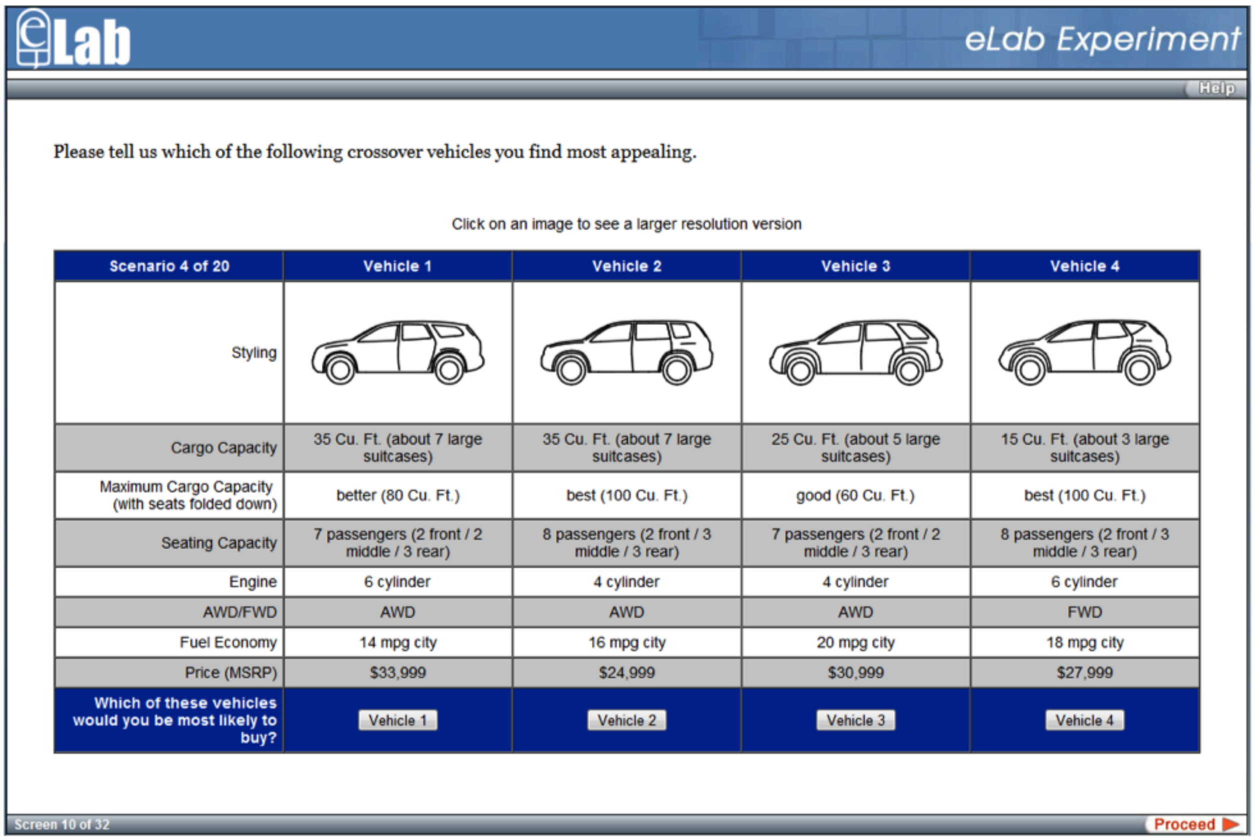


Figure 1: A choice task that includes images in addition to verbal attributes.

is insufficient to estimate effects for all images, particularly when accounting for heterogeneity in preferences for images. Extending this standard approach to larger numbers of images would require either more respondents or more questions for each respondent and rapidly becomes infeasible in commercial conjoint studies. We propose an alternative, highly-scalable approach for modeling the effect of images on product choice, which we summarize next.

1.1 Data Collection and Modeling Approach

The key idea underlying our approach is to augment the choice task in Figure 1 with direct ratings of the appeal of the images from a task like that in Figure 3. As we show, this additional data can be parsimoniously incorporated into the choice model through the covariance of a multinomial probit model (McCulloch and Rossi, 1994; Haaijer et al., 1998).

The ratings of the images can be collected from a different sample of respondents, drawn from the same population as was the conjoint sample. To further minimize the burden on respondents, a split-questionnaire can be used so that each respondent only has to rate a subset of the images (Raghunathan and Grizzle, 1995; Adigüzel and Wedel, 2008). This additional data can be used to estimate the mean appeal rating for each image across the population, as well as the heterogeneity in preferences for the images (i.e., the variance and covariance of the ratings across consumers).

To incorporate this additional information in the choice model, we include the population mean appeal rating for each image as an "attribute" in the specification of the product utility. However, as we will show, simply using the mean ratings as an attribute in the usual hierarchical multinomial logit model only accounts for differences in average preferences for the images and ignores the heterogeneity in the appeal of the images. To account for the fact that respondents vary in which images they prefer, we analyze the choice data using a multinomial probit model where the covariance structure of the error in choice utility is a function of the estimated correlations in the consumer ratings for the images. That is, if people who rate one image highly also tend to rate the other image highly, then the utility between products with those two images is correlated through the covariance in the probit utility. In this way, our approach is akin to spatial modeling approaches where similarities or distances enter the model through a covariance structure (cf. Banerjee et al., 2014). Our proposed specification allows us to decompose the estimation problem into estimating 1) a mean and covariance matrix for the appeal of the images using data collected specifically for

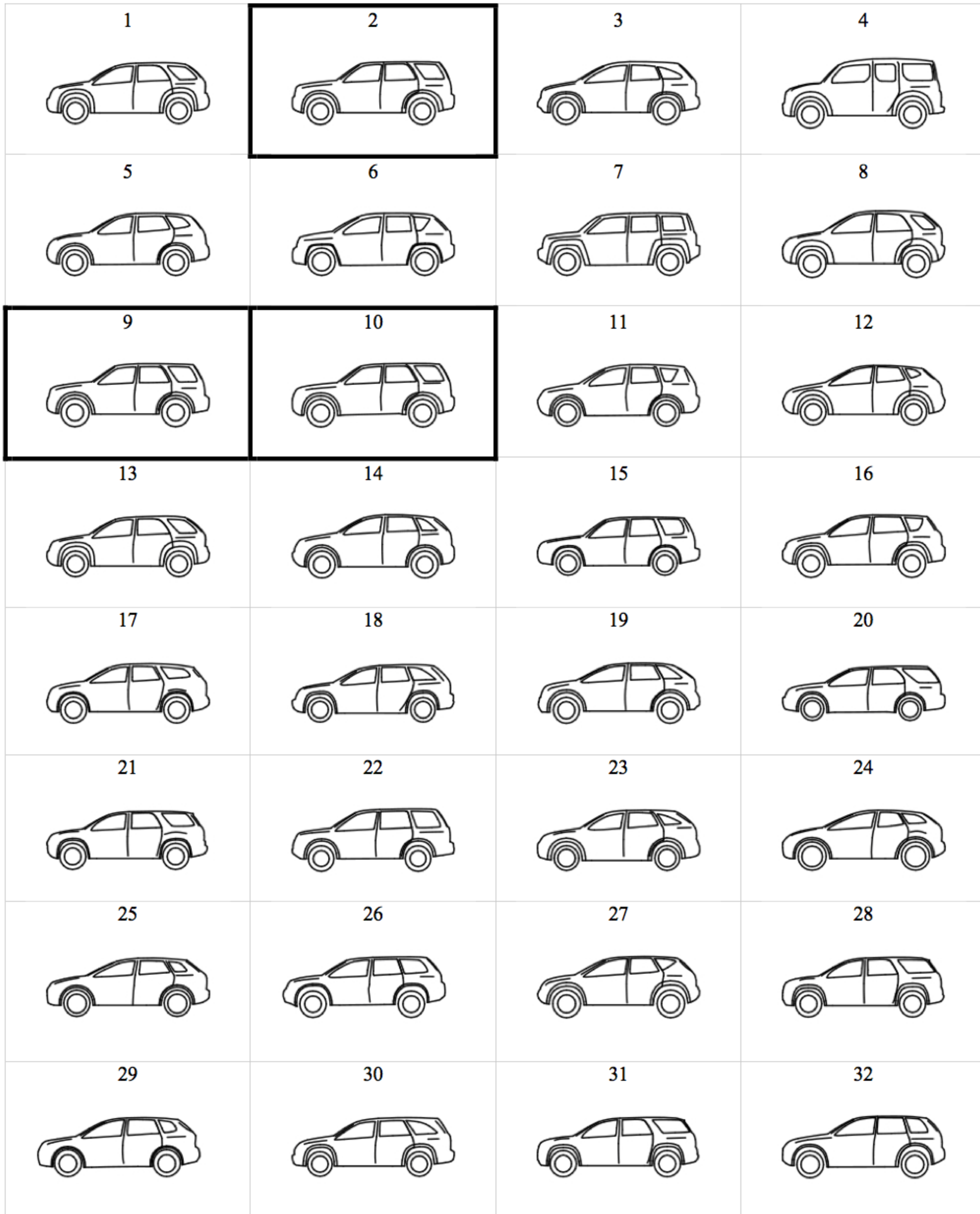


Figure 2: Images included in the choice task. Three visually similar “sister” vehicles are highlighted.

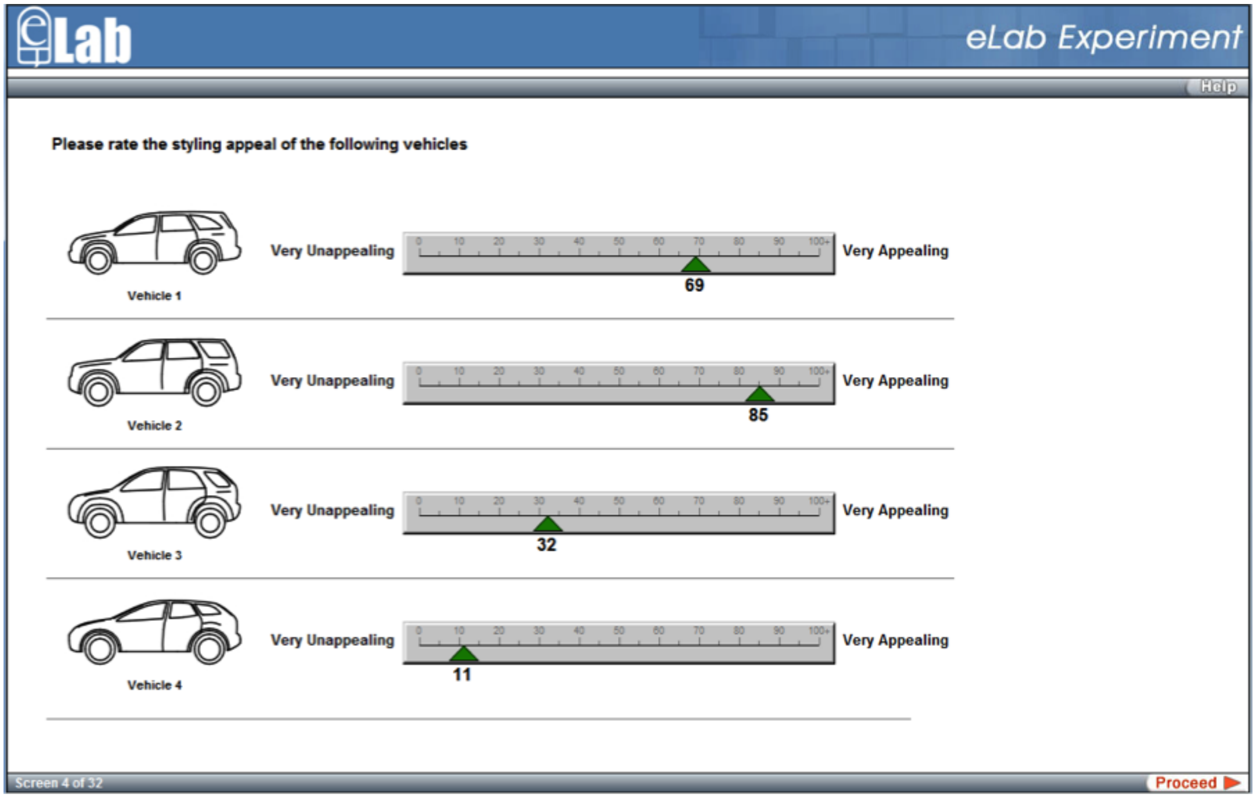


Figure 3: Product image ratings task.

that purpose, and 2) a choice model with only one additional parameter per conjoint respondent, regardless of how many images are included in the study. Incorporating the effect of the images in this way is more parsimonious than the approach of estimating an effect for each image using conjoint data only, which requires $K-1$ parameters for K images. We also show that the covariance in the appeal ratings can be replaced by a measure of "distance" between pairs of images that is computed directly from the images themselves, without reducing the model fit, further simplifying the requirements for the ratings task in Figure 3, as we will discuss. The proposed model is not only motivated by theory and a better fit to the data, but also results in substantially different predictions for product share than a model that does not account for heterogeneity in the appeal of images.

In particular, we will show that the proposed model makes substantially different predictions for new product entries when those products have an image that looks similar (and so appeals to the same subgroup of respondents) as other products already in the market. This has important implications; since it is expensive to develop new product designs, companies frequently consider developing product extensions that will look similar to existing products, yet will have different price, brand and features. For example, in our application, we apply this method to estimate the effect of product images relative to other important features such as fuel economy, seating and price for midsize crossover vehicles. In this market, there are a number of "sister vehicles" that share a common engineering platform and have similar product styling, thus saving on design and tooling costs, but are marketed under different brands at different price points. For example, the three similar-looking vehicles highlighted in Figure 2 are the 2008 models of the Mazda Tribute, the Mercury Mariner and the Ford Escape, which were all produced in the same factory, but marketed by three distinct brands owned by two independent companies. In order to engage in meaningful cost-benefit analysis, designers must be able to make accurate share projections for this set of vehicles (cf. Krishnan and Ulrich, 2001; Desai et al., 2001) in a way that properly accounts for cannibalization. A standard choice model that ignores the styling similarity between these three sister vehicles, will over-estimate the combined share for the three products and underestimate substitution between them. In our application, we show how our proposed model yields substantially different predictions about the combined share for sister vehicles than the hierarchical multinomial logit model commonly used in practice. As we will show, product designers who use naïve approaches common in practice are at risk of coming to different conclusions about whether a sister vehicle will produce enough incremental sales to justify the development costs.

1.2 Related Literature

Our approach draws on the tradition of using probit models to capture non-IIA substitution patterns in choice models. These models have typically been used to estimate non-IIA substitution patterns between brands or SKUs (Chintagunta, 1992; McCulloch and Rossi, 1994), whereas we use the probit covariance structure to capture non-IIA substitution patterns between images. A major challenge with estimating a probit model from choice data alone is that the number of parameters in the covariance structure grows large as the number of alternatives presented to consumers grows. For example, while Chintagunta (1992) and McCulloch and Rossi (1994) both estimate full-covariance probit models from observed purchases, both limit their empirical applications to estimating the covariance structure across 6 brands, while in our application we use a probit covariance structure across 32 images. Our approach to identifying the probit covariance structure relies on collecting additional data (for example, ratings for each image) which is used to inform the covariance structure. The resulting model, which can be scaled to accommodate a large number of images, fully accounts for the heterogeneity in respondent's preferences for the images and heterogeneity in how important the images are relative to other attributes. Thus, our work stands in contrast to approaches that do not collect outside data, but instead rely solely on restrictions of the covariance structure to increase parsimony. For example, Elrod and Keane (1995) impose a latent factor structure on the covariance and Kamakura and Srivastava (1984) assume that the covariance is a function of the attribute similarities between products. Unlike these approaches, which restrict the range of possible substitution patterns, our approach allows for any pattern of substitution between products, and this substitution pattern is informed by the additional data.

Our work is also related to the literature in marketing on the visual design of products, most of which is outside of the literature on choice modeling and conjoint. Much of the literature on visual design has attempted to decompose images into specific features which are predictive of the appeal of the image. For example, Veryzer and Hutchinson (1998) use lab studies to show that the rated appeal of a product design is a function of the cohesiveness between design elements ("unity") and the prototypicality of the design in the category. Similarly, Landwehr et al. (2011) characterize product designs by their prototypically, as well as their "complexity" and show that these dimensions are related to real-world product sales. Similarly, Orsborn et al. (2009) develop a method to describe SUV images in terms of a long vector of dimensions and then use principal components to reduce the dimension of that vector, so that the first few principal components can be used as attributes of the product alternatives in a choice model.

While describing the image in terms of specific features and measuring preferences for those features is useful in helping designers understand what makes an image appealing, these approaches cannot fully account for the effect of images on choice when making predictions. This is because we do not yet fully understand the full set of features that makes an image appealing. In contrast to these other approaches, we do not attempt to decompose the images and instead focus on accounting for non-IIA substitution patterns in choices between images. This allows us to account for the effect of images on product choices without trying to explain what features of those images make them appealing or unappealing and closer or more distant substitutes. Our approach has the added benefit that we do not need to assemble a set of images that spans the design space, as is required to define prototypicality or compute principal components of design variation. In our case study we apply the method to conjoint data that includes vehicle images and show that our approach fits the choice data better and makes more accurate predictions than these alternative approaches to modeling the effect of design on choice.

2 Model Development

As described above, a key component of our approach is to collect ratings of the overall appeal of each image included in the design in an additional survey task (see Figure 3). We model this vector of ratings for each of $k = 1, \ldots, K$ images with a multivariate normal distribution across the population of consumers, i:

$$z_i = (z_{i1}, \ldots, z_{iK}) \sim N_K(\bar{z}, \Sigma) \tag{1}$$

The model in (1) could be replaced with a cut-point model to accommodate ordinal ratings (Rossi et al., 2001); in our application we collected ratings on a near-continuous scale and the normal model is an appropriate approximation and simplifies the exposition.

Using ratings data from any sample of the target population, we can estimate the population average appeal for each image, $\bar{z}$, as well as the population covariance, Σ. The additional ratings data can be collected from the same respondents who answer the conjoint survey, or, if minimizing survey length is important, it can be fielded with a different sample of respondents from the same population. When there is a large number of images to be included in the study, the survey length can be further reduced (at the expense of requiring more respondents) by collecting the ratings data according to a split-questionnaire design where each respondent rates a subset of the images. In our case study, each conjoint respondent rated one of 8 subsets of 16 of the images; subsets were designed so that each pair of images occurred together in at least one of the subsets. Thus, despite some elements of z_i being missing (by design) for some respondents, $\bar{z}$ and Σ can be estimated using the standard Bayesian MCMC algorithm for the multivariate normal model with data augmentation for the missing elements of z_i (Raghunathan and Grizzle, 1995; Adigüzel and Wedel, 2008). In our application, we standardized the ratings before estimation and use conjugate, weakly-informative priors on $\bar{z}$ and Σ. The point estimates of $\bar{z}$ and Σ are then used in fitting the multinomial probit models.

In addition to collecting the ratings data, we also collect choice data where one of the images is included with each product profile (see Figure 1). To motivate our model for the observed choices, we begin by assuming that respondent i's utility for product j follows a random-utility specification that is a function of both the usual product attributes, x_{ij}, and the (unobserved) subject-specific appeal of the image shown with alternative j, z_{ij}:

$$U_{ij} = x'_{ij}\beta_i + \lambda_i z_{ij} + \varepsilon_{ij} \tag{2}$$

where β_i is a vector of parameters that reflect the importance of the attributes (that are described in words) to subject i's overall utility and λ_i is a scalar representing the importance of image appeal to respondent i (relative to the other attributes). As is usual, we assume that ε_{ij} is an IID error term with variance σ^2 and that customer i chooses the alternative with the greatest realized utility.

However, this is not the model we estimate, since z_{ij} is not observed for each respondent i. Instead, we rewrite the utility in (2) as a function of the average appeal for each image, $\bar{z}_j$ where $\bar{z}_j$ is the element of $\bar{z}$ corresponding to the image shown in alternative j. The utility can then be re-written as:

$$\begin{aligned} U_{ij} &= x_{ij}\beta_i + \lambda_i \bar{z}_j + \lambda_i(z_{ij} - \bar{z}_j) + \varepsilon_{ij} \\ &= x_{ij}\beta_i + \lambda_i \bar{z}_j + \eta_{ij} \end{aligned} \tag{3}$$

where the new error term $\eta_{ij} = \lambda_i(z_{ij} - \bar{z}_j) + \varepsilon_{ij}$ incorporates the heterogeneity among respondents in the appeal of the images. Assuming z_{ij} is normally distributed, the resulting vector of errors $\eta_i = (\eta_{i1}, \ldots, \eta_{iJ})$ follows a multivariate Normal distribution with zero mean (since both $z_{ij} - \bar{z}_j$ and ε_{ij} have mean zero). The covariance of η_i is

$$cov(\eta_i) = \lambda_i^2 \Sigma_j + \sigma^2 I \tag{4}$$

where σ^2 is the variance of the IID error terms ε_{ij} and $\lambda_i^2 \Sigma_j$ is the covariance induced by incorporating $\lambda_i(z_i j - \bar{z}_j)$ into the error term, which is simply the marginal of Σ for the images included in task j.[2]

[2] Unlike in Zeithammer and Lenk (2006) the "missing dimensions" of the multivariate normal in the choice task

Our key insight is that (3) can be estimated as a multivariate probit model (with a very specific error structure) using our estimates $\hat{\bar{z}}$ and $\hat{\Sigma}$ from the image ratings data. Conditional on those estimates, we constrain the covariance of the probit model to be equal to $\lambda_i^2 \hat{\Sigma}_j + \sigma^2 I$. The resulting model is:

$$\begin{aligned} U_{ij} &= x_{ij}\beta_i + \lambda_i \hat{\bar{z}}_j + \eta_{ij} \\ \eta_i &\sim N_J(0, \lambda_i^2 \hat{\Sigma}_j + \sigma^2 I) \end{aligned} \tag{5}$$

where $\hat{\bar{z}}_j$ and $\hat{\Sigma}_j$ are point estimates of the mean and covariance of the image ratings in the ratings task, x_{ij} are the observed product attributes and β_i, λ_i and σ are parameters to be estimated from the choice tasks. This probit model allows for flexible substitution patterns between images; the parameter λ_i is identified as respondents make repeated choices from collections of products with images that vary in terms of similarity, as measured by $\hat{\Sigma}$. Note that the parameter λ_i both scales the contribution of $\hat{\Sigma}$ to the covariance and is the coefficient on the mean appeal rating ($\hat{\bar{z}}$) in the deterministic portion of the utility. Thus, our proposed model requires just one additional individual-level parameter in the choice model to estimate the effect of images on choice.[3] Following the usual hierarchical modeling approach, the individual-level parameters, β_i and λ_i, are assumed to follow a multivariate normal distribution across the population.

We estimate the probit model by Bayesian MCMC using standard, weakly-informative, conjugate priors on the population-level parameters (Rossi et al., 2005). As λ_i appears in both the mean and covariance structure, the standard data augmentation based estimation approach for multinomial probit models cannot be used in this context. Rather, we draw realizations of the model parameters from the posterior distribution using the Metropolis-Hastings algorithm, where the model likelihood is computed using the GHK simulator Geweke (1991); Keane (1994). Convergence was assessed by initiating multiple chains using random starting values and using the Gelman-Rubin statistic to confirm that both chains had converged to the same location.

To summarize, our proposed model is a hierarchical multinomial probit model where the utility

do not present any estimation challenges since $\hat{\Sigma}$ is estimated from the ratings data and treated as known when estimating the choice model in (5).

[3] We explored alternative model specifications that allowed estimation of separate parameters to scale the covariance structure and/or weight the mean appeal of the alternative. We found little improvement in fit beyond this more parsimonious and theoretically-motivated model we present here. Results are available from the authors upon request.

is specified as a function of the product attributes and the population mean appeal rating of the image plus an error term that has the covariance structure specified in (5). This formulation accounts for individual differences in which images are most appealing and in how important the images are to product choice overall, yet requires only one additional individual-level parameter, λ_i, to accommodate all of the images. The only added cost of this approach is that of collecting ratings data on each of the images. If necessary, this data can be collected from a different sample of respondents as the conjoint respondents (but, of course, from the same target population) to reduce survey length and respondent burden.

The idea that heterogeneous perceptions of features like styling can induce correlations in the utility structure can be attributed to Hauser and Simmie (1981) who noted that any time there is a transformation from a set of physical design dimensions to a set of perceptual dimensions (in our case from the image to the appeal rating) and that transformation is not modeled directly in the utility function, then a covariance between the utilities of the alternatives is induced (page 44, equation 2). While estimating a full-covariance structure from conjoint data is certainly one way to handle this problem Chintagunta (1992); McCulloch and Rossi (1994), it requires $\frac{K(K-1)}{2}$ additional parameters, where K is the number of images. Our approach offloads this burden to a more appropriate method by relying on additional image ratings to inform the covariance structure. (As we will discuss in Section 4, there are also other, less-costly data sources that can also be used to inform this covariance structure.) Decomposing data collection in this way makes it possible to reduce overall respondent burden and obtain better information about the covariance structure of utility.

3 Application: Conjoint Study for Midsize Crossover Vehicles

The goal of this application is to illustrate the proposed method and to compare the predictions made by our proposed model to several alternative approaches. We do this in the context of a typical conjoint study for midsize crossover sport utility vehicles (SUVs) that included line drawings of the vehicle designs (see Figure 1) as well as a number of traditional attributes like fuel economy and price. (See Table 1 for the complete list of attributes.)

Table 1: The attributes and levels included in the application are typical of a choice model for vehicle purchases.

Attribute	**Level (as described to respondents)**
AWD/FWD	AWD FWD
Fuel Economy	14 mpg city 16 mpg city 18 mpg city 20 mpg city 26 mpg city 30 mpg city
Engine	4 cylinder 6 cylinder 4 cylinder hybrid
Seating	5 passengers (2 front / 3 rear) 7 passengers (2 front / 2 middle / 3 rear) 8 passengers (2 front / 3 middle / 3 rear)
Cargo Capacity	15 cu. ft. (about 3 large suitcases) 25 cu. ft. (about 5 large suitcases) 35 cu. ft. (about 7 large suitcases)
Maximum Cargo Capacity (with seats folded down)	good (60 cu. ft.) better (80 cu. ft.) best (100 cu. ft.)
Price (MSRP)	$21,999 $24,999 $27,999 $30,999 $33,999 $36,999

3.1 Survey Design

The ratings task and the choice questions were both completed by 258 respondents, who were recruited from an online research panel and indicated that they intended to purchase a new car or truck in the next three years and that they would consider a "crossover sport utility vehicle" for their next purchase.

3.1.1 Image Ratings

For the image ratings task (Figure 3), each respondent was asked to rate the "styling appeal" of 16 of the 32 images shown in Figure 2 on a continuous scale with end anchors labeled "Very Unappealing" and "Very Appealing." Each respondent rated one of 8 sets of 16 vehicles; the sets were designed so that each pair of vehicles occurred in exactly 4 of the 8 sets. Thus, the split-questionnaire design provided ample information from which to estimate the covariance in ratings, Σ, while only requiring respondents to rate half of the images.

Because a primary goal of the study was measuring response to vehicle design, the drawings were also carefully constructed to be comparable representations of the basic vehicle shape and proportions. Measuring preference for shape and proportion relative to other attributes like cargo capacity and seating is critical early in the product design process when the basic architecture of the vehicle must be established. (Color and trim details can be determined later in the design process.) The drawings were all scaled to appear the same size, so that respondents would not make inferences about size or roominess based on the images. Any brand information (e.g., a brand logo) was removed from each line drawing, so that the respondents would make minimal inferences about other attributes based on the images. While we chose to use line drawings in this study, we note that the method we propose could be used with any set of representations, including full-color photographs or virtual reality experiences.

To identify the image effect, λ_i, it is critical that the images vary in how appealing they are and how similar they are to each other. For this demonstration of the method, we chose 32 images that represent the midsize crossover SUVs available in the United States in 2008; this group varies substantially in their shape and their similarity to one another (see Figure 2). While the designs tested in this case study happen to represent products that were on the market, the method could

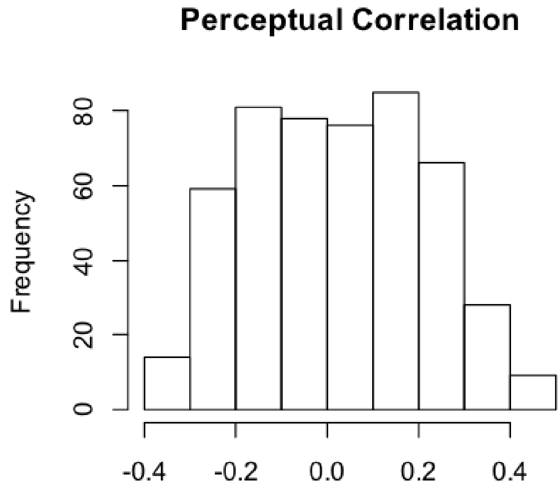


Figure 4: Distribution of cross-image correlation used in the proposed model shows substantial variation across pairs of images.

also accommodate images of new prototype designs as well. In applications where designers are testing new design concepts, the number of images to test could easily exceed 30 or 50 images, making it nearly impossible to estimate fixed effects for each image from the choice tasks alone.

Using the ratings, we estimated the normal multivariate model in (1) using a Bayesian data augmentation approach that accounts for the missing observations for each respondent. For example, the greatest estimated positive correlation in ratings is between image 4 and image 7 (correlation = 0.625), meaning that respondents who rate the boxy image 4 highly also tend to rate the boxy image 7 highly. Similarly, there is a positive correlation in the appeal ratings for images 2 and 9, which represent the sister vehicles Ford Escape and Mazda Tribute which both have traditional SUV styling (correlation = 0.489), while we find a negative correlation in ratings is between the image 2 and image 19, which represents the Ford Edge and has a more modern "crossover" look (correlation = -0.406). The distribution of correlations between pairs of images is summarized in Figure 4.

3.1.2 Choice Tasks

In addition to rating 16 images, each respondent also answered 20 choice scenarios like that in Figure 1. Each question offered the choice from among 4 product profiles that were described in terms of the attributes listed in Table 1 along with one of the 32 pictures in Figure 2. For the traditional attributes, the profiles were selected based on an optimally-efficient Bayesian experimental design (Sandor and Wedel, 2001; Sándor and Wedel, 2005) using weakly informative, zero-centered prior of the form proposed by Lenk and Orme (2009). A total of 12 alternative versions of the conjoint survey were created and respondents were randomly assigned to one version. The 32 line drawings were rotated across tasks in a way that balanced the number of times that any pair of images occurred together in the same task.

3.2 Model Fit

First, to demonstrate how including styling in the model improves the fit of the choice model, we compare our proposed model (Model 2) to two naive models (Model 0 and Model 1). Our first benchmark specification is a hierarchical independence multinomial probit model (i.e., a probit model with a covariance equal to the identity matrix) that includes all of the attributes *except styling* (Model 0). The independence probit is identical to the standard hierarchical multinomial logit model widely used in practice (up to the distributional form of the error term) and so closely represents what a practitioner would estimate if he completely ignored the effect of the images on choice and estimated a hierarchical multinomial logit model for the other attributes. Unsurprisingly, we find that this model fits our data substantially worse than our proposed model (log-marginal density[4] = -3847 versus -3551) strongly suggesting that ignoring the effect of styling on choice entirely is not reasonable. To guard against over-fitting, we also compute the average hit rate[5] for one hold-out question that was not used in estimation. Although the differences in hit rate are more modest, we find that the independence probit that ignores styling performs worse than our proposed model (out-of-sample average hit rate = 0.500 v. 0.534).

[4]We estimate the log-marginal density using the Newton-Raftery estimator. Because this estimator is unstable, we computed it for several sub-samples from the posterior draws and the ranking of the models remained the same. The median and average deviance of the draws also ranked the models in the same order.

[5]We define the average hit rate as the average likelihood of the chosen alternative at the posterior mean of the parameters. This metric favors a model that makes accurate predictions with high certainty over a model that makes accurate predictions with lower certainty.

Table 2: Our proposed model (Model 2) fits the choice data better than naive benchmarks.

Model	**In-Sample**	**Out-of-Sample**	
	Log-Marginal Density	Hit Count	Hit Rate
Model 0 Independence probit ignoring images	-3847.14	150	0.500
Model 1 Independence probit with mean image rating in the utility specification	-3701.79	150	0.510
Model 2 **Proposed model** **with structured covariance**	**-3551.01**	**153**	**0.534**

A benchmark that more closely represents what might typically be done in practice is an independence probit that includes the average styling rating as an attribute in the mean utility specification in (5) but ignores the covariance and sets $\eta_{ij} \sim N(0, \sigma^2 I))$. The comparison between this model (Model 1) and the proposed model (Model 2) represents the additional gain in performance achieved by accounting for the heterogeneity in the appeal of the images through the structured covariance. Model 1 also fits substantially worse than our full specification (log-marginal density = -3702 versus -3551, out-of-sample average hit rate = 0.510 versus 0.534). The fit of Models 0, 1 and 2 are compared in Table 2. In total, the in- and out-of-sample model fit suggests that our proposed model fits the data better than models that ignore the non-IIA substitution patterns between images, resulting from heterogeneity in preferences for the images.

We also find that the other parameter estimates, the part-worths of the conventional attribute levels, are not substantially different between Model 1 and Model 2; the improvement in fit comes almost entirely from the addition of the covariance between images. Estimates of the population-level parameters for our proposed model (Model 2) and the independence probit with styling ratings in the mean (Model 1) are reported in Table 3.

We attempted to fit a model where we estimated a fixed effect for each image as an attribute with 32 levels, requiring 31 additional parameters, but found that the posterior for the individual-level parameters for the images was highly diffuse, to the point where we had trouble obtaining a sample from the posterior using MCMC with diffuse priors. To see how unreasonable this approach is, consider that each respondent only sees each image on average 2.5 times in the choice tasks,

Table 3: Estimates of part-worths are similar between the benchmark independence probit (Model 1) and our proposed model (Model 2)

Estimated Population Mean	**Model 1 Independence Probit**		**Model 2 Proposed Model**	
	Est	**95% HPD**	**Est**	**95% HPD**
FWD (versus AWD)	-0.19	[-0.31,-0.05]	-0.20	[-0.31,-0.08]
Fuel Economy (MPG)	1.41	[1.33,1.85]	1.57	[1.32,1.78]
6 Cylinder Engine (versus 4 Cylinder)	0.51	[0.37,0.75]	0.57	[0.41,0.73]
4 Cylinder Hybrid Engine (versus 4 Cylinder)	0.08	[-0.13,0.20]	0.09	[-0.05,0.24]
7 Passengers (versus 6)	0.35	[0.21,0.54]	0.35	[0.20,0.50]
8 Passengers (versus 6)	0.34	[0.19,0.52]	0.36	[0.20,0.50]
25 cu. ft. Cargo Capacity (versus 15)	0.19	[0.09,0.35]	0.22	[0.11,0.34]
35 cu. ft. Cargo Capacity (versus 15)	0.26	[0.15,0.42]	0.28	[0.12,0.44]
Better Maximum Cargo Capacity (versus Good)	0.05	[-0.05,0.20]	0.08	[-0.05,0.19]
Best Maximum Cargo Capacity (versus Good)	0.27	[0.17,0.43]	0.29	[0.17,0.39]
$24,999 (versus $21,999)	-0.19	[-0.32,-0.01]	-0.17	[-0.30,-0.04]
$27,999 (versus $21,999)	-0.47	[-0.61,-0.24]	-0.43	[-0.59,-0.26]
$30,999 (versus $21,999)	-0.99	[-1.27,0.77]	-0.97	[-1.16,0.78]
$33,999 (versus $21,999)	-1.42	[-1.74,-1.16]	-1.47	[-1.72,-1.20]
$36,999 (versus $21,999)	-1.94	[-2.38,-1.62]	-2.03	[-2.32,1.75]
Average Styling Rating	0.34	[0.26,0.43]		
λ_i (weight for Average Styling Rating and covariance structure)			0.35	[0.24,0.49]

which gives very little information from which to estimate heterogeneous effects for each image. If a respondent never chooses an alternative with a particular image, which is quite common in our data, then the individual's parameter for that image is not identified by the data and depends entirely on the population-level distribution. We were unable to estimate this model without placing highly-restrictive hyper-priors on the population-level covariance of the individual-level parameters (β_i).

3.3 Model Predictions

In the previous subsection, we showed that the proposed Model 2 fits the data better both in-sample and out-of-sample, but the improvement might be considered modest. However, the differences in predictions between Model 1 and our proposed Model 2 are substantial and important in practice. For example, consider the situation where a company is considering introducing a new product to the market that is similar to an existing product. An example of a scenario like this is described in Table 4, where the existing products are described by in columns 1-6. We look at this example from the perspective of the company that produces vehicle 5 in Table 4 and is interested in launching a the new vehicle described in the final column which will compete with vehicle 5 and the other competitors.

The new vehicle, shown in the last column of Table 4, has image 13 in Figure 2, which is similar to the image for vehicle 5 (image 1 in Figure 2). Our model recognizes that these two images have correlated appeals ratings (correlation = 0.282). (Note these are moderately correlated images; some pairs have correlations in appeal rating close to 0.5.)

Figure 5a shows the shares before and after the launch of the new vehicle as predicted by Model 1, which only includes average image appeal in the mean specification and ignores the heterogeneity in image appeal. While the Independence of Irrelevant Alternatives property does not strictly hold for this hierarchical model, we do find that the share predictions show an IIA tendency with share for the new vehicle coming from other vehicles somewhat in proportion to their initial shares. This can be seen clearly in the source-of-volume predictions (on the right in Figure 5a), which show the proportion of share for the new vehicle that comes from each of the existing vehicles. While Model 1 does predict some competition between vehicle 5 and the new vehicle, the new vehicle still gains about half its share from other vehicles. The model optimistically predicts that together, vehicle 5

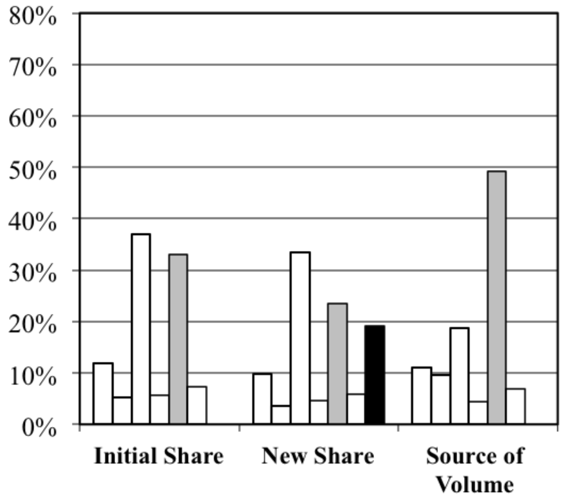
Table 4: A hypothetical new product introduction of a "sister vehicle" to vehicle 5.

Attribute	Veh 1	Veh 2	Veh 3	Veh 4	**Veh 5**	Veh 6	**New**
FWD	0	0	0	0	0	0	0
Fuel Economy	26	18	26	16	20	18	20
6 Cylinder Engine	0	1	0	0	1	0	1
4 Cylinder Hybrid	0	0	1	0	0	0	0
7 Passengers	0	1	0	0	0	0	1
8 Passengers	1	0	1	0	1	1	0
25 cu. ft. cargo	0	0	0	1	0	0	0
35 cu. ft. cargo	1	0	0	0	1	0	1
Better Maximum Cargo	0	0	0	1	0	1	0
Best Maximum Cargo	1	0	0	0	1	0	1
$24,999	0	0	1	0	0	0	0
$27,999	0	0	0	0	1	0	1
$30,999	0	1	0	0	0	0	0
$33,999	1	0	0	0	0	0	0
$36,999	0	0	0	0	0	0	0
Ave. Image Rating ($\hat{\bar{z}}$)	0.07	0.10	0.85	-3.46	1.12	0.08	0.72
Image Number	2	5	3	4	1	6	13

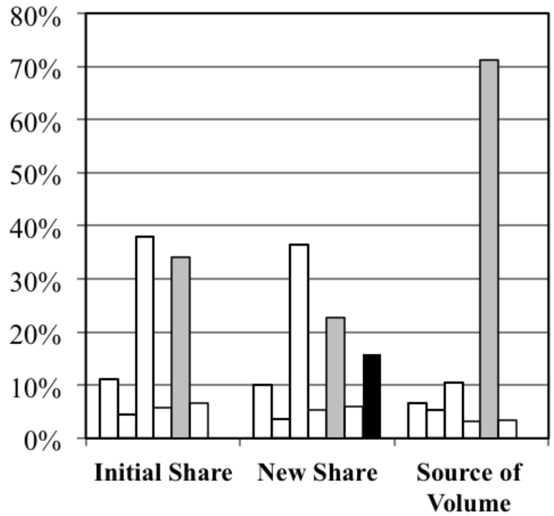
and the new vehicle will achieve 42.3% share.

Our proposed Model 2, in contrast, accounts for the increased competition between vehicle 5 and the new vehicle due to their similar images. The predicted probit covariance matrix for the 7 vehicles after the hypothetical new product launch is shown in Table 5 and reflects the strong predicted competition between vehicle 5 and the new vehicle. The share predictions for the structured covariance probit shown in Figure 5b reflect this strong competition, predicting that 71.2% of the volume for the new vehicle will come from vehicle 5. (Note, as well, that Model 1 and Model 2 predict an almost identical share for vehicle 5 before introduction of the new vehicle.) The combined share of both vehicles together is predicted to be 38.5%, which may not be enough of an increase (over the base share of vehicle 5, which is 34.0%) to justify the additional cost of producing and marketing the new vehicle. While profit calculations would require cost data and are beyond the scope of this paper, the results show that the difference in predictions between Model 1 and Model 2 can be substantial enough to affect critical for business decisions such as whether to launch a sister vehicle.

To see how this example of launching a new similar-looking product compares to launching a new dissimilar looking product, we ran a second counterfactual simulation with the same settings as



(a) Model 1: Independence Probit



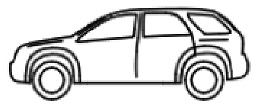
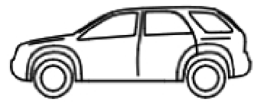
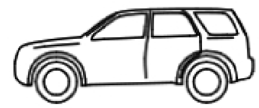
(b) Model 2: Proposed Model

Figure 5: Share predictions from the benchmark independence probit and our proposed model for the hypothetical new vehicle launch described in Table 4. Bars represent the vehicles in Table 4 in the same order. Bars for Vehicle 5 are shaded in grey and the bar for the new vehicle is black.

Table 5: The average predicted probit correlation matrix for the hypothetical scenario described in Table 4 shows strong competition between vehicle 5 and the new vehicle

	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 4	Vehicle 5	Vehicle 6	New Vehicle
Vehicle 1	1.00						
Vehicle 2	-0.28	1.00					
Vehicle 3	-0.35	0.34	1.00				
Vehicle 4	0.08	-0.04	-0.11	1.00			
Vehicle 5	-0.19	0.18	0.26	-0.10	1.00		
Vehicle 6	-0.17	0.16	0.16	0.14	-0.12	1.00	
New Vehicle	-0.07	0.07	0.13	-0.05	**0.48**	0.00	1.00

Figure 6: Predicted product line market share for introduction of a similar versus dissimilar product based on the market described in Table 4.

Model	Predicted Share for Vehicle 5 before New Entry	Predicted Combined Share for Vehicle 5 and New Vehicle for Alternative Product Images	
			
	Image 1	Image 13 (similar)	Image 10 (Dissimilar)
Model 1 Independence Probit	32.9%	42.3%	42.7%
Model 2 Proposed Model	34.0%	38.5%	43.1%

in Table 4, except assuming that the new vehicle has a more distinct image (image 10 in Figure 2). The combined market share for vehicle 5 and the new vehicle predicted by the independence probit (Model 1) are 42.3% when the new vehicle has an image that is similar to vehicle 5 (image 13) and 42.7% when the new vehicle has an image that is dissimilar (image 10). This shows that when we ignore the similarity between images (and the resulting heterogeneity in preferences for images), we do not see a substantial difference in the combined shares for vehicle 5 and the new vehicle. By contrast, our proposed model (Model 2), predicts a substantial difference in the combined share for vehicle 5 and the new vehicle, with 38.5% when the new vehicle has a similar image and 43.1% when it has a dissimilar image. Thus, we can see that there is a substantial difference in the predictions between the two models. A company who plans to launch vehicle 5 and the new vehicle together as a product line would likely choose the similar-looking design if they used Model 1, since similar designs are usually less costly to produce. However, using predictions from Model 2 the company would be able to see the substantial difference in combined share for the product line and (depending on costs) would be more likely to choose to make the investment to differentiate the designs. Combined share for vehicle 5 and the new vehicle for both models under both scenarios is summarized in Figure 6.

4 Alternative Model Based on Physical Dimensions

In this section, we offer an alternative approach to quantifying the differences between images and using these images to approximate the probit covariance. This approach computes the distances between images directly from the images themselves, reducing the need for the appeal ratings data (but not completely eliminating it, as we will discuss shortly.) We do this by quantifying the distance between images based on the features of those images and then using those distances to approximate the probit covariance. We find that this approach performs as well as the model described in the previous section, while requiring less additional survey data from outside the conjoint survey.

The line drawings presented in Figure 2 can be summarized using a shape-space, which is a system for encoding the shape of an object using a minimal number of points. In fact, all of the images in Figure 2 were generated from a shape space, defining the shape of a midsize crossover vehicle based on a vector of 343 key points in 2-D space (see Smith et al., 2007, for more details). Using these points, we can compute a distance between two images and use that distance to approximate the heterogeneity in preferences, $\hat{\Sigma}$ in (5). The key assumption in doing this is that images which are physically similar (as determined by the shape space) are likely to appeal to the same subset of people, and so will be more substitutable.

After removing control points that do not lie on the curves that make up the line drawing, we computed the simple Euclidian distance between the two vectors of points that define any two vehicle images in the shape space. For example, using this metric, the distance between the images 2 and 9 (which are sister vehicles Ford Escape and Mazda Tribute) is 0.029, while the distance between images 2 and 19 (which are the Ford Escape and the Ford Edge) is much greater at 0.065. The distribution of the squared distances is shown in Figure 7. Comparing the distance based on the shape space and estimated correlation in appeal ratings, we find that the distance measure based on the physical dimensions is negatively associated with the correlation in appeal ratings (correlation = -0.586 over 496 pairs of images).

We scaled the physical distance measure to lie in the unit interval and used it to approximate $\hat{\Sigma}$ in (5), rather than using the covariance in appeal estimated from the ratings data. We then estimated the model in (5) and we find the model based on the physical distances between images

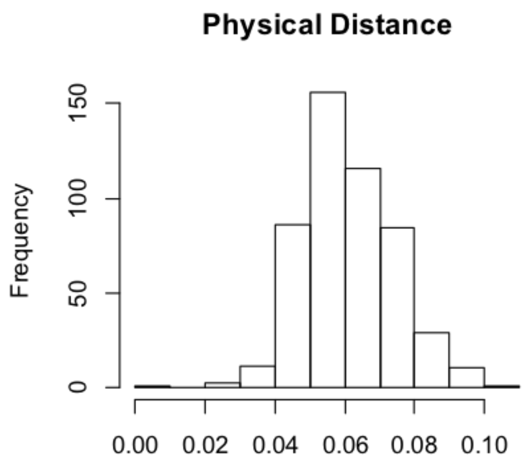


Figure 7: Distribution of physical distances between images.

(Model 3) fits the data as well as the one that uses the estimated covariance based on the ratings (Model 2). In fact, the fit statistics for Model 3 slightly outperform Model 2 (log-marginal density = -3508 v. -3551, out of sample average hit-rate = 0.538 v. 0.534). This suggests that the physical distance metric can be used to form a reasonable approximation to the covariance matrix; that is, two vehicles that are "close" according to the shape-space distance seem to appeal to the same subset of customers and are closer substitutes in the choice task.

We note that Model 3 still uses the average appeal rating from the task in Figure 3 as an "attribute" in the mean utility, so this approach would still require collecting ratings data for each image. However, since we do not need to estimate the covariance matrix Σ from the ratings data, the ratings for each vehicle could be collected from a different, smaller sample of the target market. This would allow the use of short, inexpensive internet based surveys (e.g., Google consumer surveys) to gather the image ratings, with the caveat that care needs to be taken to avoid methods biases. The Model 2 approach described in the previous section requires a survey where each respondent rates a substantial subset of the images so that the covariance is identified, requiring longer surveys for each respondent.

4.1 Comparison to Other Approaches Based on Physical Dimensions

While Model 3 uses the physical dimensions of the image to compute a distance metric between images, several alternative ways to characterize product images so that they can be included in a choice model have been proposed in the literature. Unlike our covariance-based approach, these alternatives primarily focus on summarizing the image, which has a large number of dimensions, using a smaller number of dimensions that can be incorporated directly as attributes in the mean specification of a model (i.e., as elements in x_{ij}) . For example, Landwehr et al. (2011) build on information-processing theory to model consumer choice of images based on two aspects of vehicle styling: the styling “prototypicality” and the styling “complexity”. The prototypicality is computed directly from physical dimensions that describe line drawings (in their case, 50 characteristic feature points of frontal vehicle designs) as the distance between the centerpoint of a group of images and the target image. The complexity is gauged by the size of the image file after compression. Then, using German sales registration data for the first six months of 2007, they regressed each model’s sales in that period on prototypicality, complexity, their interaction, and other typical marketing mix variables. They conclude that, all else equal, vehicles with images that are prototypical, but complex, have higher sales.

To see how well prototypicality and complexity predict choice in our conjoint study, we fit an independence probit model with these two features as attributes in the mean utility specification (Model 4) and show and find that it fits substantially worse than Model 3 (see Table 6). While prototypically and complexity are clearly associated with choice, they don’t seem to be a comprehensive set of all the ways that images affect choice. While our approach does not speak to what makes an image appealing, as Landwehr et al. (2011) do, it accommodates a larger array of substitution patterns (through the covariance) and fits the data better. This means that either of our proposed approaches (Model 2 or Model 3) is preferable when it is important to accurately predict the effect of images on choice. It also suggests that one might combine the two approaches by using prototypicality and complexity as attributes in the utility specification and then basing the covariance on the residuals of the appeal ratings after prototypically and complexity are regressed out.

A second benchmark is proposed by Orsborn et al. (2009). They reduce the large number of

Table 6: Alternative approaches to modeling the effect of images on choice do not fit the data as well as our proposed models.

Model	**In-Sample**	**Out-of-Sample**	
	Log-Marginal Density	Hit Count	Hit Rate
Model 2 **Proposed model** **with covariance based on ratings**	**-3551.01**	**153**	**0.534**
Model 3 **Proposed model** **with covariance based on distance**	**-3508.21**	**153**	**0.538**
Model 4 Complexity and prototypicality	-3695.70	151	0.513
Model 5 Principal components	-3611.12	151	0.521

dimensions of the image by identifying the first few principal components of variation among the images of a set of competitive vehicles. Those principal components can then be used as attributes in the utility specification. We benchmark our approach against this alternative by computing the principal components for the images in Figure 2, based on the points in the shape-space. Our proposed model fits substantially better (LMD=-3551 v. -3611, hit rate = 0.534 v. 0.521). Table 6 summarizes the model fit for Models 2 through 6.

While both of these approaches based on the literature provide some predictive power in our application, our more flexible approach does a better job at capturing departures from IIA substitution and fits the data better. We also note that both of these alternative approaches to characterizing images require the researcher to collect a set of images that represent the products currently on the market (to assess prototypicality or to estimate principal components). Our approach, by contrast, does not require the researcher to establish a set of images that represents the market and only requires that the analyst be able to define a distance metric between the images to be tested. This makes the approach particularly useful in product design applications where designers wish to test a large number of new images and may not be concerned with why particular images are more appealing.

4.2 Making Predictions for New Images After the Model is Estimated

Throughout the design process, product designers are constantly creating new designs, making it difficult to anticipate the full set of images for which the company might want predictions at the time the conjoint study is fielded. Using Model 3, which relies on the distance metric computed directly from the images, it is straightforward to make predictions for new images that were not included in the conjoint survey. With the standard approach of estimating fixed-effects for each image, this cannot be done without collecting additional conjoint data. However, since the distance-based estimate of the covariance is computed directly from the line drawings, we can easily compute a estimate of the new covariance matrix, Σ, for any new line drawings. The only additional data requirement is to estimate the mean appeal ratings for the new drawing, which is something the company would likely want to do anyway, and can be done quickly with a one-question survey. This makes it very easy for the company to rapidly adjust the choice model to understand how the new image will affect consumer's likelihood to purchase a proposed design.

5 Discussion

In this paper, we have presented a parsimonious way to account for the effect of a large number of images on respondents' choices in conjoint studies. Our proposed approach accounts for heterogeneity in the appeal of the images through the covariance structure in a probit model, which is informed by a separate task where respondents rate the images included in the study. Using data collected for midsize vehicles, we have shown that this approach fits better than several alternatives commonly used in practice. Further, we have shown that the proposed model makes different and more reasonable predictions about product substitution when a new product enters the market.

This approach can be adapted to a variety of marketing problems. As we illustrated in our case study, it can be used in nearly any product design context to predict how product design influences product choice. This is a critical application area in practice, where aesthetic considerations are often in direct conflict with other practical features of the product and it is important to understand how consumers will react to alternative designs. Similarly, the approach could be adapted to aid with online merchandising, helping a retailer understand which product images will make an online shopper more likely to purchase a product. The "cost" of our improved ability to account for the effect of images on product choice is that our approach requires some additional data on appeal ratings for each image; data which is relatively simple to collect requiring minimal effort from respondents.

Going beyond images, the structured covariance probit approach could be used to model non-IIA substitution patterns between levels of any attribute with a large number of levels. For instance, to include a large number of brands in a conjoint study, one could collect ratings for the appeal of specific brands and use the heterogeneity in those ratings to inform a covariance structure between those brands, following the same approach we have used for images. Our approach is particularly useful for modeling non-IIA substitution between levels of other aesthetic attributes like sound quality or taste, where each level included in the conjoint study can be represented by a (possibly large) vector of dimensions that can be used to compute a distance. For these types of attributes, the covariance can be based either on ratings or on a similarity metric (minimizing the additional data collection that is required).

There are several limitations of the approach that we should point out. For convenience, we estimate the multinomial normal model for the ratings and multivariate probit model for product choice separately. However, it would be possible to estimate these two models jointly and such an approach would propagate the posterior uncertainly in $\bar{z}$ and Σ into the choice model. This may lead to wider (and more accurate) posteriors for λ_i, particularly when the ratings data does not lead to tight posteriors for $\bar{z}$ and Σ. This would require a relatively straightforward integration of the Gibbs samplers for both models. Second, we should caution that the approach we propose only captures departures from IIA substitution between the images. There are certainly many other potential sources of covariance between alternatives and those sources could be incorporated into this framework.

We see our approach as a practical – but not the ultimate – solution to modeling and understanding preference for styling. As engineers and psychologists begin to better understand how people perceive and form preferences images, they will be able to effectively translate between the pixels in an image and key dimensions that people perceive when they see these images. Once we understand these perceptual dimensions, they can be incorporated directly in the mean utility specification. While the literature has identified some candidates for these perceptual dimensions (Veryzer and Hutchinson, 1998; Landwehr et al., 2011), we are far from a complete theory of visual perception and preference, and product designers continue to face the dilemma of how to incorporate images into choice models. While we hope that others will continue to develop our collective understanding of visual design of products and its effect on product choice, in the absence of a comprehensive theory of perception, our approach provides a practical solution to accounting for the effect of images in conjoint studies, providing product designers with more accurate predictions for how their products will fare in the market.

References

Adigüzel, Feray, Michel Wedel. 2008. Split questionnaire design for massive surveys. *Journal of Marketing Research* **45**(5) 608–617.

Banerjee, Sudipto, Bradley P Carlin, Alan E Gelfand. 2014. *Hierarchical modeling and analysis for spatial data*. CRC Press.

Bloch, Peter H. 1995. Seeking the ideal form: Product design and consumer response. *The Journal of Marketing* 16–29.

Bloch, Peter H, Frederic F Brunel, Todd J Arnold. 2003. Individual differences in the centrality of visual product aesthetics: Concept and measurement. *Journal of Consumer Research* **29**(4) 551–565.

Chintagunta, Pradeep K. 1992. Estimating a multinomial probit model of brand choice using the method of simulated moments. *Marketing Science* **11**(4) 386–407.

Cox, Dena, Anthony D Cox. 2002. Beyond first impressions: The effects of repeated exposure on consumer liking of visually complex and simple product designs. *Journal of the Academy of Marketing Science* **30**(2) 119–130.

Creusen, Marielle EH, Jan PL Schoormans. 2005. The different roles of product appearance in consumer choice. *Journal of Product Innovation Management* **22**(1) 63–81.

Desai, Preyas, Sunder Kekre, Suresh Radhakrishnan, Kannan Srinivasan. 2001. Product differentiation and commonality in design: Balancing revenue and cost drivers. *Management Science* **47**(1) 37–51.

Elrod, Terry, Michael P Keane. 1995. A factor-analytic probit model for representing the market structure in panel data. *Journal of Marketing Research* 1–16.

Geweke, John. 1991. Efficient simulation from the multivariate normal and student-t distributions subject to linear constraints and the evaluation of constraint probabilities 571–578.

Green, Paul E, Abba M Krieger, Yoram Wind. 2001. Thirty years of conjoint analysis: Reflections and prospects. *Interfaces* **31**(3_supplement) S56–S73.

Green, Paul E, Venkat Srinivasan. 1990. Conjoint analysis in marketing: new developments with implications for research and practice. *The Journal of Marketing* 3–19.

Haaijer, Rinus, Michel Wedel, Marco Vriens, Tom Wansbeek. 1998. Utility covariances and context effects in conjoint mnp models. *Marketing Science* **17**(3) 236–252.

Hauser, John R, Patricia Simmie. 1981. Profit maximizing perceptual positions: An integrated theory for the selection of product features and price. *Management Science* **27**(1) 33–56.

Holbrook, Morris B, William L Moore. 1981. Feature interactions in consumer judgments of verbal versus pictorial presentations. *Journal of Consumer Research* **8**(1) 103–113.

Kamakura, Wagner A, Rajendra K Srivastava. 1984. Predicting choice shares under conditions of brand interdependence. *Journal of Marketing Research* 420–434.

Keane, Michael P. 1994. A computationally practical simulation estimator for panel data. *Econometrica: Journal of the Econometric Society* 95–116.

Kotler, Philip, G Alexander Rath. 1984. Design: A powerful but neglected strategic tool. *Journal of Business Strategy* **5**(2) 16–21.

Krishnan, Viswanathan, Karl T Ulrich. 2001. Product development decisions: A review of the literature. *Management Science* **47**(1) 1–21.

Landwehr, Jan R, Aparna A Labroo, Andreas Herrmann. 2011. Gut liking for the ordinary: Incorporating design fluency improves automobile sales forecasts. *Marketing Science* **30**(3) 416–429.

Lenk, Peter, Bryan Orme. 2009. The value of informative priors in bayesian inference with sparse data. *Journal of Marketing Research* **46**(6) 832–845.

Louviere, Jordan J, David A Hensher, Joffre D Swait. 2000. *Stated choice methods: analysis and applications*. Cambridge University Press.

McCulloch, Robert, Peter E Rossi. 1994. An exact likelihood analysis of the multinomial probit model. *Journal of Econometrics* **64**(1) 207–240.

Orme, Bryan K. 2006. *Getting started with conjoint analysis*. Madison, WI: Research Publishers LLC.

Orsborn, Seth, Jonathan Cagan, Peter Boatwright. 2009. Quantifying aesthetic form preference in a utility function. *Journal of Mechanical Design* **131**(6) 061001.

Page, Christine, Paul M Herr. 2002. An investigation of the processes by which product design and brand strength interact to determine initial affect and quality judgments. *Journal of Consumer Psychology* **12**(2) 133–147.

Raghunathan, Trivellore E, James E Grizzle. 1995. A split questionnaire survey design. *Journal of the American Statistical Association* **90**(429) 54–63.

Rossi, Peter E, Greg M Allenby, R McCulloch. 2005. *Bayesian Statistics and Marketing*. John Wiley and Sons, Chichester, UK.

Rossi, Peter E, Zvi Gilula, Greg M Allenby. 2001. Overcoming scale usage heterogeneity: A bayesian hierarchical approach. *Journal of the American Statistical Association* **96**(453) 20–31.

Sandor, Zsolt, Michel Wedel. 2001. Designing conjoint choice experiments using managers' prior beliefs. *Journal of Marketing Research* **38**(4) 430–444.

Sándor, Zsolt, Michel Wedel. 2005. Heterogeneous conjoint choice designs. *Journal of Marketing Research* **42**(2) 210–218.

Smith, Randall, Richard Pawlicki, István Kókai, Jörg Finger, Thomas Vetter. 2007. Navigating in a shape space of registered models. *IEEE transactions on visualization and computer graphics* **13**(6) 1552–1559.

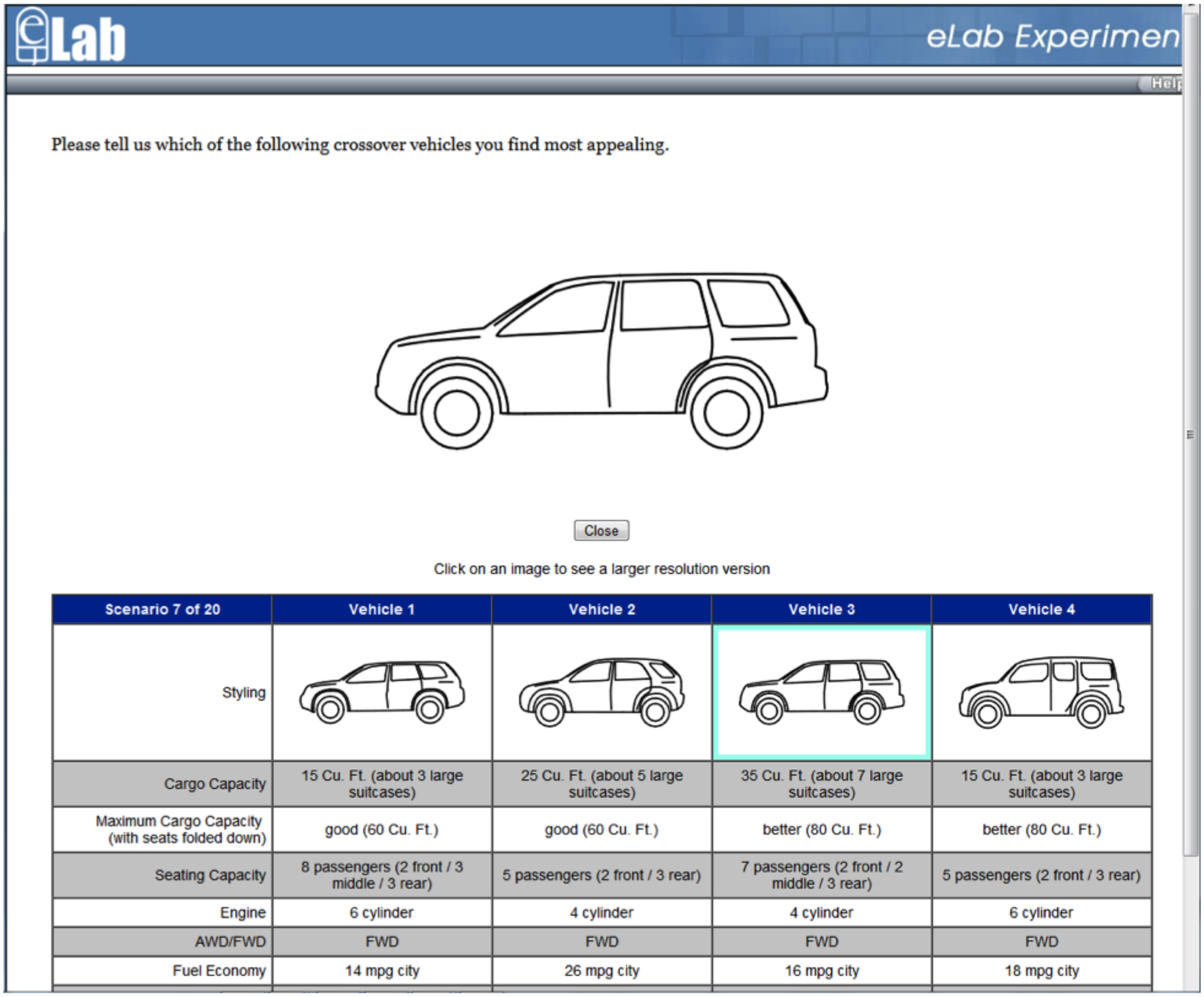
Veryzer, Robert W, J Wesley Hutchinson. 1998. The influence of unity and prototypicality on aesthetic responses to new product designs. *Journal of Consumer Research* **24**(4) 374–394.

Vriens, Marco, Gerard H Loosschilder, Edward Rosbergen, Dick R Wittink. 1998. Verbal versus realistic pictorial representations in conjoint analysis with design attributes. *Journal of Product Innovation Management* **15**(5) 455–467.

Wittink, Dick R, Philippe Cattin. 1989. Commercial use of conjoint analysis: An update. *The Journal of Marketing* 91–96.

Yamamoto, Mel, David R Lambert. 1994. The impact of product aesthetics on the evaluation of industrial products. *Journal of Product Innovation Management* **11**(4) 309–324.

Appendix



Scenario 7 of 20	Vehicle 1	Vehicle 2	Vehicle 3	Vehicle 4
Styling				
Cargo Capacity	15 Cu. Ft. (about 3 large suitcases)	25 Cu. Ft. (about 5 large suitcases)	35 Cu. Ft. (about 7 large suitcases)	15 Cu. Ft. (about 3 large suitcases)
Maximum Cargo Capacity (with seats folded down)	good (60 Cu. Ft.)	good (60 Cu. Ft.)	better (80 Cu. Ft.)	better (80 Cu. Ft.)
Seating Capacity	8 passengers (2 front / 3 middle / 3 rear)	5 passengers (2 front / 3 rear)	7 passengers (2 front / 2 middle / 3 rear)	5 passengers (2 front / 3 rear)
Engine	6 cylinder	4 cylinder	4 cylinder	6 cylinder
AWD/FWD	FWD	FWD	FWD	FWD
Fuel Economy	14 mpg city	26 mpg city	16 mpg city	18 mpg city

Figure 8: To ensure that the styling could be viewed clearly, respondents were able to view a larger version of each picture, by clicking on the thumbnail image in the choice task.