

# Market Structure Mapping with Disentangled Visual Characteristics

(Authors' names blinded for peer review)

Demand models typically use structured data for estimating the value of product characteristics. However, for several product categories such as automobiles, consumers emphasize that visual characteristics of the product are significant demand drivers. Since visual characteristics are typically in high-dimensional unstructured data (e.g., product images), this poses a challenge to incorporate them in demand models. We introduce a method that enables estimation of demand using visual characteristics, by building on the BLP demand model with recent advances in disentangled representation learning. Our method also overcomes the challenge of not having supervised signals, which are required for good disentanglement, by using the demand model as supervisory signal. We discover independent and human interpretable visual characteristics directly from product image data, while simultaneously estimating equilibrium demand in a competitive automobile market in the UK. We conduct a counterfactual analysis using a recent dramatic change in the visual design language of BMW cars, and show our predicted results align with actual changes in BMW market share. To our best knowledge, this work is the first to link visual product characteristics with demand—in other words, to quantify the economic value of design.

*Key words:* visual analytics, deep learning, demand models

---

Exterior look/design is the top reason shoppers avoid a particular vehicle (30%), followed by cost (17%).

— *JD Power Avoider Study 2015*

## 1. Introduction

## 2. Literature Review

Our work integrates two areas of research spanning the fields of machine learning and marketing. The first area encompasses recent advances in machine learning for unstructured data, specifically focusing on “disentangled” representation learning. The second area is market structure mapping, a widely used method in competitive marketing strategy that leverages various data sources to understand the relative positions of products and inform key business decisions. However, prior work on market structure mapping has largely ignored the visual appearance of products. In this paper, we bridge these two research streams by discovering interpretable visual product characteristics from images using disentangled representation learning, and incorporating this information into the market structure mapping process. The following sections discuss each of these research areas in further detail, highlighting their relevance to our work.

### 2.1. Disentangled Representation Learning

Representation learning is a machine learning sub-field that theorizes that high-dimensional data is generated from low-dimensional factors. According to [Bengio et al. \(2013\)](#), the goal is to learn data representations that simplify the extraction of useful information for building predictive models. This paper focuses on a branch of representation learning called disentangled representation learning, which aims to isolate meaningful factors of variation in data [Bengio et al. \(2013\)](#). Take the dSprites dataset (Higgins et al., 2017) as an example, which contains 2D images of objects with different shapes, sizes, colors, and positions. Disentangled representation learning seeks to separate these factors, identifying shape, size, color, and position as four latent dimensions. Notably, disentanglement identifies only the real factors of variation, regardless of the dimensionality of the latent space. We use this method to automatically learn (discovery) and quantify human-interpretable visual characteristics of products without human labeling or intervention.

One key challenge of any disentanglement method is that, with purely unsupervised methods, there is no theoretical guarantee for learning unique disentangled representations [Locatello et al. \(2019\)](#). In other words, we need some form of relevant supervision to

identify independent and semantically interpretable visual characteristics. To address this challenge, Locatello et al. (2020) showed that a small number of labelled examples with even potentially imprecise and incomplete labels is sufficient to perform model selection to learn disentangled representations. However, since our aim to identify the visual characteristics automatically without specifying them in advance, we can not use this approach. Instead, in this paper, we address this theoretical challenge by obtaining alternative supervisory signals derived from a classic model of market equilibrium (Berry et al. 1995). The model of market equilibrium provides us with product fixed effects that can be used as supervisory signals. Implicitly, we assume that product fixed effects capture consumer preferences on unobservable product characteristics including visual characteristics. We compare this alternative source for supervisory signals with signals obtained from structured product characteristics to study whether they alone or together allow us to overcome the well-known challenge in the deep learning literature.

## 2.2. Market Structure Mapping

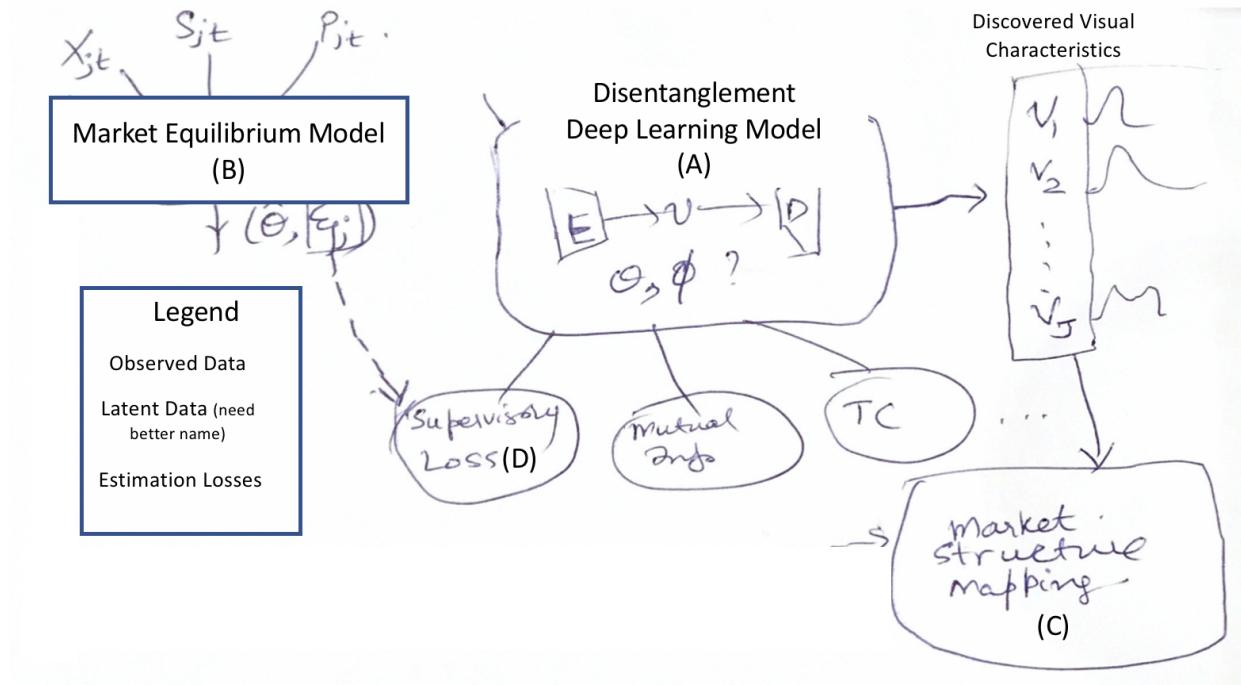
Market structure mapping is one of the primary and commonly used methods in competitive marketing strategy (Rao et al. 1986). Firms use it to understand the relative positions of their products with respect to their rivals in order to inform brand positioning; new product development; and product, advertising, and pricing strategies (Urban et al. 1984, DeSarbo et al. 1993, Bergen and Peteraf 2002, Lattin et al. 2003, DeSarbo et al. 2006). Market structure methods in marketing have used a variety of data sources, such as panel-level scanner data (Erdem 1996), consumer search data (Ringel and Skiera 2016, Kim et al. 2011), online product reviews (Lee and Bradlow 2011, Tirunillai and Tellis 2014), social media engagement data (Liu et al. 2020, Yang et al. 2022), co-occurrence of products in shopping baskets (Gabel et al. 2019) as well as co-occurrence of products in product reviews (Netzer et al. 2012).

The input to commonly used methods to map the market structure is a similarity matrix between all pairs of products. Existing work on market structure mapping has ignored the visual appearance of products. Although it can be overcome using the survey approach by asking consumers to provide visual similarity scores between different pairs of products, it would still be very inefficient and not automatic because we would need to cover all  $n^2$  pairs and would also suffer from a bias from human raters if they focus on different aspects of design. Moreover, the absence of underlying visual characteristics in this approach would

not inform managers on the reason why two products are located close together or further apart on the market structure map. Without understanding the specific visual characteristics driving perceived similarity, managers cannot make informed decisions about product design changes or positioning strategies. In this work, we discover interpretable visual product characteristics from product images to inform the market structure mapping process. By identifying the key visual dimensions that shape consumer perceptions, our approach has the potential to provide insights for managers looking to differentiate their products or reposition them in the market.

### 3. Methodology

**Figure 1** Overview of Methodology



Our method for discovering and mapping the visual product characteristics consists of two main components, as illustrated in Figure TODO. First, we employ a disentanglement-based approach using Variational Autoencoders (VAEs) to obtain visual product characteristics that are independent and human-interpretable. However, disentanglement methods typically require ground truth visual characteristics as supervisory signals, which are not available in our case since we aim to discover these characteristics. To address this challenge, we propose using supervisory signals from a model of market equilibrium that

leverages traditional structured data. Next, we combine the structured characteristics and the discovered visual characteristics to create competitive market structure maps using Multidimensional Scaling (MDS). In the following subsections, we discuss each of these components in detail, starting with the disentanglement approach using VAEs (Section 3.1), followed by the market structure mapping methodology (Section 3.2).

### 3.1. Disentanglement with Variational Autoencoder

Our approach to discovering visual product characteristics builds upon the framework of Variational Autoencoders (VAEs) (Kingma and Welling 2014), a class of deep generative models that learn to encode input data into a latent space, and simultaneously enable the generation of new data samples from this latent space. VAEs have been widely used for disentangled representation learning, which aims to identify statistically independent and semantically meaningful factors of variation in the data (Bengio et al. 2013).

We consider a dataset  $\mathbf{m}$  of product images, which we assume are generated from a distribution parameterized by a set of visual characteristics  $\mathbf{v}$ . Our goal is to learn a low-dimensional representation of these visual characteristics that captures the most salient factors of variation in the product images. To this end, we incorporate additional regularization terms following the  $\beta$ -TCVAE approach (Chen et al. 2018) and a supervised loss term to overcome the challenge posed by Locatello's theorem (Locatello et al. 2019) into the VAE objective function:

$$\begin{aligned}
 \underbrace{L(\theta, \phi; \mathbf{m}, \mathbf{v})}_{\text{Disentanglement Loss}} &= \underbrace{-\mathbf{E}_{q_\phi(\mathbf{v}|\mathbf{m})} [\log p_\theta(\mathbf{m}|\mathbf{v})]}_{\text{Reconstruction Loss}} + \underbrace{I_q(\mathbf{v}, \mathbf{m})}_{\text{Mutual Information Loss}} + \lambda_1 \underbrace{KL \left[ q(\mathbf{v}) \parallel \prod_{j=1}^J q(v_j) \right]}_{\text{Total Correlation Loss}} \\
 &+ \underbrace{\sum_{j=1}^J KL [q(v_j) \parallel p(v_j)]}_{\text{Dimension-Wise KL Divergence Loss}} + \lambda_2 \underbrace{P(\widehat{\mathbf{y}(\mathbf{v})}, \mathbf{y})}_{\text{Supervised Loss}}
 \end{aligned} \tag{1}$$

In this objective function, the VAE consists of an encoder network  $q_\phi(\mathbf{v}|\mathbf{m})$ , which maps the high-dimensional product images  $\mathbf{m}$  to a lower-dimensional latent representation  $\mathbf{v}$ , and a decoder network  $p_\theta(\mathbf{m}|\mathbf{v})$ , which reconstructs the images from the latent representation. Both the encoder and decoder are parameterized by deep neural networks with parameters  $\phi$  and  $\theta$ , respectively.

The VAE assumes a generative model where the latent visual characteristics  $\mathbf{v}$  are first sampled from a prior distribution  $\rho(\mathbf{v})$ , set to an isotropic unit Gaussian  $\mathcal{N}(0, \mathbf{I})$ . The product images are then generated from the decoder distribution  $\rho_\theta(\mathbf{m}|\mathbf{v})$ . During training, the true posterior distribution  $p(\mathbf{v}|\mathbf{m})$  is approximated by the encoder distribution  $q_\phi(\mathbf{v}|\mathbf{m})$ , which is typically assumed to be a multivariate Gaussian with diagonal covariance, i.e.,  $\log q_\phi(\mathbf{v}|\mathbf{m}) = \log \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_d, \boldsymbol{\sigma}_d^2 \mathbf{I})$ , where  $\boldsymbol{\mu}_d$  and  $\boldsymbol{\sigma}_d$  are the mean and standard deviation outputs of the encoder network.

The objective function in Equation 1 is designed to encourage the VAE to learn a disentangled representation of the visual characteristics present in the product image data. It consists of several key loss terms:

- Reconstruction Loss ( $-\mathbf{E}_{q_\phi(\mathbf{v}|\mathbf{m})} [\log \rho_\theta(\mathbf{m}|\mathbf{v})]$ ): This term measures the discrepancy between the original input images  $\mathbf{m}$  and the reconstructed images  $p_\theta(\mathbf{m}|\mathbf{v})$  generated by the decoder network. Minimizing this loss ensures that the learned latent representation  $\mathbf{v}$  captures sufficient information to faithfully reconstruct the original images.
- Mutual Information Loss ( $I_q(\mathbf{v}, \mathbf{m})$ ): It quantifies the amount of information shared between the latent representation  $\mathbf{v}$  and the input images  $\mathbf{m}$ . Penalizing this term encourages the model to learn a compact representation that discards irrelevant details and focuses on the most salient aspects of the images.
- Total Correlation Loss ( $KL \left[ q(\mathbf{v}) \parallel \prod_{j=1}^J q(v_j) \right]$ ): This term measures the dependence between the individual dimensions of the latent representation  $\mathbf{v}$ . By penalizing the KL divergence between the joint distribution  $q(\mathbf{v})$  and the product of its marginals  $\prod_{j=1}^J q(v_j)$ , this loss term promotes statistical independence among the learned visual characteristics. The hyperparameter  $\lambda_1$  controls the strength of this penalty.
- Dimension-Wise KL Loss ( $\sum_{j=1}^J KL [q(v_j) \parallel \rho(v_j)]$ ): This term minimizes the KL divergence between each individual latent dimension  $q(v_j)$  and its corresponding prior distribution  $p(v_j)$ , typically assumed to be a standard Gaussian. Penalizing this divergence regularizes the latent space and encourages the learned representation to conform to a smooth and continuous distribution.
- Supervised Loss ( $MSE(\widehat{\mathbf{y}}(\mathbf{v}), \mathbf{y})$ ): The supervised loss term measures the discrepancy between the predicted supervisory signals  $\widehat{\mathbf{y}}(\mathbf{v})$  and the ground truth signals  $\mathbf{y}$ . By minimizing this loss, the model is encouraged to learn visual characteristics that are predictive of

these signals. The hyperparameter  $\lambda_2$  balances the importance of this supervised objective with the other unsupervised loss terms.

Our methodological contribution lies in leveraging supervised learning via a model of market equilibrium to address a fundamental challenge in disentangled representation learning. Locatello et al. (2019) proved that unsupervised disentanglement learning is fundamentally impossible for arbitrary generative models without inductive biases on both the models and the data. This result highlights the need for appropriate inductive biases or supervisory signals to learn disentangled representations in practice. Consequently, recent efforts in the deep learning literature have focused on improving disentanglement methods by utilizing benchmark datasets with known ground truth labels corresponding to each visual characteristic (Locatello et al. 2020). However, the very visual characteristics we aim to discover are precisely these ground truth labels.

Inductive biases, in the context of disentangled representation learning, refer to the assumptions, preferences, or prior knowledge that a learning algorithm incorporates to guide the learning process and generalize beyond the training data. These biases can be implicit or explicit and are essential for learning meaningful and interpretable representations from limited data. Some examples of inductive biases include architectural choices (e.g., convolutional neural networks for image data), prior distributions for latent variables, regularization techniques, and data augmentation.

This is a different category of signals.

In our work, we propose a novel approach that uses a model of market equilibrium (Berry et al. 1995) to provide supervisory signals and introduce inductive biases. We posit that consumers have preferences over visual product characteristics, which in turn impact their choices and ultimately affect market outcomes. Thus, metrics based on market outcomes are likely to be correlated with visual product characteristics, implying that they can serve as relevant supervisory signals to help discover and obtain disentangled visual characteristics. However, since other factors might also impact market outcomes, we use a model of market equilibrium to isolate the unobserved product characteristic, which includes the impact of all unstructured product characteristics (among other factors).<sup>1</sup> Rather than just using the market shares corresponding to each product as the supervisory signal, we use

<sup>1</sup> It is also possible to use only the demand model to obtain these product-level fixed effects.

the estimated product-level effects as the signal. By leveraging these supervisory signals, we aim to guide the learning algorithm towards discovering meaningful and disentangled visual representations, even in the absence of explicit labels for each factor of variation. The details of the model of market equilibrium are provided in Appendix C.

While it may seem that we have several possible visual characteristics to discover but only one signal for each product, the machine learning literature has shown that even weak supervision with a signal that has some correlation to the ground truth will help obtain a disentangled representation in practice (Locatello et al. 2020). This underlying motivation leads us to include the estimated product-level effects obtained from the market equilibrium model, since the effect is likely *weakly correlated* with each of the visual characteristics. Other aspects (e.g., unobserved quality) may also affect this variable.<sup>2</sup> This approach helps to constrain the space of possible solutions and make the learning problem more tractable, enabling the model to generalize better and learn interpretable representations from limited data.

Another choice of supervisory signals for disentangling visual characteristics is structured product characteristics, as explored by Sisodia et al. (2023). Structured product characteristics, such as brand, material, performance attributes, and price, can be informative for guiding the learning of disentangled representations due to their potential correlation with visual characteristics. For example, a product's brand or the materials used in its construction can significantly influence its visual appearance. A luxury brand like Louis Vuitton or a high-end material like leather may be associated with specific visual design elements that set them apart from other brands or materials. Similarly, the price point of a product can often be reflected in its visual design, with higher-priced items frequently exhibiting more refined or elaborate visual features compared to lower-priced alternatives.

Using structured product characteristics as supervisory signals offers several advantages. These characteristics are often readily available in marketing data and can be easily incorporated into the learning process. They provide a direct link to specific visual attributes of a product, allowing the model to learn representations that are aligned with meaningful product features. However, it is important to note that the effectiveness of structured product characteristics may depend on the relevance and informativeness of the available

<sup>2</sup> However, this approach might be problematic if we want to discover visual characteristics for which consumers have no preferences at all, since the signal would be uncorrelated with those visual characteristics in that case.

attributes for the visual aspects of the product. In some cases, the structured characteristics may not capture all the important visual dimensions, potentially limiting their ability to fully guide the disentanglement process.

In contrast, using a model of market equilibrium to obtain product fixed effects as supervisory signals has its own strengths and limitations. Product fixed effects capture the overall impact of unobserved product characteristics, including visual attributes, on consumer preferences and market outcomes. By leveraging these fixed effects, we can guide the learning algorithm to discover visual characteristics that are more directly related to consumer choices and market performance. However, product fixed effects may also be influenced by factors beyond visual characteristics, such as unobserved quality or brand reputation, which could introduce noise in the supervisory signal.

Ultimately, the effectiveness of different supervisory signals, whether structured product characteristics or product fixed effects from a model of market equilibrium, may vary depending on the specific empirical setting and the nature of the products being studied. In our work, we empirically test the performance of using product fixed effects alone, structured product characteristics alone, and a combination of both as supervisory signals to determine the optimal approach for disentangling visual characteristics in our specific context. By comparing the quality of the learned representations and their predictive power in downstream tasks, we can gain insights into the relative strengths and limitations of each type of supervisory signal for our particular application.

Our model architecture, detailed in Appendix A, is a modified version of the one used by Burgess et al. (2017). We adapt the architecture to work with  $128 \times 128$  pixel images and incorporate the supervisory signals from our model of market equilibrium and structured product characteristics. The architecture consists of an encoder neural network, a decoder neural network, and a supervised neural network, which work together to learn disentangled visual representations and predict the supervisory signals.

In addition to the model parameters that are learned during training, we also make modeling choices in the form of hyperparameters. These hyperparameters impact the estimation process but are not directly estimated with the model. The separation between parameters and hyperparameters is common in machine learning, as hyperparameters are typically set before training and control the learning process, while parameters are learned from data during training (Goodfellow et al. 2016, Murphy 2012, Bishop 2006). Examples

of hyperparameters include the learning rate, batch size, and regularization strengths like  $\lambda_1$  and  $\lambda_2$  in our model. The underlying logic is that hyperparameters define the model's capacity, regularization, and optimization settings, which need to be tuned separately from the model parameters to achieve the best performance and generalization. While there are techniques for automatically searching for optimal hyperparameter values, such as grid search, random search, and Bayesian optimization (Bergstra and Bengio 2012, Snoek et al. 2012), hyperparameters are typically not learned directly during the main model training process.

In our model, we have two key hyperparameters:  $\lambda_1$  and  $\lambda_2$ . The hyperparameter  $\lambda_1$  controls the weight of the total correlation loss within the disentanglement loss (Chen et al. 2018). This term encourages the model to learn statistically independent latent factors. The hyperparameter  $\lambda_2$  represents the weight of the supervised loss when incorporated into the overall loss function. A higher value of  $\lambda_2$  prioritizes the model's ability to predict the vector of supervisory signals, relative to other loss terms like mutual information or reconstruction loss. However, placing too much emphasis on the supervised loss could potentially reduce the quality of disentanglement. On the other hand, setting  $\lambda_2$  to zero implies that we are not addressing the impossibility theorem (Locatello et al. 2019) and thus have no theoretical guarantees for discovering disentangled visual characteristics.

To select the optimal values for  $\lambda_1$  and  $\lambda_2$ , we follow the approach proposed by Locatello et al. (2020). We perform a grid search over a range of values for these hyperparameters and select the combination that yields the lowest 10-fold cross-validated supervised loss for each vector of supervisory signals. This approach allows us to find the best balance between the disentanglement and supervised objectives, tailored to each specific set of supervisory signals.

To compare the quality of disentangled representations produced by different vectors of supervisory signals, we use the Unsupervised Disentanglement Ranking (UDR) metric proposed by Duan et al. (2020). UDR is an automated method that assesses the robustness of disentangled representations to variance at different starting points without requiring access to the ground truth data generative process. It relies on the assumption that for a particular dataset, a disentangling VAE will converge on the same disentangled representation up to certain isomorphic transformations. We select the hyperparameters  $\lambda_1$  and  $\lambda_2$  based on the lowest supervised loss for each vector of supervisory signals and then choose

the supervisory signals that yield the highest UDR score. The details of the UDR algorithm are provided in Appendix B.

### 3.2. Market Structure Mapping

Market structure maps are graphical representations of the positioning of products within a given market. These maps serve as strategic tools for businesses to better understand the competitive landscape and inform marketing, product development, and overall business strategy. By mapping the market, companies can identify gaps, potential opportunities for new products, and the relative positioning of their competitors.

We employ Multidimensional Scaling (MDS), a technique that aims to find a lower-dimensional representation of products such that the distances between them in the lower-dimensional space closely match their original dissimilarities. This is achieved by minimizing a stress function, which measures the discrepancy between the original dissimilarities and the distances in the lower-dimensional space.

Why we care about interpretability of visual characteristics?, Title - Interpretable Positioning Maps, Resnet or Disentanglement - Which one is the ground truth?

To apply MDS, we construct a dissimilarity matrix of pairwise distances between products based on their characteristics. We consider three different sets of product characteristics: structured characteristics alone, visual characteristics discovered through our disentanglement approach, and a combination of both. The dissimilarity matrix is constructed by calculating the Euclidean distances between products based on their normalized characteristics. We normalize each characteristic by subtracting its mean value across all products and dividing by its standard deviation to ensure that all characteristics contribute equally to the dissimilarity measure.

The resulting market structure map provides a visual representation of the competitive landscape, with products positioned in a two-dimensional space based on their similarity or dissimilarity in terms of the considered characteristics. By comparing these market structure maps, we can gain insights into the role of visual characteristics in shaping market structure and competition. The inclusion of visual characteristics can reveal patterns of differentiation and similarity that may not be evident when considering only structured attributes.

An alternative method to create market structure maps based on visual characteristics is to conduct consumer surveys about perceptions of visual similarity. These surveys provide

a dissimilarity matrix that can be used as input for MDS. However, this approach has several limitations compared to our proposed method. First, it relies on subjective human judgments and may suffer from biases introduced by the respondents, who may focus on different aspects of design. Second, it does not provide insights into why products are visually similar, as it considers only pairwise distances without accounting for the underlying visual characteristics. Third, the survey-based approach is not scalable, as it requires collecting data on all pairwise comparisons, which grows quadratically with the number of products ( $n^2$  pairs for  $n$  products in a market).

In contrast, our approach leverages the visual characteristics discovered through disentanglement, which are objective, human-interpretable, and automatically extracted from product images. This allows us to create market structure maps that provide a more comprehensive and interpretable representation of the competitive landscape while being scalable to large product datasets. By examining the underlying visual characteristics, we can understand why products are visually similar or dissimilar and identify opportunities for differentiation based on specific aspects of product design.

#### 4. Empirical Setting and Results

In this section, we present the empirical context, data, and main findings of our study on leveraging visual product characteristics in market structure analysis. We focus on the automobile industry in the United Kingdom (UK) from 2008 to 2017, which is well-suited for this analysis given the importance of visual design in consumer purchase decisions and the competitive nature of the market.

We begin by describing our dataset, which combines information on automobile characteristics, market shares, and images. Next, we discuss the visual characteristics discovered through our disentangled representation learning approach, comparing the performance of supervised and unsupervised methods. To validate the interpretability and quantification of these visual characteristics, we present the results of human subject surveys.

We then construct market structure maps using multidimensional scaling (MDS) based on structured characteristics alone, visual characteristics alone, and a combination of both. These market structure maps allow us to gain insights into the competitive landscape of the UK automobile market, including the differentiation between market segments, the interplay between visual and structured characteristics, and the role of visual characteristics in shaping brand identity and within-brand differentiation.

Finally, we validate our approach by comparing it to a benchmark method based on consumer forums, demonstrating the value of incorporating visual information in market structure analysis. Overall, our results highlight the importance of considering interpretable visual characteristics alongside traditional structured attributes when examining competitive dynamics in the automobile industry.

#### 4.1. Data

We compiled a data set covering 2008 through 2017 consisting of automobile characteristics, market shares and their images from the United Kingdom (UK). We obtain information on sales (in 1000's) and images of the automobiles from DVM-CAR (Huang et al. 2021). Market research studies have shown that up to 70% of consumers identify and judge automobiles by the appearance of headlights and grille located on the face of the automobile.<sup>3</sup> So we only select the images of the front face of the automobiles and ignore other views. Since our sales data comes at the make-model level, we choose the average of product characteristic across trims. We use the make-model fixed effect learned from the BLP demand model in addition to the structured product characteristics to construct supervisory signals for the disentanglement model.

We collected manufacturer suggested retail prices (MSRP), and characteristics of all automobiles sold in the UK from 2008-2017 from Parker's. We have product characteristics for weight, horsepower, length, width, and miles per gallon. The price variable is the list price (in £1000's) for the entry-level trim. Prices in all years are deflated to 2015 UK using the consumer price index. We supplemented the Parker's information with additional information, including vehicle country of production and company ownership information. We also supplemented additional information from the Office of National Statistics, UK. We gathered the price of ultra low sulphur petrol per gallon and ultra low sulphur diesel per gallon as well as the number of households in the UK. Similar to Berry et al. (1995), we calculated miles per UK pound (MP£) as miles per gallon divided by the price per gallon. We measure the market size as the number of households in the UK. We use 'HP/Weight', 'MP£', and 'Space' to construct BLP instruments.

In Table 1, we display summary statistics for the products at the make-model-year level. There are 2439 observations in our sample and a total of 379 distinct models. The variables

<sup>3</sup> URL: <https://www.wsj.com/articles/SB114195150869994250>

include quantity (in units of 1000), price (in £000 units), the ratio of horsepower to weight (in HP per 10 lbs.), the number of ten mile increments one could drive for one £ of gasoline (MP£), tens of miles per gallon (MPG), and size (measured as length times width). We provide sales-weighted means for each variable. We see that automobiles have improved in terms of both power and fuel efficiency over these ten years.

**Table 1 Descriptive Statistics of Structured Data**

Market	No. of Observations	Quantity	Price	HP/Wt	Space	MPG	MP£
2008	233	6.158	21.398	0.416	1.245	4.578	0.760
2009	247	6.159	21.089	0.411	1.229	4.838	0.905
2010	243	6.655	21.584	0.414	1.247	5.022	0.837
2011	231	6.876	21.784	0.421	1.261	5.183	0.782
2012	244	7.028	21.533	0.422	1.264	5.422	0.825
2013	241	8.075	21.351	0.423	1.268	5.573	0.878
2014	251	8.608	21.697	0.431	1.277	5.702	0.962
2015	251	9.148	22.754	0.443	1.290	5.787	1.126
2016	253	9.225	24.067	0.457	1.305	5.692	1.149
2017	245	8.687	24.834	0.465	1.318	5.502	1.053
All	2439	7.685	22.352	0.433	1.274	5.391	0.948

In Figure 2, we display images of 25 automobiles present in our dataset. Note that, we converted color images of size  $128 \times 128$  to grayscale for our study (sales are also not available separately by color). Moreover, our goal is to extract visual characteristics that are related to the shape of the automobile and not related to the color. For each image, we have its associated make, model, year, structured product characteristics and price.

#### 4.2. Discovered Visual Characteristics

We learn the visual characteristics of each make-model sold in the UK between 2008 and 2017 using disentanglement representation learning. We compare the unsupervised approach to learn visual characteristics with supervised approaches. In the supervised approach, we train the learned visual characteristics to predict the supervisory signal associated with each make-model.<sup>4</sup> We follow the hyperparameter selection approach and the UDR metric described in the Methodology section (Section 3.1). From Table 2, we find that the visual characteristics learned from supervising on ‘Make-Model Fixed Effects’ obtained from the model of market equilibrium achieve the best disentanglement in terms of UDR.

<sup>4</sup> We use the following supervisory signals:

1. ‘Price’ of the make-model
2. Combination of structured product characteristics of the make-model (‘HP/Weight’, ‘MPG’, ‘Space’)
3. Make-model fixed effects learned from the demand model (see Appendix D for the demand model estimates)

**Figure 2** Sample of Automobile Images**Table 2 Comparison of Different Supervisory Approaches**

Number of Signals	Supervisory Signals	$\lambda_1$	$\lambda_2$	UDR
1	Make-Model FE	50	40	0.642
0	Unsupervised $\beta$ -TCVAE	50	0	0.608
3	HP/Weight, MPG, Space	50	20	0.614
1	Price	50	50	0.587
1	Unsupervised VAE	1	0	0.071
1	Unsupervised AE	0	0	0.073

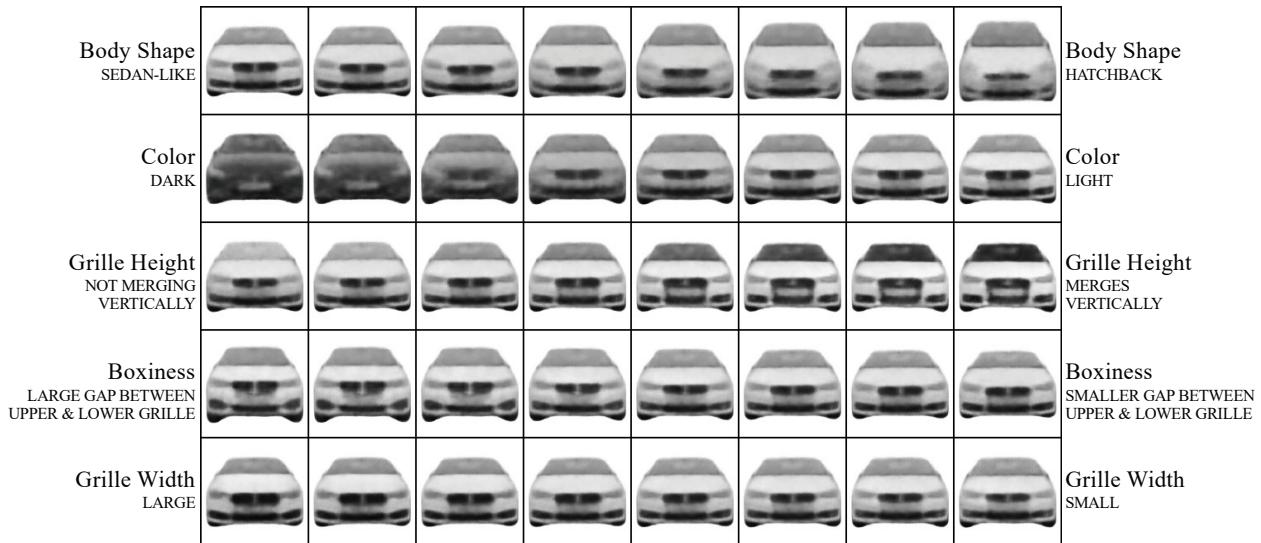
$$\lambda_1 \in [1, 5, 10, 20, 30, 40, 50] \text{ and } \lambda_2 \in [0, 1, 5, 10, 20, 30, 40, 50].$$

We show the discovered visual characteristics in Figure 3 corresponding to the model with  $\lambda_1 = 50$ ,  $\lambda_2 = 40$  and ‘Make-Model Fixed Effects’ supervisory signal. Each row in the image corresponds to a visual characteristic. In each row, we change the value of one visual characteristic while fixing the value of all the other characteristics. Note that since we use a generative deep learning-based method, we can change the underlying learned visual characteristics and generate counterfactual images. The ability to generate counterfactual images allows us to interpret each visual characteristic as it isolates the effect of change in one visual characteristic while keeping the other characteristics fixed. We find five informative visual characteristics of an automobile’s front view, while the rest were uninformative (i.e., changing the visual characteristic produces no change in the image). These informative characteristics are:

1. Body Shape: Automobiles scoring low on this characteristic have a narrower, more angular, and less rounded shape, resembling a sedan. Those scoring high have a wider, less angular, and more rounded shape, resembling a hatchback.
2. Color: Automobiles scoring low on this characteristic are darker, while those scoring high are lighter.
3. Grille Height: As the score of this visual characteristic increases, the grille becomes more prominent, larger, and more defined, with the top and bottom parts beginning to merge.
4. Boxiness: Automobiles scoring low on this characteristic have a high degree of boxiness, characterized by a taller, more upright, and narrower shape. Those scoring high have a lower degree of boxiness, with a lower, flatter, and wider appearance.
5. Grille Width: Automobiles scoring low on this characteristic have a wider, more prominent grille, while those scoring high have a narrower, less pronounced grille.

To validate the interpretability and quantification of the discovered visual characteristics, we conducted surveys with human respondents. The results show that the majority of respondents agreed with each other on the interpretation and with the algorithm on the quantification of the visual characteristics. Detailed information about these surveys can be found in Appendix E.

**Figure 3 Discovered Visual Characteristics**



Left to Right: Vary one visual characteristic, keeping all others fixed

We analyzed the correlations between structured and visual product characteristics and found that they are weakly correlated with each other, as shown in Table 3. The visual product characteristics are largely uncorrelated with each other, with the exception of a weak correlation between boxiness and body shape. In contrast, the structured product characteristics exhibit correlations with each other. Notably, the structured product characteristics and visual product characteristics are only weakly correlated. This suggests that visual characteristics provide additional information not captured by structured characteristics, highlighting the potential value of incorporating visual information in market structure analysis.

**Table 3 Correlation Matrix**

	Structured Characteristics				Visual Characteristics			
	Price	MPG	HP/Weight	Space	Boxiness	Body Shape	Grille Height	Grille Width
Price	1.00							
MPG	-0.60	1.00						
HP/Weight	0.74	-0.48	1.00					
Space	0.67	-0.47	0.36	1.00				
Boxiness	0.05	0.04	0.27	0.08	1.00			
Body Shape	-0.51	0.26	-0.53	-0.36	-0.07	1.00		
Grille Height	0.14	0.05	0.17	0.06	0.02	-0.08	1.00	
Grille Width	-0.11	0.09	-0.08	-0.17	-0.06	-0.07	0.00	1.00

#### 4.3. Visual Market Structure Map

We apply Multidimensional Scaling (MDS) to create market structure maps based on structured product characteristics alone, visual characteristics alone, and a combination of both, as described in the Methodology section (Section 3.2). The structured product characteristics used are Price', HP/Weight', MPG', and Space', while the visual product characteristics are Body Shape', Boxiness', Grille Height', and Grille Width'. Before implementing MDS, we normalize each characteristic by subtracting its mean value across all make-models and dividing by its standard deviation.

Figure 4 shows the market structure of automobiles belonging to the ‘2013’ market. We select B, D and J segment because they have high market share (Segment A (Minicars): 9%, Segment B (Subcompact): 24%, Segment C (Compact): 23%, Segment D (Mid-Size): 9%, Segment E (Mid-Size Luxury): 3%, Segment J (SUV): 27%, Segment M (MPV): 4%) and high variation in characteristics across segments. We show models belonging to the top 5 makes by market share for B, D and J segment in the ‘2013’ market in the maps. Table 4 shows the structured as well as visual product characteristics for the models belonging

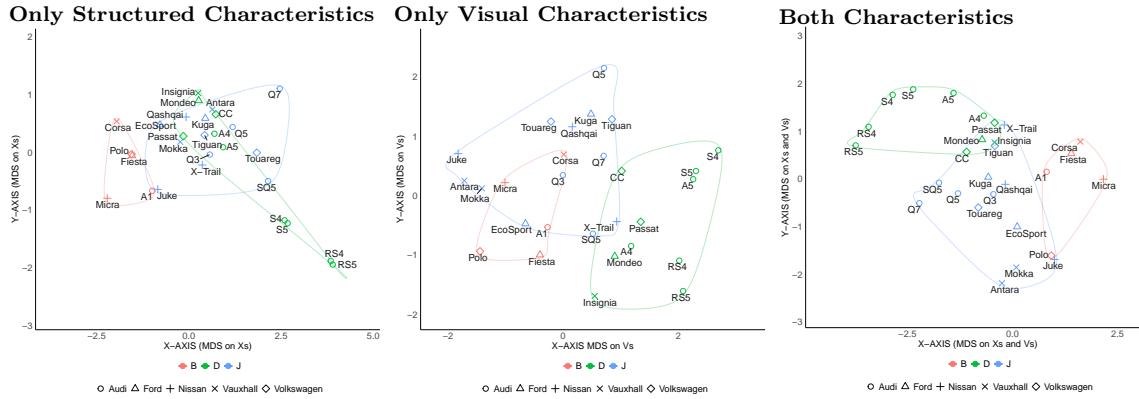
to Segment B, D and J of the top 5 makes by total shares: Audi, Ford, Nissan, Vauxhall, and Volkswagen.

**Table 4 Product Characteristics**

Make Model	Structured Characteristics				Visual Characteristics			
	Price	MPG	HPWT	Space	Boxiness	Body Shape	Grille Height	Grille Width
<b>Segment B (Subcompact)</b>								
Audi A1	19.33	5.98	0.49	1.17	-0.21	0.21	-0.16	-0.66
Ford Fiesta	16.18	6.37	0.41	1.22	0.95	1.47	-0.14	-0.34
Nissan Micra	12.73	5.72	0.39	0.98	-0.61	1.78	-0.21	0.12
Vauxhall Corsa	14.22	5.82	0.31	1.20	-0.53	1.27	1.31	0.54
Volkswagen Polo	15.29	5.53	0.37	1.17	0.17	-0.22	-3.67	-0.78
<b>Segment D (Mid-Size)</b>								
Audi A4	31.47	5.27	0.51	1.49	1.38	0.06	1.52	-0.20
Audi A5	34.44	5.10	0.54	1.47	0.86	-2.21	1.53	0.42
Audi RS4	56.57	2.60	1.12	1.49	1.78	-1.60	1.40	-0.45
Audi RS5	64.84	2.60	1.11	1.46	1.59	-2.56	1.03	-1.21
Audi S4	41.11	3.60	0.86	1.49	1.11	-3.11	0.99	0.99
Audi S5	44.84	3.48	0.85	1.46	1.19	-1.46	1.99	0.84
Ford Mondeo	23.90	5.34	0.44	1.56	1.06	-0.01	1.23	-0.55
Vauxhall Insignia	25.51	5.53	0.43	1.57	1.79	0.87	0.93	-0.73
Volkswagen CC	28.44	5.05	0.48	1.56	0.75	-1.56	-0.71	0.69
Volkswagen Passat	24.69	5.10	0.42	1.35	0.74	-0.75	1.53	-0.23
<b>Segment J (SUV)</b>								
Audi Q2	27.79	5.58	0.47	1.31	-2.21	-2.39	0.92	0.64
Audi Q3	31.88	5.16	0.47	1.37	-3.00	-2.52	1.85	-0.66
Audi Q5	40.71	4.78	0.61	1.55	-2.44	-1.83	1.33	0.36
Audi Q7	59.43	3.97	0.56	1.73	-2.65	-1.09	0.76	-0.33
Audi SQ5	49.44	3.40	0.85	1.55	-0.57	-0.75	1.19	-1.17
Audi SQ7	77.99	3.80	0.83	1.74	0.01	-0.33	0.30	-0.79
Ford EcoSport	17.08	5.32	0.39	1.28	-0.65	-0.17	-1.02	-0.86
Ford Edge	36.96	4.72	0.44	1.63	-1.34	-0.56	-0.90	-0.10
Ford Kuga	29.29	4.33	0.42	1.47	-1.41	0.13	-1.36	0.56
Nissan Juke	20.41	5.08	0.49	1.27	-1.81	-0.48	-2.40	0.54
Nissan Qashqai	24.59	5.77	0.39	1.41	-0.94	-0.67	-1.29	0.06
Nissan X-Trail	28.19	5.06	0.43	1.32	1.14	0.56	-0.82	-0.32
Vauxhall Mokka	23.94	5.15	0.45	1.35	-2.65	0.04	-0.64	-1.35
Volkswagen Tiguan	30.42	4.39	0.45	1.42	-0.48	-1.41	0.24	0.52
Volkswagen Touareg	44.87	4.20	0.50	1.64	-0.24	-1.95	-1.24	0.42

#### 4.4. Insights

*Do visual and structured characteristics provide two independent dimensions of variation or does one restrict the other because they are strongly correlated?* We operationalize this by calculating the average pairwise distance of each make-model to rival make-models in a particular

**Figure 4 (Color Online) Segment B, D & J: Market Structure Map (MDS)**

segment in the structured space as well as the visual space. Next, we calculate the correlation between the average pairwise distances in the structured space and the visual space. Table 5 reports these correlations for the UK automobile market in 2013. A low correlation indicates similarity in structured space does not relate well to the visual similarity. Across all segments, we find that when products are close in structured space, visual space allows the competition to be relaxed. This suggests that even when products are close in terms of structured characteristics, they can still be differentiated in the visual space. In other words, the visual design of products offers a new dimension for brands to position themselves and stand out from their competitors, without being restricted by their structured similarities.

Bentley and Lamborghini - Anecdote, picture from blog, quote from blog, characteristics from us

**Table 5 Correlation Between Distances in Structured Space & Distances in Visual Space**

Segment	Correlation
A (Minicars)	-0.078
B (Subcompact)	-0.080
C (Compact)	0.145
D (Mid-size)	0.074
E (Mid-size Luxury)	0.154
J (SUV)	0.102
M (MPV)	0.031

*Does differentiation across segments increase when visual information is included?* From the market structure map using only structured product characteristics, we note that Segment B is clearly separated from both Segment D and J. However, Segment D and J overlap.

From the market structure map using only visual product characteristics, we note that Segment B and J overlap. However, Segment D is separated from Segment B and J. Interestingly, when we account for both structured and visual product characteristics, all the 3 Segments separate out much more. This means that if one considers only type of characteristic, then the market appears more competitive. However, if one includes both the characteristics, then the market is less competitive. In other words, differentiation across product categories increases when visual information is included.

We verify this empirically by calculating the ratio of non-overlapping area to the total area bounded by the vertices in each of the three market structure maps. We find that in a map using only structured characteristics, 15.1% of the area is overlapping. Interestingly, when we create a market structure map using both structured and visual characteristics, then only 1.5% of the area is overlapping. Note that 39.4% of the area is overlapping in the market structure map created using only visual characteristics.<sup>5</sup>

*Is the competition in the visual space a strategic substitute or complement to the competition in the structured space?* We study this by finding the centroid for each make within a particular segment in a particular year. We refer to them as own-make centroid. Next, we calculate the distance to all other rival-make centroids in that segment. Finally, we find the closest make for each make in a particular segment. We do this for both structured space as well as visual space. Table 6 shows the closest makes in the structured space as well as visual space for the top 10 selling makes in the J segment in the ‘2013’ market.

**Table 6    Closest Within-Segment Rivals in Structured Space & Visual Space**

Make	Quantity Sold	Closest Rival in Structured Space	Closest Rival in Visual Space
<b>Segment B (Subcompact)</b>			
Ford	113390	Volkswagen	Peugeot
Vauxhall	76413	Dacia	Fiat
Volkswagen	39453	Ford	Mitsubishi
Peugeot	37896	Volkswagen	Ford
MINI	31062	Suzuki	DS
<b>Segment D (Mid-Size)</b>			
BMW	39612	Audi	Volkswagen
Mercedes-Benz	31068	Infiniti	BMW
Audi	29955	BMW	Lexus
Vauxhall	24382	Ford	Suzuki
Volkswagen	17964	SKODA	BMW
<b>Segment J (SUV)</b>			
Nissan	82431	Kia	Jeep
Kia	22480	Nissan	Hyundai
Audi	21888	BMW	BMW
Vauxhall	20461	Chevrolet	Ssangyong
BMW	19336	Audi	Audi

<sup>5</sup> Please note that the area calculations are an approximation of those in Figure 4. We construct a convex hull around the coordinates and then calculate the area

*Within Make-Segment Product Differentiation: Role of Visual Space* We operationalize this by comparing the area share of each make in a particular segment in the market structure maps created using only structured space with the ones using only visual space. We consider the top 4 most selling models of a make in a segment. We show the area share of makes belonging to the Segment J in the ‘2013’ market. On the one hand, we can see that Audi has models close together in the visual space but far apart in the structured space. On the other hand, we can see that Nissan has models close together in the structured space but far apart in the visual space.

**Table 7 Area Share of a Make in Structured Space & Visual Space**

Make	Quantity Sold	Models	Area Share (Structured)	Area Share (Visual)
<b>Segment J (SUV)</b>				
Audi	21888	4	17.90%	9.43%
BMW	19336	4	6.35%	5.51%
Jeep	1842	4	9.38%	21.92%
Kia	22480	3	3.87%	4.84%
Mitsubishi	5375	3	0.84%	1.23%
Nissan	82431	4	7.41%	17.33%

#### 4.5. Validation

We use [Netzer et al. \(2012\)](#)‘s forum-based market structure approach as a benchmark to validate our own methodology, which incorporates visual product information alongside typical structured product attributes. [Netzer et al. \(2012\)](#) introduced an approach for deriving market structure from consumer forums. Their key insight is that the co-occurrence of product mentions within forum messages can serve as a proxy for the products’ similarity or substitutability in consumers’ minds. They demonstrate that market structure maps derived from this co-occurrence data align well with those constructed from more traditional data sources like consumer surveys and purchase data. We hypothesize that the inclusion of visual features can enhance the representation of market structure, as visual design is a key driver of consumers’ perceptions and preferences in many product categories.

To test this hypothesis, we employ two validation techniques. First, we compute the correlations between the inter-product distances derived from the forum data and the inter-product distances derived from the structured product characteristics alone; between the inter-product distances derived from the forum data and the inter-product distances

derived from the visual product characteristics alone; and between the inter-product distances derived from the forum data and the inter-product distances derived from both the structured product characteristics and visual product characteristics. If visual information is indeed valuable, we expect the correlation between the distances from forum data and the distances from the combination of structured and visual characteristics to be significantly higher than that between the distances from forum data and the distances from the structured characteristics alone.

Second, we use the distances from structured characteristics alone, visual characteristics alone, as well as a combination of both structured and visual characteristics to predict the distances from forum data. If visual characteristics are informative, the model containing them should exhibit lower error in recovering the “true” market structure reflected in the forums.

To validate our approach, we collected a dataset of online discussions from the Honest John car forum (see URL <https://www.honestjohn.co.uk/forum/threads.htm?f=2>). The data, downloaded on May 1, 2024, covers forum activity from 2008 to 2017 and includes 429,198 threads, 1,175,467 messages, and 3,464,060 sentences. For each year in the data, we first construct a co-occurrence matrix where each cell  $(i, j)$  represents the number of times car models  $i$  and  $j$  are mentioned together in the same message. We then normalize this matrix using the lift metric in Equation 2.

$$\text{lift}(i, j) = \frac{P(i, j)}{P(i) \times P(j)} \quad (2)$$

where  $P(i, j)$  is the probability of models  $i$  and  $j$  co-occurring in a message, and  $P(i)$  and  $P(j)$  are the marginal probabilities of each model being mentioned. This normalization adjusts for the differing overall popularity of make models in the discussions.

After computing the lift matrix, we convert it into a distance matrix by taking the reciprocal of each lift value (see Equation 3). This transformation ensures that higher co-occurrence (and thus presumed similarity) corresponds to smaller distance, while lower co-occurrence (and thus presumed dissimilarity) corresponds to larger distance. For any lift values that are zero (indicating no co-occurrence), we set the corresponding distance to a large number (10 times the maximum finite distance in the matrix) to maintain the

distance interpretation. Next, we perform multidimensional scaling (MDS) on the distance matrix to derive a two-dimensional representation of the market structure.

$$\text{dist}(i, j) = \frac{1}{\text{lift}(i, j)} \quad (3)$$

To assess the contribution of visual attributes in capturing market structure, we examine the mean correlations across 10 years between inter-product distances derived from forum data and those based on structured attributes alone (0.0489), visual attributes alone (-0.0263), and the combination of structured and visual attributes (0.0559). The combined structured and visual attributes yield the highest correlation with forum distances, surpassing the correlation based on structured attributes alone.

To further investigate, we use structured, visual, and combined distances to predict forum-derived distances with random forest regression models. Training separate models for each year (80/20 train/test split), we evaluate predictive accuracy using mean absolute error (MAE). The 10-year average MAEs are: structured distances only (8.0394), visual distances only (8.0577), combined structured and visual distances (8.0505), and both structured and visual distances as separate predictors (7.4462). Using structured and visual distances as separate predictors yields the lowest MAE, suggesting they contain complementary information that improves recovery of the forum-based market structure when used together.

## 5. Discussion & Conclusion

## References

- Bengio Y, Courville A, Vincent P (2013) Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence* 35(8):1798–1828.
- Bergen M, Peteraf MA (2002) Competitor identification and competitor analysis: a broad-based managerial approach. *Managerial and decision economics* 23(4-5):157–169.
- Bergstra J, Bengio Y (2012) Random search for hyper-parameter optimization. *Journal of machine learning research* 13(2).
- Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica: Journal of the Econometric Society* 841–890.
- Berry S, Levinsohn J, Pakes A (1999) Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review* 89(3):400–430.
- Bishop CM (2006) Pattern recognition and machine learning. *Springer google schola* 2:1122–1128.
- Burgess C, Higgins I, Pal A, Matthey L, Watters N, Desjardins G, Lerchner A (2017) Understanding disentangling in  $\beta$ -vae. *Workshop on Learning Disentangled Representations at the 31st Conference on Neural Information Processing Systems*.
- Chen RTQ, Li X, Grosse RB, Duvenaud DK (2018) Isolating sources of disentanglement in variational autoencoders. *Advances in Neural Information Processing Systems*, 2615–2625.
- DeSarbo WS, Grewal R, Wind J (2006) Who competes with whom? a demand-based perspective for identifying and representing asymmetric competition. *Strategic Management Journal* 27(2):101–129.
- DeSarbo WS, Manrai AK, Manrai LA (1993) Non-spatial tree models for the assessment of competitive market structure: an integrated review of the marketing and psychometric literature. *Handbooks in operations research and management science* 5:193–257.
- Duan S, Matthey L, Saraiva A, Watters N, Burgess C, Lerchner A, Higgins I (2020) Unsupervised model selection for variational disentangled representation learning. *International Conference on Learning Representations*.
- Dumoulin V, Visin F (2016) A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285* .
- Erdem T (1996) A dynamic analysis of market structure based on panel data. *Marketing science* 15(4):359–378.
- Gabel S, Guhl D, Klapper D (2019) P2v-map: Mapping market structures for large retail assortments. *Journal of Marketing Research* 56(4):557–580.
- Goodfellow I, Bengio Y, Courville A (2016) *Deep learning* (MIT press).
- Huang J, Chen B, Luo L, Yue S, Ounis I (2021) Dvm-car: A large-scale automotive dataset for visual marketing research and applications. *arXiv preprint arXiv:2109.00881* .

- Kim JB, Albuquerque P, Bronnenberg BJ (2011) Mapping online consumer search. *Journal of Marketing research* 48(1):13–27.
- Kingma DP, Ba J (2014) Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* .
- Kingma DP, Welling M (2014) Auto-encoding variational bayes. *International Conference on Learning Representations*.
- Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25.
- Lattin JM, Carroll JD, Green PE (2003) *Analyzing multivariate data*, volume 1 (Thomson Brooks/Cole Pacific Grove, CA).
- LeCun Y, Bengio Y, et al. (1995) Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks* 3361(10):1995.
- Lee TY, Bradlow ET (2011) Automated marketing research using online customer reviews. *Journal of Marketing Research* 48(5):881–894.
- Liu L, Dzyabura D, Mizik N (2020) Visual listening in: Extracting brand image portrayed on social media. *Marketing Science* 39(4):669–686.
- Locatello F, Bauer S, Lučić M, Rätsch G, Gelly S, Schölkopf B, Bachem OF (2019) Challenging common assumptions in the unsupervised learning of disentangled representations. *International Conference on Machine Learning*, 4114–4124.
- Locatello F, Tschannen M, Bauer S, Rätsch G, Schölkopf B, Bachem O (2020) Disentangling factors of variations using few labels. *International Conference on Learning Representations*.
- Maas AL, Hannun AY, Ng AY, et al. (2013) Rectifier nonlinearities improve neural network acoustic models. *Proc. icml*, volume 30, 3 (Atlanta, GA).
- Murphy KP (2012) *Machine learning: a probabilistic perspective* (MIT press).
- Netzer O, Feldman R, Goldenberg J, Fresko M (2012) Mine your own business: Market-structure surveillance through text mining. *Marketing Science* 31(3):521–543.
- Rao VR, Sabavala DJ, et al. (1986) Measurement and use of market response functions for allocating marketing resources. (*No Title*) .
- Ringel DM, Skiera B (2016) Visualizing asymmetric competition among more than 1,000 products using big search data. *Marketing Science* 35(3):511–534.
- Sisodia A, Burnap A, Kumar V (2023) Generative interpretable visual design: Using disentanglement for visual conjoint analysis. *Available at SSRN 4151019* .
- Snoek J, Larochelle H, Adams RP (2012) Practical bayesian optimization of machine learning algorithms. *Advances in neural information processing systems* 25.

- Tirunillai S, Tellis GJ (2014) Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent dirichlet allocation. *Journal of marketing research* 51(4):463–479.
- Urban GL, Johnson PL, Hauser JR (1984) Testing competitive market structures. *Marketing Science* 3(2):83–112.
- Yang Y, Zhang K, Kannan P (2022) Identifying market structure: A deep network representation learning of social engagement. *Journal of Marketing* 86(4):37–56.

## Electronic Companion Supplement

### Appendix A: Neural Net Architecture

Figure EC.1 shows the detailed neural net architecture of our model. We modify the architecture proposed by Burgess et al. (2017) to accommodate  $128 \times 128$  pixel images and incorporate the supervisory signals from our model of market equilibrium.

The encoder neural network uses a sequence of convolutional neural network (CNN) layers to learn high-level representations of the input images. CNNs are well-suited for working with image data, as they can effectively capture spatial hierarchies and learn translation-invariant features (LeCun et al. 1995, Krizhevsky et al. 2012). We stack multiple CNN layers in the encoder to progressively learn more complex and abstract visual concepts. The output of the final CNN layer is then flattened and passed through two fully-connected (FC) layers. The first FC layer reduces the dimensionality of the flattened representation, while the second FC layer further compresses the information into a compact set of latent visual characteristics, with a maximum of  $J$  dimensions.

The decoder neural network is designed to reconstruct the original image from the latent visual characteristics. Its architecture is essentially the transpose of the encoder network, consisting of FC layers followed by a sequence of transposed convolutional layers (Dumoulin and Visin 2016). The decoder takes the  $J$ -dimensional latent visual characteristics as input and gradually upsamples and expands the representation until it reaches the original image size of  $128 \times 128$  pixels. Finally, the supervised neural network takes the discovered visual characteristics as input and predicts the vector of supervisory signals, which serve as labels for training the model. The supervised network allows the model to learn visual characteristics that are predictive of the supervisory signals, guiding the disentanglement process.

### Appendix B: UDR Algorithm

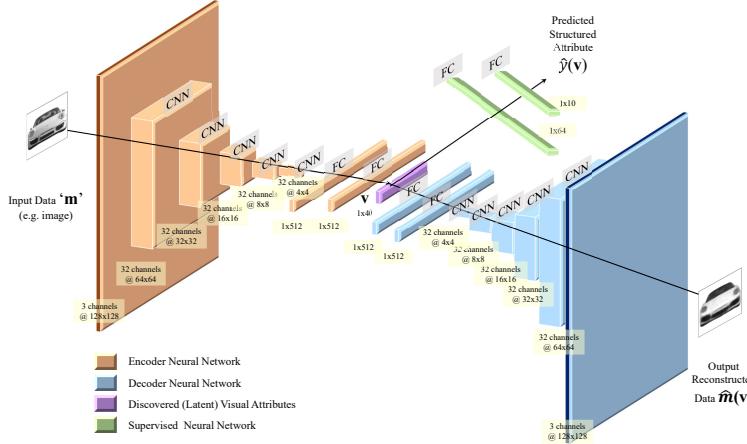
The Unsupervised Disentanglement Ranking (UDR) metric, proposed by Duan et al. (2020), assesses the similarity of disentangled representations learned by different models. The key idea behind UDR is that if two models have learned similar disentangled representations, their informative latent dimensions should exhibit strong correlations.

To compute UDR for a pair of models  $i$  and  $j$ , we first calculate the correlation matrix  $R$ , where each entry  $R(a, b)$  represents the correlation between latent dimensions  $a$  and  $b$  from models  $i$  and  $j$ , respectively (Equation EC.1). We then identify the most correlated latent dimension in model  $j$  for each dimension  $a$  in model  $i$ , denoted as  $r_a$  (Equation EC.2).

$$R(a, b) = \text{cor}(v_i(a), v_j(b)) \quad (\text{EC.1})$$

$$r_a = \max_{b \in V(j)} \text{corV}(a, b) \quad (\text{EC.2})$$

The UDR score for the pair of models,  $UDR_{ij}$ , is computed using Equation EC.3. This equation consists of two symmetric terms, each focusing on one model. The first term considers each informative latent dimension

**Figure EC.1** Model Architecture

Notes: The encoder neural net for the VAEs consisted of 5 convolutional layers, each with 32 channels,  $4 \times 4$  kernels, and a stride of 2. This was followed by 2 fully connected layers, each of 512 units. The latent distribution consisted of one fully connected layer of 40 units parameterizing the mean and log standard deviation of 20 Gaussian random variables. The decoder neural net architecture was the transpose of the encoder neural net but with the output parameterizing Bernoulli distributions over the pixels. Leaky ReLU activations were used throughout, which help alleviate the vanishing gradient problem and improve the model's ability to learn complex representations (Maas et al. 2013). We used the Adam optimizer (Kingma and Ba 2014) with the learning rate 5e-4 and parameters  $\beta_1 = 0.9$  and  $\beta_2 = 0.999$ . We set batch size equal to 64. We train for 200 epochs to ensure convergence.

$b$  in model  $j$  and calculates the ratio of the squared correlation of its most similar dimension in model  $i$  ( $r_b^2$ ) to the sum of correlations between  $b$  and all dimensions in model  $i$ . This ratio is then multiplied by an indicator function  $I_{KL}(b)$ , which equals 1 if dimension  $b$  is informative (determined by its KL divergence from the prior distribution) and 0 otherwise. The second term follows the same process, but with the roles of models  $i$  and  $j$  reversed.

$$UDR_{ij} = \frac{1}{d_i + d_j} \left[ \sum_{b \in Z(j)} \frac{r_b^2}{\sum_{a \in Z(i)} R(a, b)} I_{KL}(b) + \sum_{a \in Z(i)} \frac{r_a^2}{\sum_{b \in Z(j)} R(a, b)} I_{KL}(a) \right] \quad (\text{EC.3})$$

The final UDR score is normalized by the total number of informative dimensions in both models ( $d_i + d_j$ ) to ensure that having more informative dimensions does not automatically lead to a higher score. A perfect one-to-one correspondence between the informative dimensions of the two models would result in a UDR score of 1.

## Appendix C: Model of Market Equilibrium

We use the BLP demand model (Berry et al. 1995) with the specification laid out in Berry et al. (1999) to estimate a model of market equilibrium. This model allows us to obtain product-level fixed effects, which serve as supervisory signals for our disentanglement learning approach. Table EC.1 presents the notation used in the demand model.

**Table EC.1 Table of Notation in Demand Model**

Symbol	Meaning
$j$	Products
$t$	Markets
$i$	Consumers
$f$	Firms
$T$	Number of Markets
$J_t$	Number of products in market $t$
$I_t$	Number of consumers in market $t$
$F_t$	Number of firms in market $t$
$\zeta$	Model Parameters
$\zeta_1$	Linear demand-side parameters
$\zeta_2$	Non-linear common parameters
$\zeta_3$	Linear supply-side parameters
$p_{jt}$	Price
$c_{jt}$	Marginal Cost
$x_{jt}$	Observed product characteristic
$U_{ujt}$	Indirect utility
$\delta_{jt}$	Mean utility
$\mu_{ijt}$	Heterogeneous utility
$\epsilon_{ijt}$	Idiosyncratic taste shock
$d_{ijt}$	Choice indicator
$s_{ijt}$	Choice probability
$s_{jt}$	Market share
$\xi_{jt}$	Demand-side structural error
$\xi_j$	Product-level fixed effects
$\omega_{jt}$	Supply-side structural error
$Z^D$	Demand Instruments
$Z^S$	Supply Instruments
$W$	Weighting matrix
$g$	Sample Moments

*Consumers:* In each market  $t = 1, \dots, T$ , there are  $J_t$  differentiated goods and  $I_t$  consumers. For each market, we observe average quantities, prices and product characteristics for all  $J_t$  products. The indirect utility of consumer  $i$  from purchasing product  $j$  in market  $t$  is a function of observed product characteristics  $\mathbf{x}_{jt}$ , unobserved product-market characteristics  $\xi_{jt}$ , price  $p_{jt}$ , consumer characteristics  $\nu_{it}$ .  $y_{it}$  is the income of the consumer  $i$  in market  $t$ ,  $\nu_{it}^k$  represents consumer  $i$ 's taste for characteristic  $k$  in market  $t$ , and finally,  $\epsilon_{ijt}$  denotes a mean-zero idiosyncratic taste shock. The indirect utility is specified in Equation (EC.4).

$$U_{ijt} = \mathbf{x}_{jt} \bar{\beta} - \alpha p_{jt} / y_{it} + \xi_{jt} + \sum_k (\sigma_\beta^k x_{jt}^k \nu_{it}^k) + \epsilon_{ijt} \quad (\text{EC.4})$$

Price  $p_{jt}$  is typically endogenous, and based on the unobserved product-market characteristics  $\xi_{jt}$ , and hence correlated with it. The unobserved product-market characteristics  $\xi_{jt}$  can reflect hard to quantify aspects of

the product such as quality or style. The unobserved product characteristics can be decomposed into product fixed effect  $\tilde{\xi}_j$  and rest of the unobserved product-market characteristics  $\Delta\xi_{jt}$ . This decomposition is written in Equation (EC.5).

$$\xi_{jt} = \Delta\xi_{jt} + \tilde{\xi}_j \quad (\text{EC.5})$$

This decomposition is important since the product fixed effects  $\tilde{\xi}_j$  is used as the supervisory signal for the disentanglement learning model. If the visual data at the model level is time varying at a lower frequency (e.g. every 5 years) than the sales data (every year), then it would be possible to model separate fixed effects at this lower frequency. We do not use such data in our estimation, hence model just product-level fixed effects. Each consumer  $i$  in market  $t$  has unit demand, and chooses from the set  $J_t = \{0, 1, \dots, J_t\}$ , including the outside good denoted by  $j = 0$ , which represents no purchase and is given by  $U_{i0t} = \epsilon_{i0t}$ . Consumers select the alternative (including outside good) with the highest utility:

$$d_{ijt} = \begin{cases} 1 & \text{if } U_{ijt} > U_{ilt} \text{ for all } l \neq j \\ 0 & \text{otherwise} \end{cases} \quad (\text{EC.6})$$

Note that as in BLP, we can decompose the indirect utility in Equation (EC.4) into a mean utility,  $\delta_{jt}$ , and a deviation from that mean,  $\mu_{ijt}$ .

$$\begin{aligned} \delta_{jt}(\mathbf{x}_{jt}, p_{jt}, \Delta\xi_{jt}, \tilde{\xi}_j; \zeta_1) &= \mathbf{x}_{jt}\bar{\beta} + \Delta\xi_{jt} + \tilde{\xi}_j \\ \mu_{ijt}(\mathbf{x}_{jt}, p_{jt}, \nu_{ijt}, y_i; \zeta_2) &= -\alpha p_{jt}/y_i + \sum_k (\sigma_\beta x_{jt}^k \nu_{it}^k) + \epsilon_{ijt} \end{aligned} \quad (\text{EC.7})$$

The parameter vector is denoted  $\zeta = (\zeta_1, \zeta_2)$ . The vector  $\zeta_1$  contain the linear parameters or the mean preference on  $\mathbf{x}_{jt}$ , i.e.  $\bar{\beta}$ . These preferences are common across all consumers. The vector  $\zeta_2$  contain the nonlinear parameters or the standard deviation from mean preference i.e.  $\sigma_\beta$  as well as the term on the price  $\alpha$ . These nonlinear parameters introduce heterogeneity in preferences over structured product characteristics.

Using the standard assumption that  $\epsilon_{ijt}$  are i.i.d. with the Type I extreme value distribution, the probability  $s_{ijt}$  that consumer  $i$  chooses product  $j$  in market  $t$  and aggregate product market shares are given by equation (EC.8) below.

$$s_{ijt} = \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{l \in J_t} \exp(\delta_{lt} + \mu_{ilt})} \quad \text{and} \quad s_{jt} = \int \frac{\exp(\delta_{jt} + \mu_{ijt})}{\sum_{l \in J_t} \exp(\delta_{lt} + \mu_{ilt})} dF_i \quad (\text{EC.8})$$

*Firms:* We assume that automobile firms, indexed by  $f$  and part of a set  $F_t$ , play a static, full information, simultaneous move pricing game each period. Firms choose the price levels of all their models (products) with the objective of maximizing overall profit. We specify a constant marginal cost  $c_{jt}$  for a product  $j$  in market  $t$ . The pricing first order condition for vehicle  $j$  is given by Equation (EC.9).

$$s_{jt} + \sum_{j \in J_t} (p_{jt} - c_{jt}) \frac{\partial s_{jt}}{\partial p_{jt}} = 0 \quad (\text{EC.9})$$

We parameterize the marginal costs as written below in Equation (EC.10).

$$c_{jt} = \mathbf{x}_{jt}\gamma_1 + \mathbf{w}_{jt}\gamma_2 + \omega_{jt} \quad (\text{EC.10})$$

where  $\mathbf{x}_{jt}$  are product characteristics,  $\mathbf{w}_{jt}$  are observable cost-shifters and  $\omega_{jt}$  are unobserved cost-shifters. We can estimate the marginal costs for each product when we solve the supply model jointly with the demand model.

*Instruments:* In this demand model, we assume that a consumer's utility depends up on the observed product characteristics as well as unobserved (to the researcher) product characteristics. Firms observe these unobserved product characteristics and set then set prices, which implies that price is endogenous and necessitates the use of instruments.

We use BLP instruments in our analysis, other possibilities are detailed in ??.

$$Z_{BLP} = \{1, x_{jt}, w_{jt}, \sum_{j \in J_t \setminus \{j\}} 1, \sum_{j \notin J_t} 1, \sum_{j \in J_t \setminus \{j\}} x_{jt}, \sum_{j \notin J_t} x_{jt}\} \quad (\text{EC.11})$$

With the addition of demand instruments  $Z_{jt}^D$ , we construct demand-side moment conditions of the form  $E[\tilde{\xi}_{jt} Z_{jt}^D] = 0$ . Similarly, we also construct supply-side moment conditions of the form  $E[\omega_{jt} Z_{jt}^S] = 0$  using supply instruments  $Z_{jt}^S$ .

*GMM Estimation:* We construct a GMM estimator using both supply-side and demand-side moment conditions.

$$g(\theta) = \begin{bmatrix} \frac{1}{N} \sum_{jt} E[\xi_{jt} Z_{jt}^D] \\ \frac{1}{N} \sum_{jt} E[\omega_{jt} Z_{jt}^S] \end{bmatrix} \quad (\text{EC.12})$$

We construct a nonlinear GMM estimator for  $\zeta$  with some weighting matrix  $W$  in Equation (EC.13). We solve this problem twice. First, we obtain a consistent estimate of  $W$  and then an efficient GMM estimator.

$$\hat{\theta} = \min_{\theta} g(\zeta)' W g(\zeta) \quad (\text{EC.13})$$

## Appendix D: Model of Market Equilibrium Estimates

In this section, we estimate a demand model to understand the consumer preferences on structured product characteristics as well as obtain the fixed effects for each make-model. The make-model fixed effects capture the time-invariant consumer preferences on each product's unobservable characteristics such as quality and visual style. To make our model simple, we only introduce heterogeneity on two terms: price and constant. We cluster the standard errors at the make-model level. From Table EC.2, we can see that all the estimates are precise by looking at the standard errors of our estimates. Our estimates are in line with economic intuition. We find that on average consumers prefer cars with more power, higher fuel efficiency as well as larger space. We also find that consumers with higher income have lower price sensitivity.

Change the title of this section.

**Table EC.2 Parameter Estimates of Model of Market Equilibrium**

	Variable	Parameter Estimate	Standard Errors
Means ( $\bar{\beta}'s$ )			
	HP/Weight	3.30	(0.89)
	MP $\mathcal{L}$	1.00	(0.18)
	Space	0.89	(0.41)
Standard Deviation ( $\sigma'_\beta s$ )			
Term on Price	(-p/y)	-5.50	(0.83)
Supply-Side Terms			
	Constant	3.50	(0.12)
	ln(HP/Weight)	0.65	(0.08)
	ln(MPG)	-0.61	(0.06)
	ln(Space)	2.10	(0.21)
	Trend	0.02	(0.004)

## Appendix E: Validation Surveys for Interpretability and Quantification

To validate the interpretability of the discovered visual characteristics, we conducted a survey with 93 respondents after removing those who failed attention checks. Respondents were shown a sequence of five images for each visual characteristic, where the characteristic varied from left to right while keeping all other characteristics fixed. They were asked to describe how the car changed the most across the sequence of images. Figure EC.2 presents an example of the open-ended question posed to the respondents.

We then used language models (ChatGPT4 and Claude 3 Opus) to summarize the main themes that respondents agreed upon for each visual characteristic.<sup>6</sup> Based on the LLM summaries, we label the visual characteristics as follows:

Maybe put in more LLMs and show where they agree and/ differ.

1. Body Shape: The LLM summary of survey respondents states that the "car appears to change in shape, particularly becoming narrower, less angular, and more rounded with each successive image."
2. Color: Automobiles scoring low on this characteristic are darker and vice-versa.
3. Grille Height: The LLM summary of survey respondents indicates that "grilles are become larger, darker, and more defined." Although this summary also mentions the windscreens becoming darker, which is entangled with the grille height, it captures the essence of the characteristic. We interpret this as automobiles scoring low on this characteristic have less prominent grilles, while those scoring high have more prominent, and larger grilles.
4. Boxiness: The LLM summary of survey respondents describes that the "car becomes lower, flatter, and wider as the sequence progresses." We interpret this as automobiles scoring low on this characteristic have a high degree of boxiness, characterized by a taller, more upright, and narrower shape. In contrast, those scoring high on this characteristic have a lower degree of boxiness, with a lower, flatter, and wider appearance.

<sup>6</sup> LLM Prompt: Summarize the below responses and share the biggest theme that most respondents agree up on?

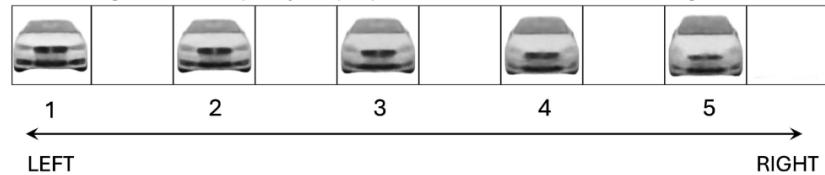
**Figure EC.2 Interpretability Validation Survey Question**

Q1/4: Look at the below image to see the various parts of a car.



Now, carefully examine each car image below from 1 to 5, going from left to right.

Note: Images are low-quality on purpose. Be sure to see all the images 1 to 5.



**How does the car change the most as you go from image 1 to 5?** Go through each part of the car one by one before deciding your response. Write it in a few words.

5. Grille Width: The LLM summary of survey respondents states that "grille width become smaller, narrower, and less pronounced as the sequence progresses." Based on this, we interpret that automobiles scoring low on this characteristic have a wider, more prominent grille, while those scoring high have a narrower, less pronounced grille.

To further validate the quantification of the visual characteristics determined by our method, we conducted a second survey (Figure EC.3). In this survey, we presented respondents with several pairs of automobile images that differed only along one visual characteristic. Respondents were asked to select the pair of automobiles that they perceived as more similar. We then compared the responses to our algorithm's quantification to assess consistency with human interpretation. For the characteristic we labeled as "Body Shape," 97% of the 104 respondents agreed with the algorithm's quantification scale. Similarly, for "Grille Height," 98% of the 107 respondents were in agreement. The characteristic we termed "Boxiness" had a 95% agreement rate among 103 respondents, while "Grille Width" saw 93% of the 104 respondents concurring with the algorithm's quantification. Overall, a strong majority of respondents (averaging 96% across four visual characteristics) agreed with the algorithm's quantification scale for the visual characteristics, demonstrating that our method's quantification aligns well with human perception.

**Figure EC.3 Quantification Validation Survey Question**

Which pair of cars in your judgment are visually more similar? Carefully check both large and small visual aspects. Do not consider any non-visual features like brand or price.



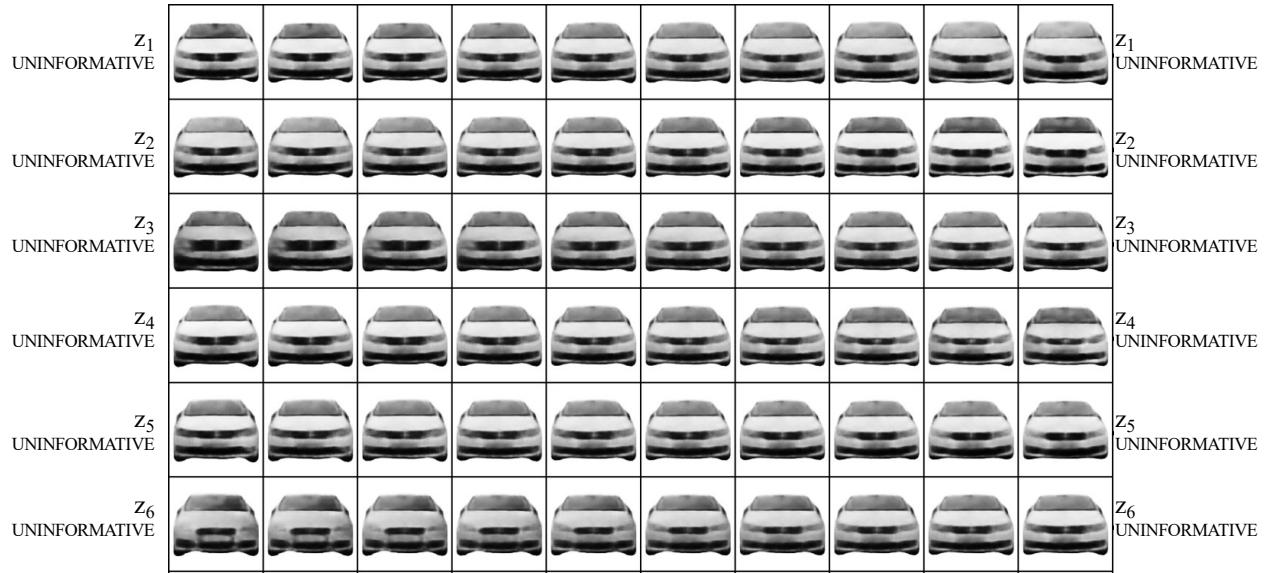
## Appendix F: Other Posterior Traversals

Figure EC.4 shows that the visual characteristics learned from the Autoencoder model are all uninformative. Similarly, Figure EC.5 shows that the visual characteristics learned from the unsupervised VAE model are all uninformative as well. Finally, Figure EC.6 shows the visual characteristics learned from the unsupervised  $\beta$ -TCVAE model. Unlike the supervised model, we can see that two of the characteristics learned from the unsupervised model are the same.

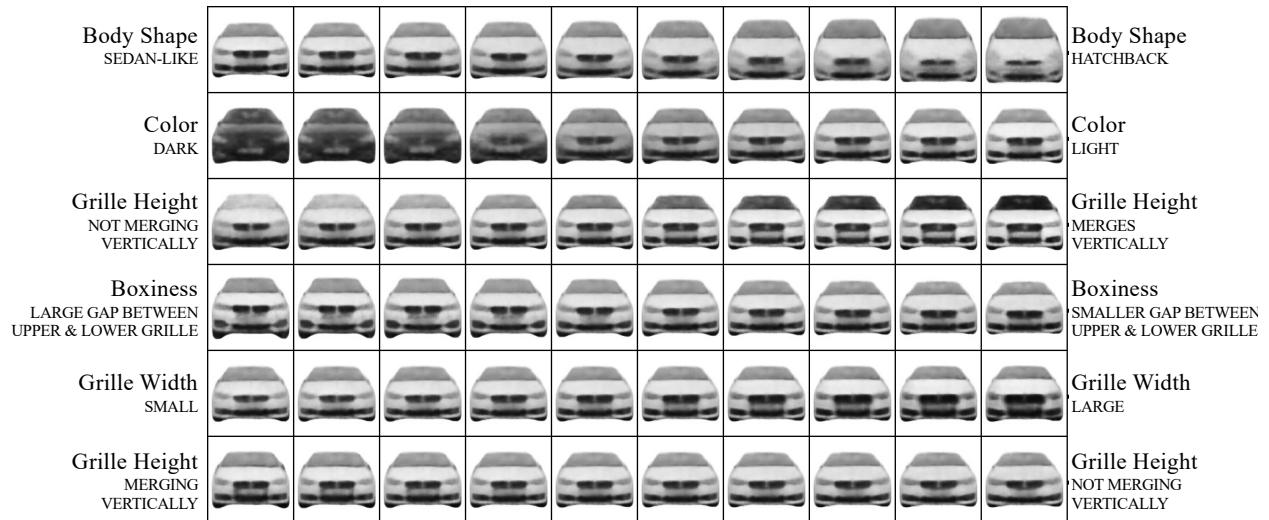
**Figure EC.4 Discovered Visual Characteristics from Autoencoder**

$z_1$ UNINFORMATIVE		$z_1$ UNINFORMATIVE
$z_2$ UNINFORMATIVE		$z_2$ UNINFORMATIVE
$z_3$ UNINFORMATIVE		$z_3$ UNINFORMATIVE
$z_4$ UNINFORMATIVE		$z_4$ UNINFORMATIVE
$z_5$ UNINFORMATIVE		$z_5$ UNINFORMATIVE
$z_6$ UNINFORMATIVE		$z_6$ UNINFORMATIVE

Left to Right: Vary one visual characteristic, keeping all others fixed

**Figure EC.5 Discovered Visual Characteristics from Unsupervised VAE**

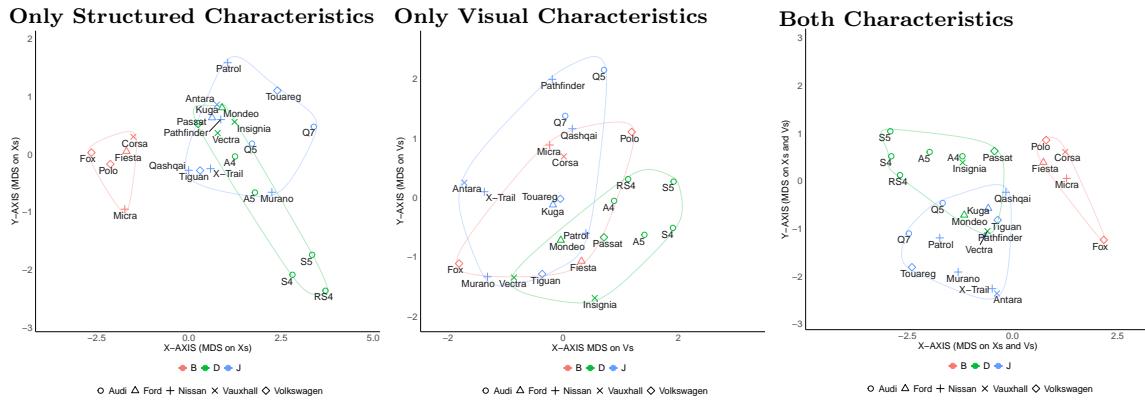
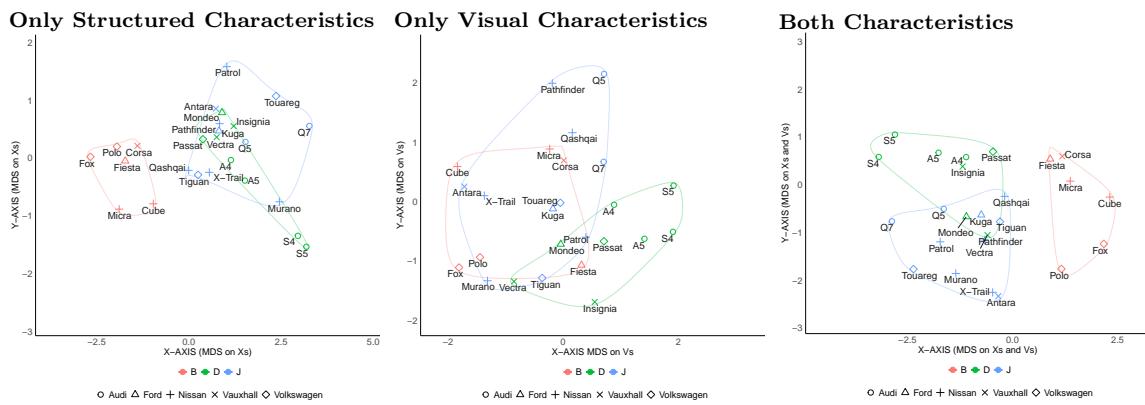
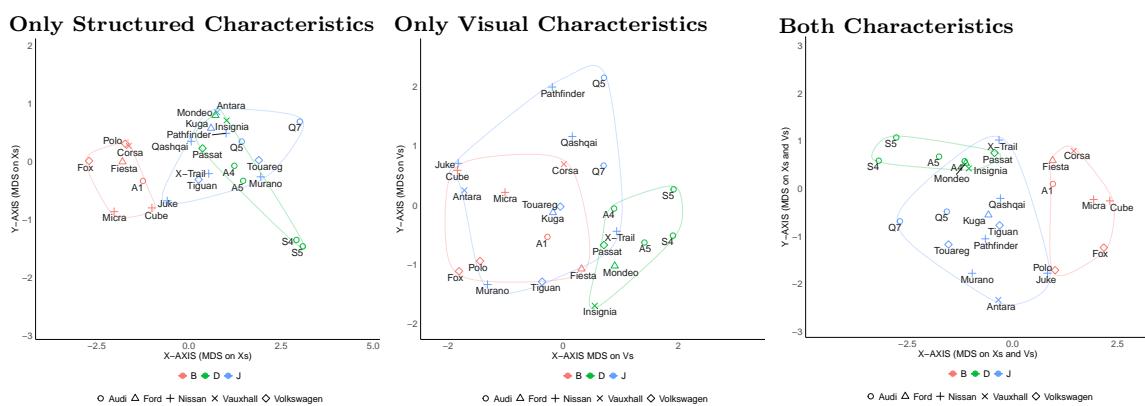
Left to Right: Vary one visual characteristic, keeping all others fixed

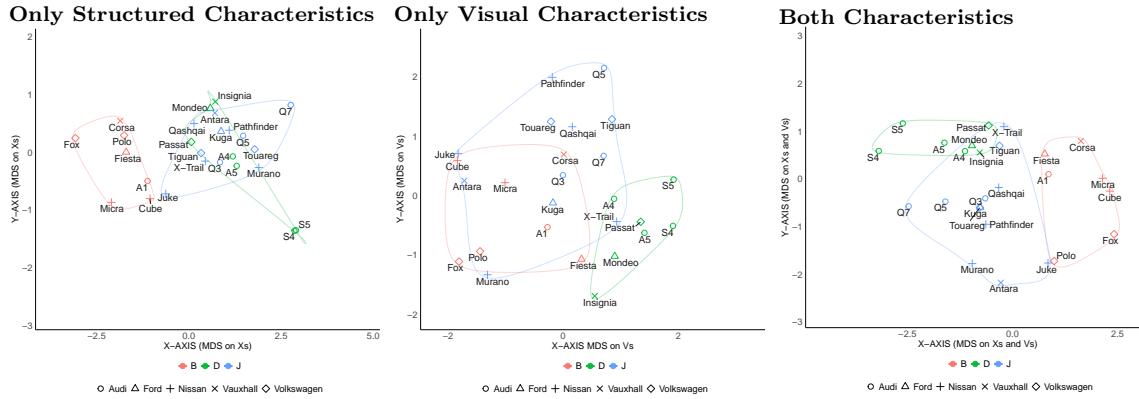
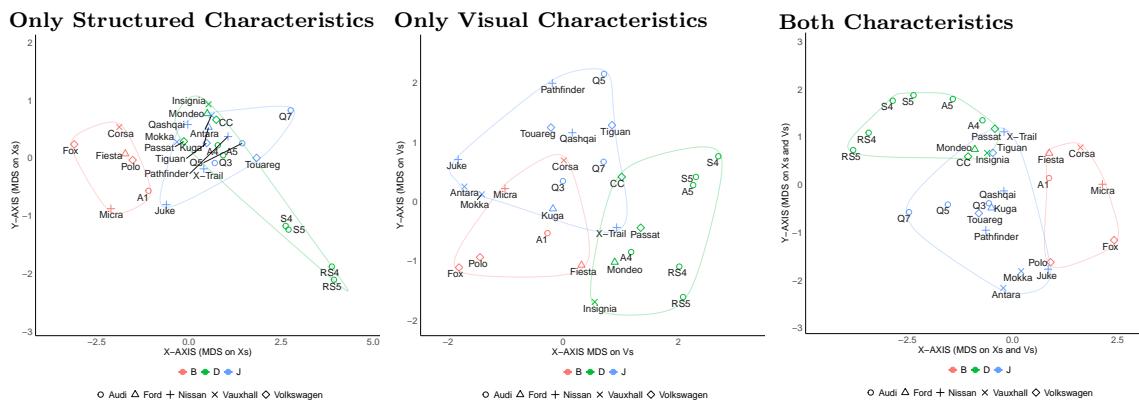
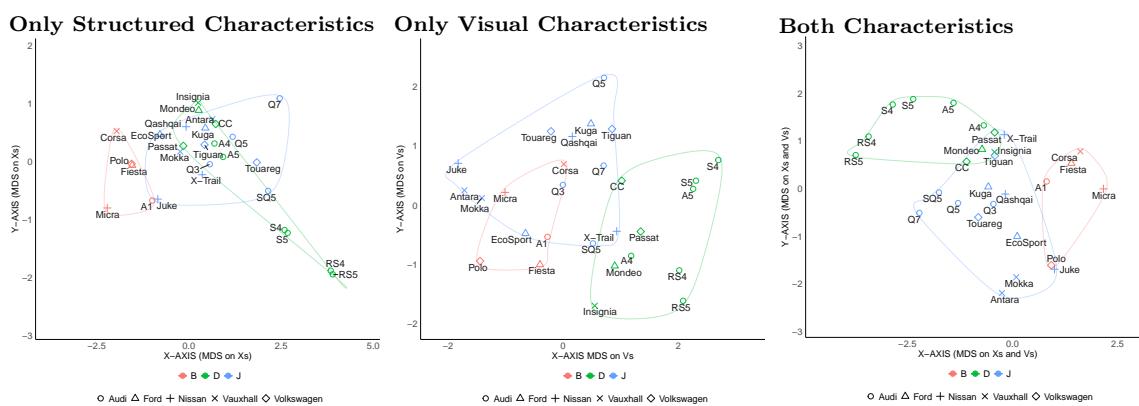
**Figure EC.6 Discovered Visual Characteristics from Unsupervised  $\beta$ -TCVAE**

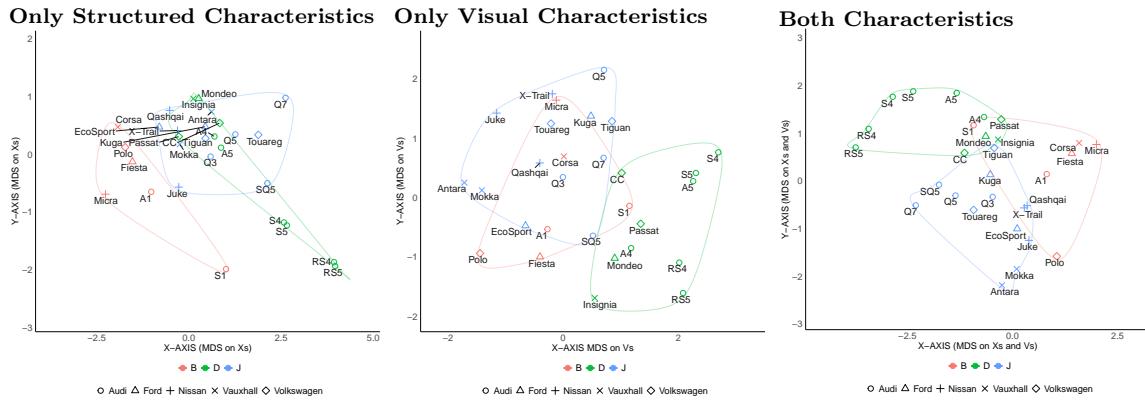
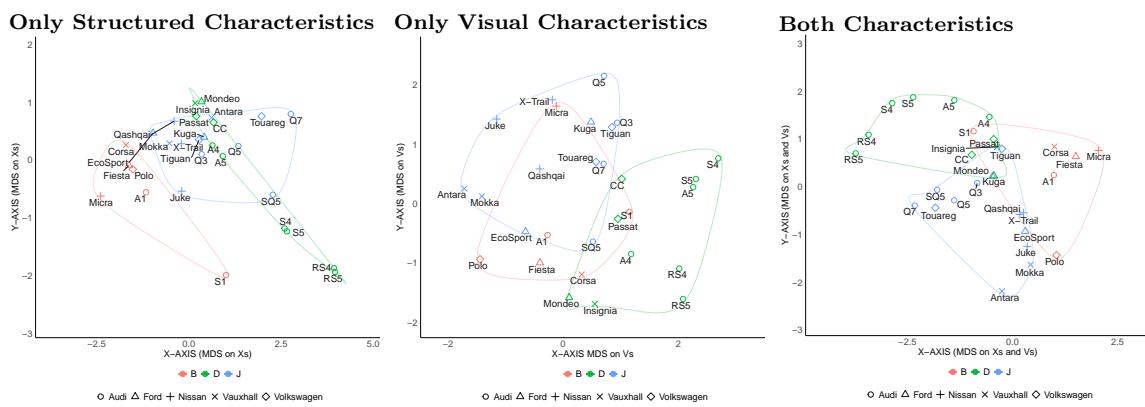
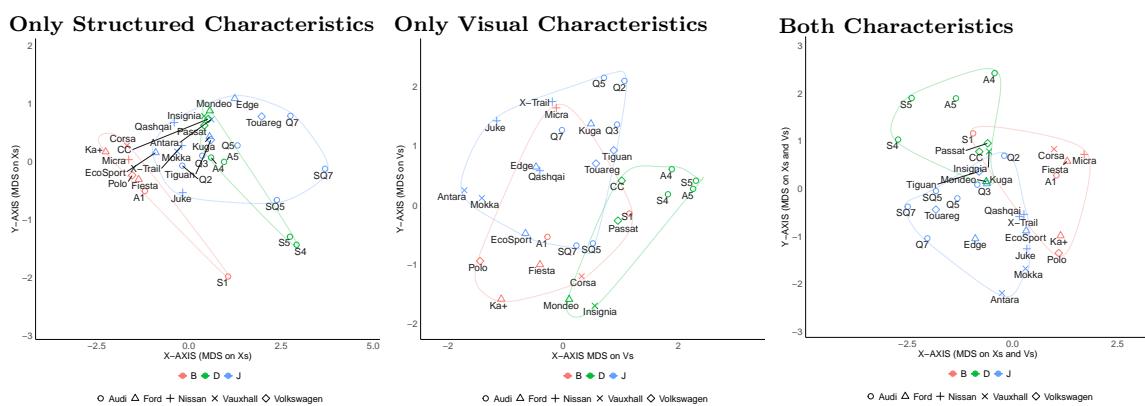
Left to Right: Vary one visual characteristic, keeping all others fixed

## Appendix G: Market Structure Maps in Different Years

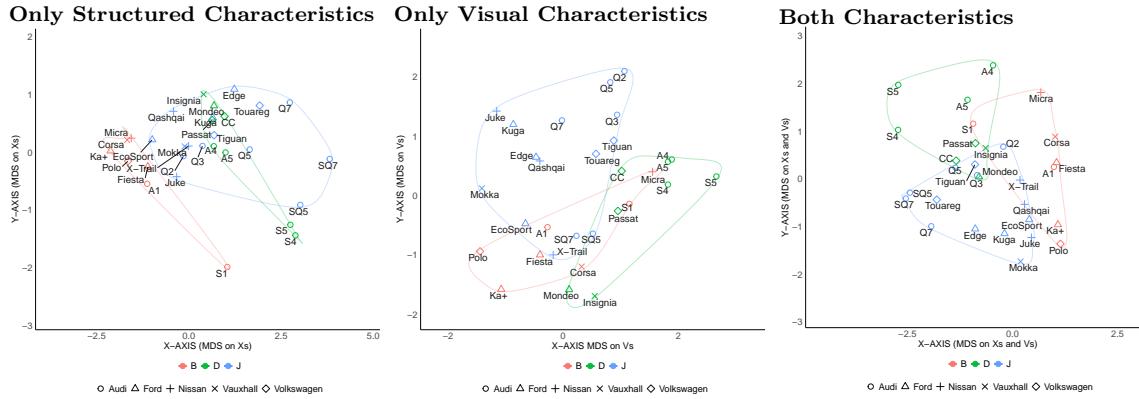
Figures EC.7 to EC.16 presents the market structure maps for different years from 2008 to 2017. These maps allow us to observe the evolution of the competitive landscape in the UK automobile market over time, based on both structured and visual product characteristics.

**Figure EC.7 (Color Online) Segment B, D & J: Market Structure Map for 2008****Figure EC.8 (Color Online) Segment B, D & J: Market Structure Map for 2009****Figure EC.9 (Color Online) Segment B, D & J: Market Structure Map for 2010**

**Figure EC.10 (Color Online) Segment B, D & J: Market Structure Map for 2011****Figure EC.11 (Color Online) Segment B, D & J: Market Structure Map for 2012****Figure EC.12 (Color Online) Segment B, D & J: Market Structure Map for 2013**

**Figure EC.13 (Color Online) Segment B, D & J: Market Structure Map for 2014****Figure EC.14 (Color Online) Segment B, D & J: Market Structure Map for 2015****Figure EC.15 (Color Online) Segment B, D & J: Market Structure Map for 2016**

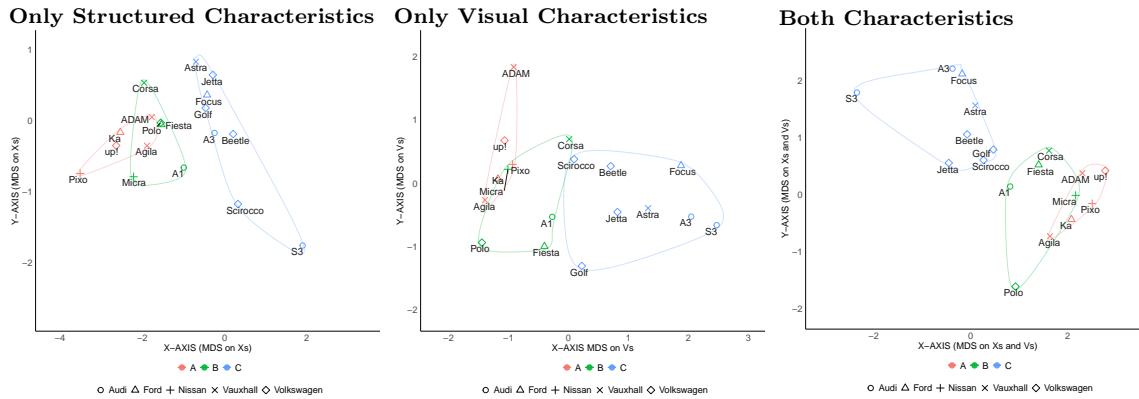
**Figure EC.16** (Color Online) Segment B, D & J: Market Structure Map for 2017

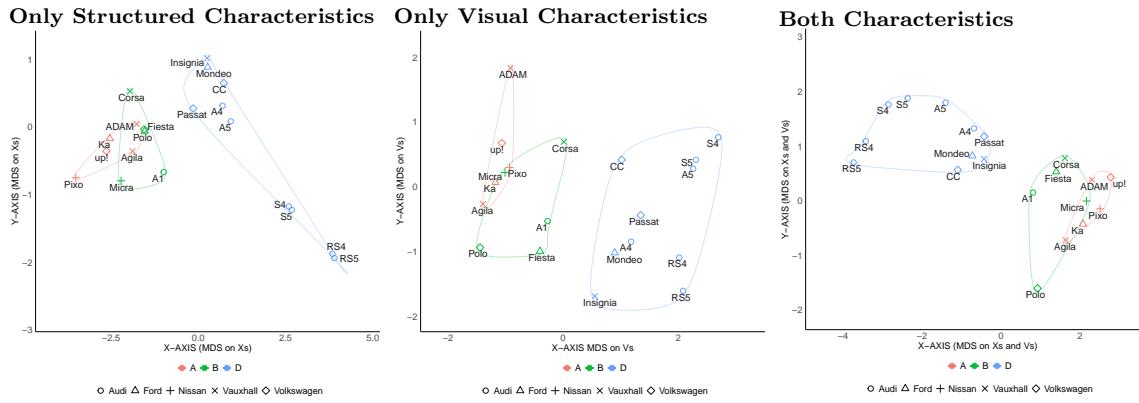
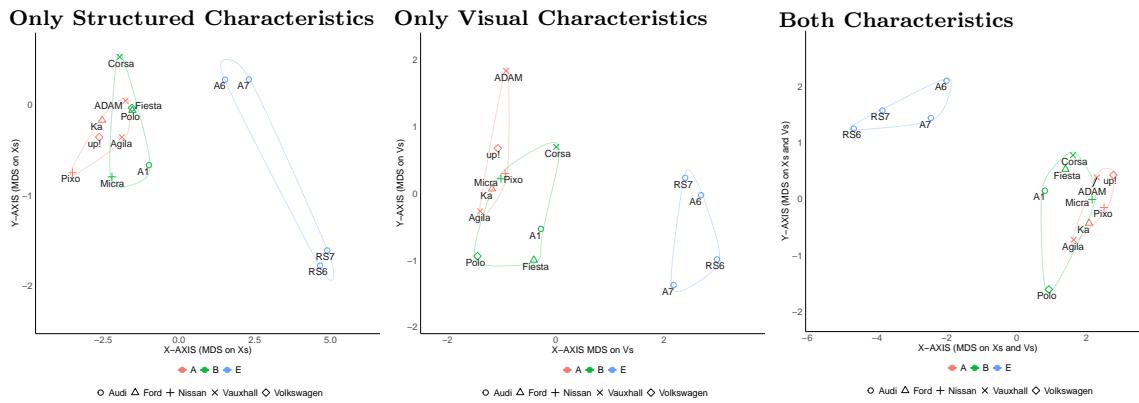
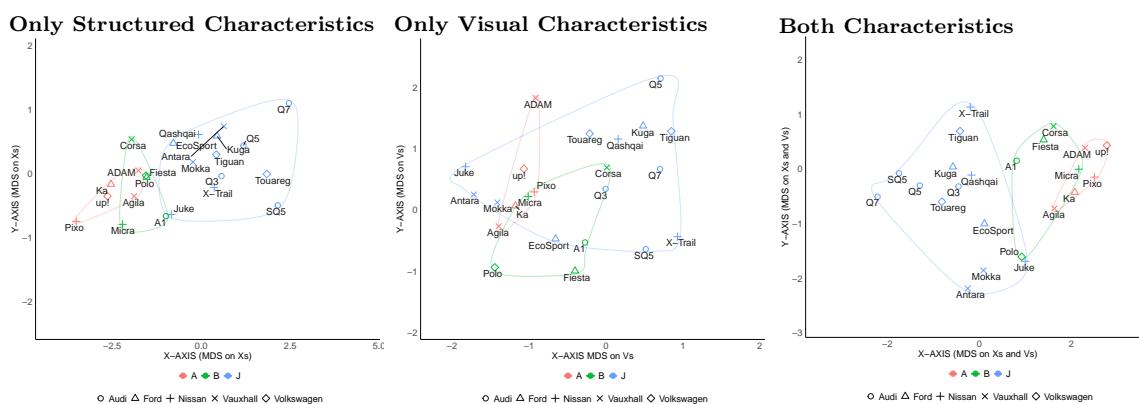


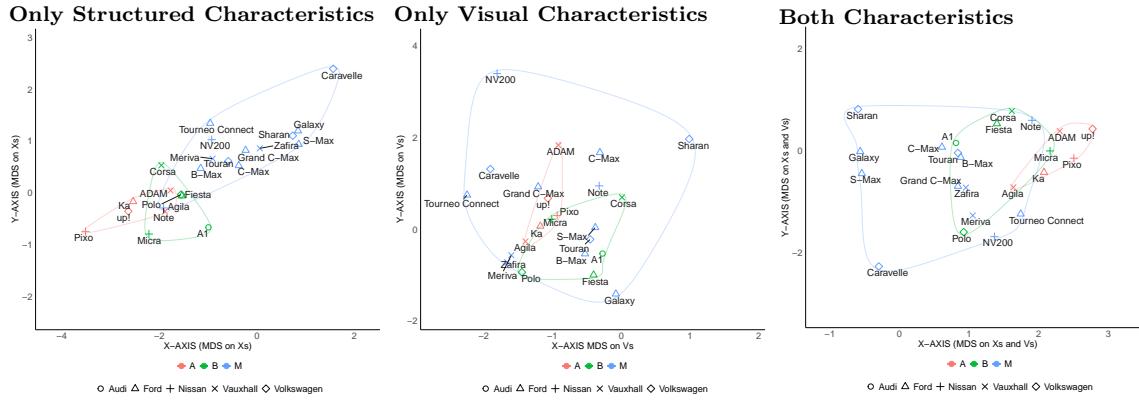
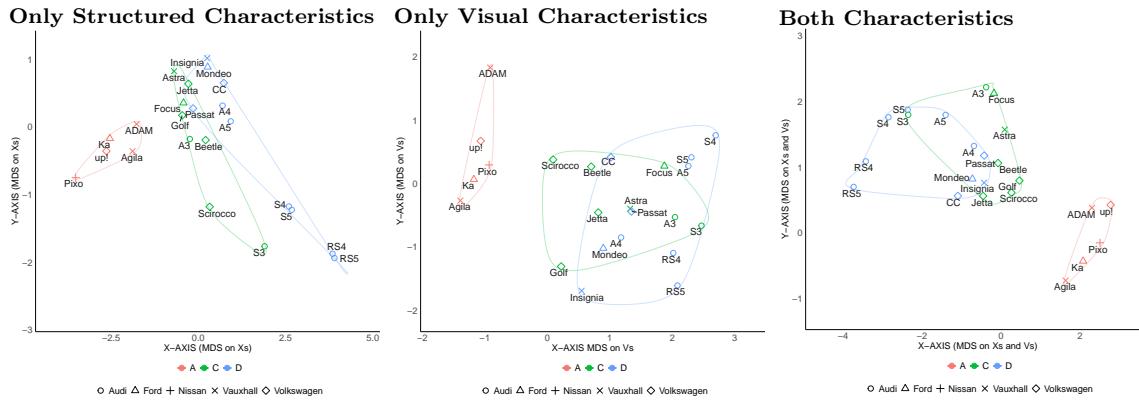
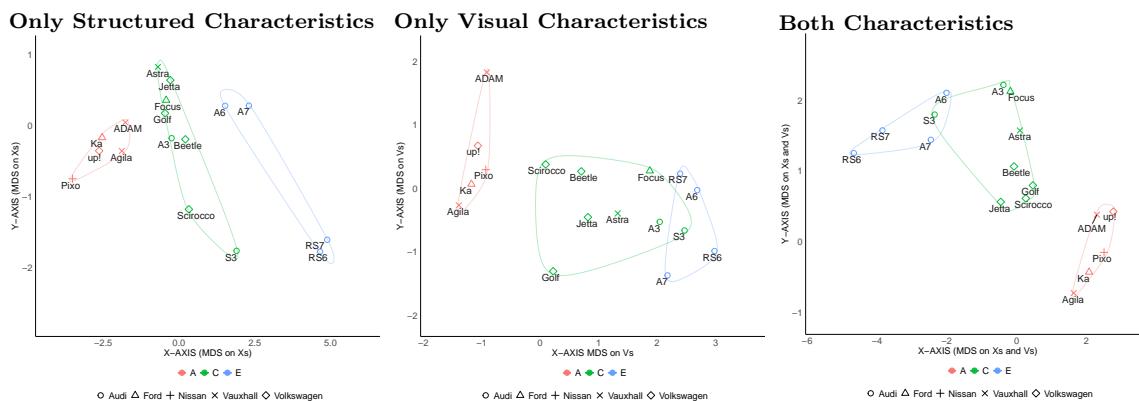
## **Appendix H: Market Structure Maps for Different Segment Combinations in 2013**

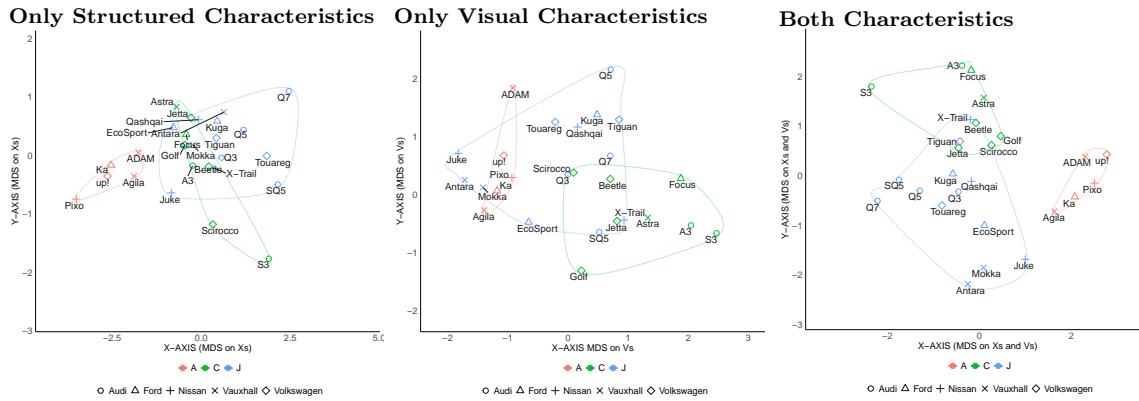
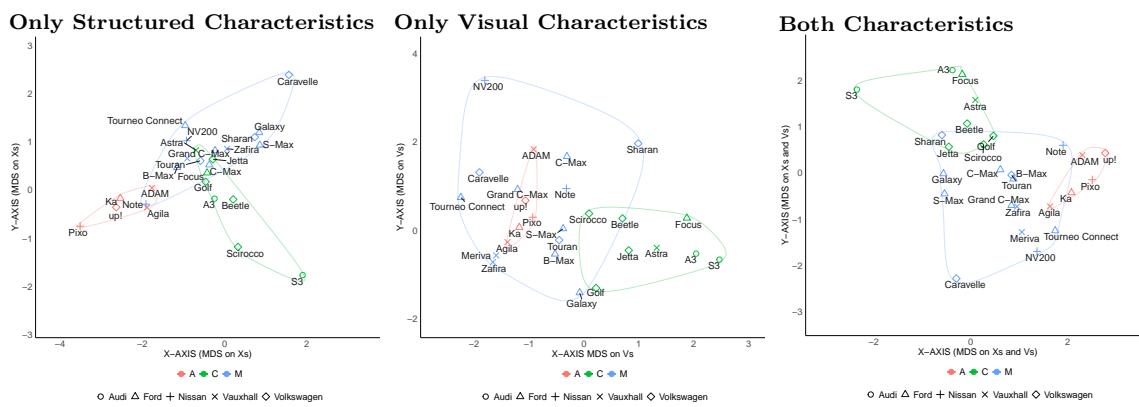
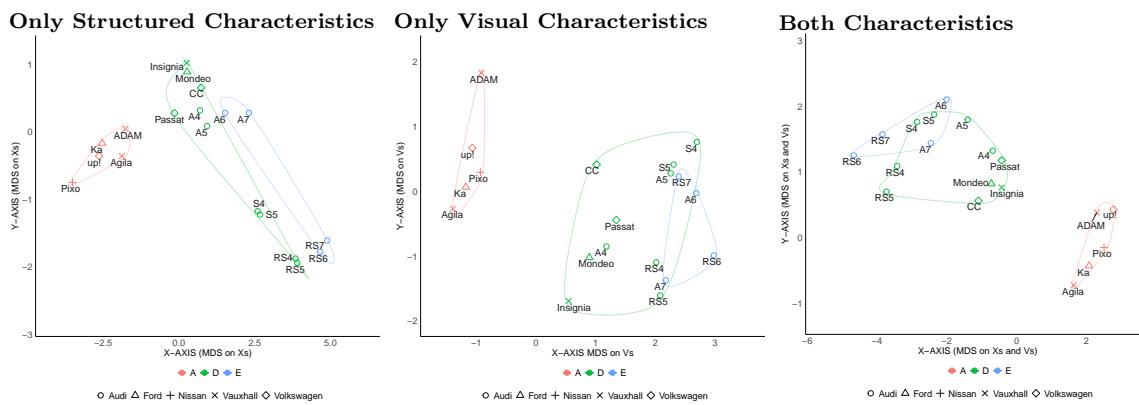
Figures EC.17 to EC.50 presents market structure maps for various combinations of automobile segments (e.g., A, B, C, D, E, J, and M) in the year 2013. By exploring different segment combinations, we can gain insights into the competitive dynamics across the entire market and identify potential cross-segment substitution patterns.

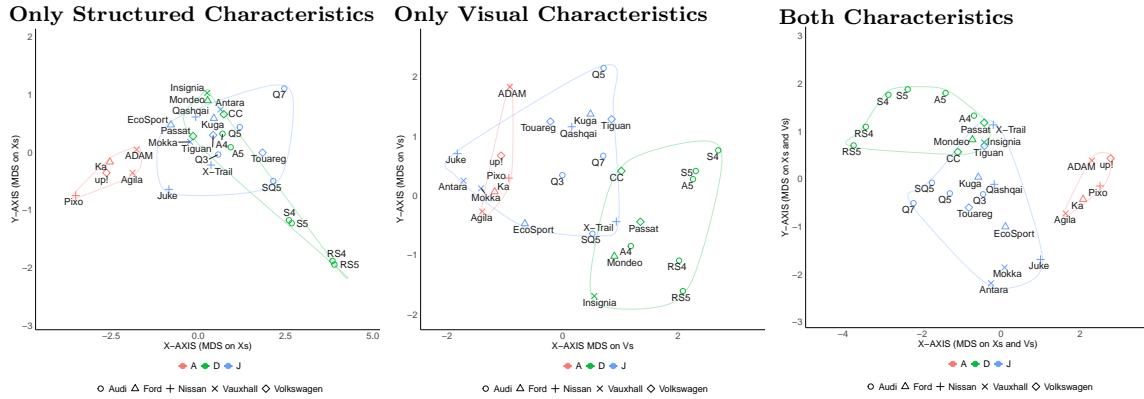
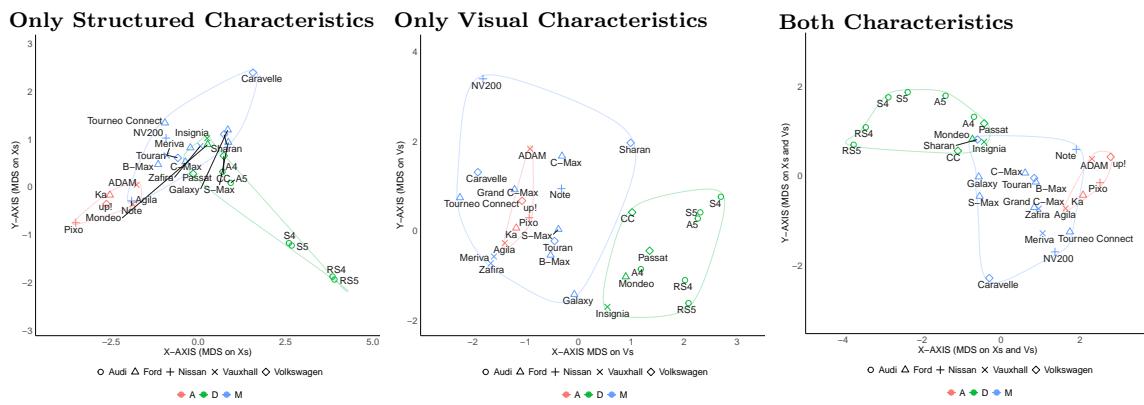
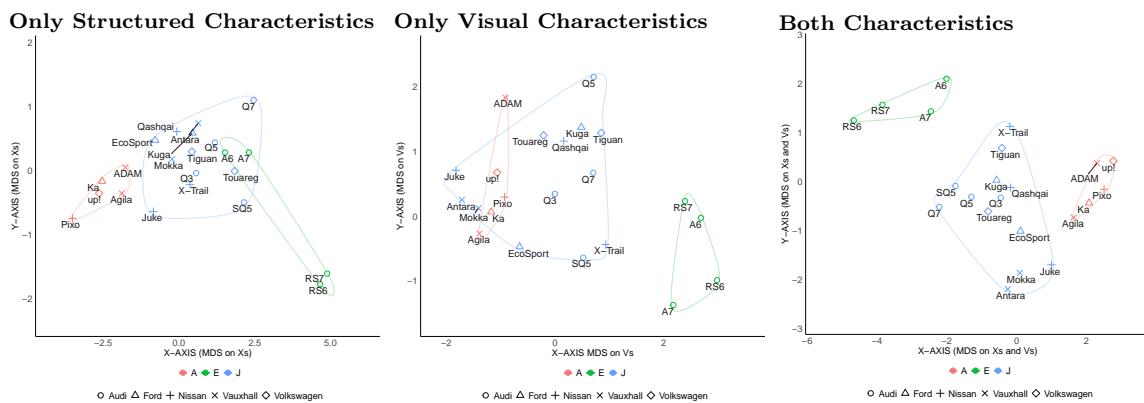
**Figure EC.17** (Color Online) Segment A, B & C: Market Structure Map for 2013

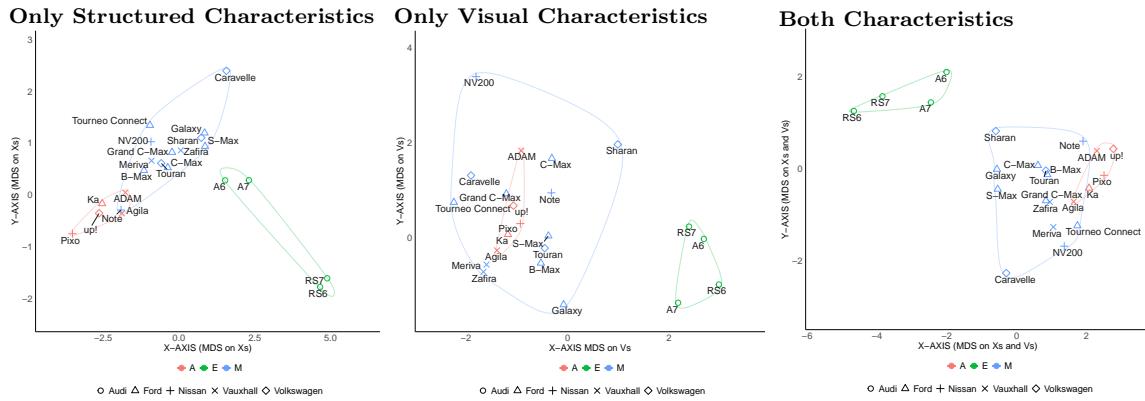
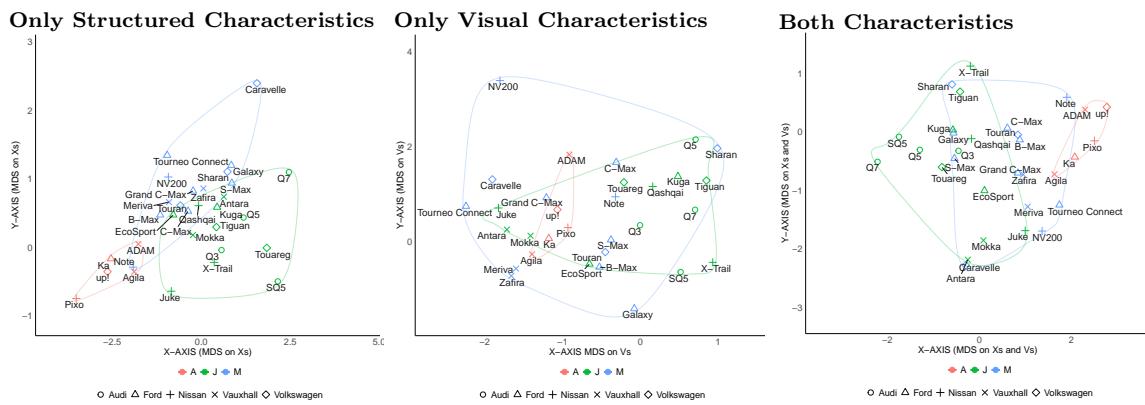
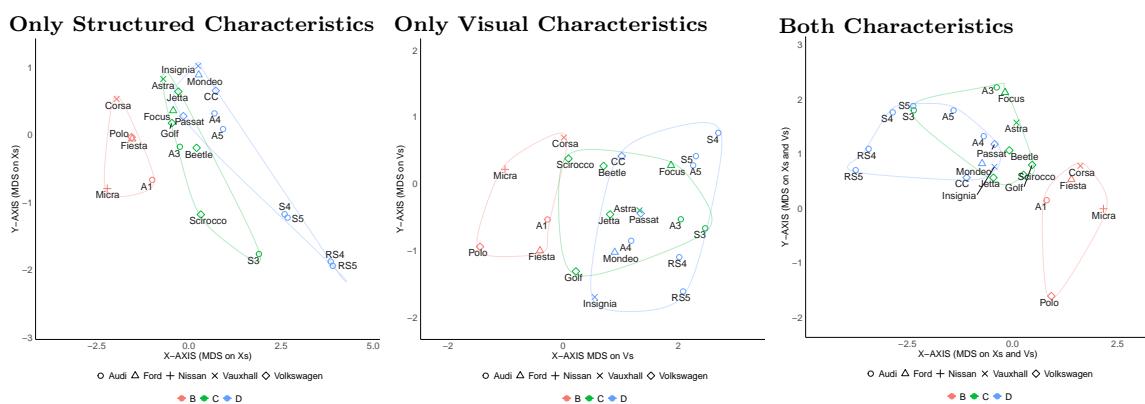


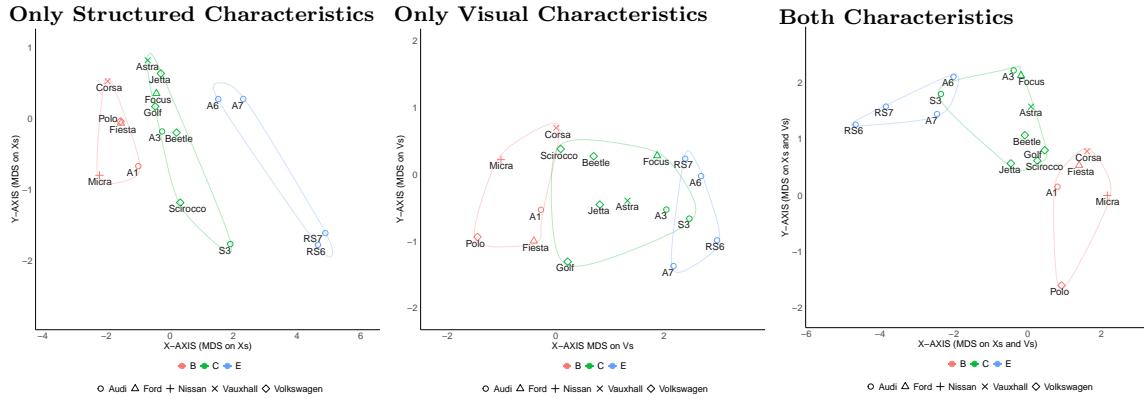
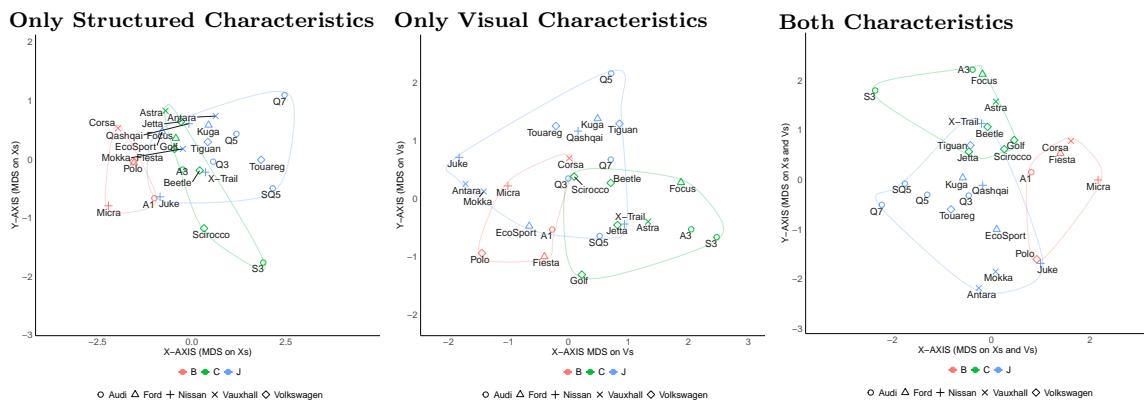
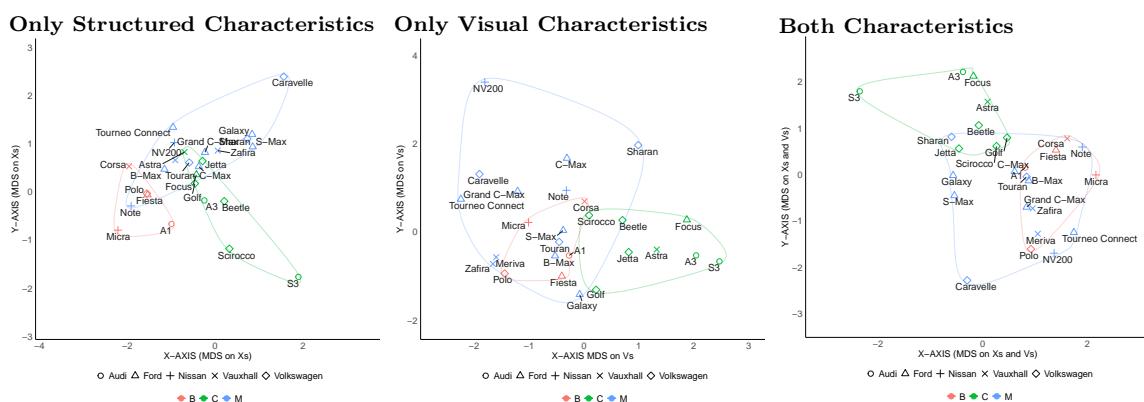
**Figure EC.18 (Color Online) Segment A, B, & D: Market Structure Map for 2013****Figure EC.19 (Color Online) Segment A, B & E: Market Structure Map for 2013****Figure EC.20 (Color Online) Segment A, B & J: Market Structure Map for 2013**

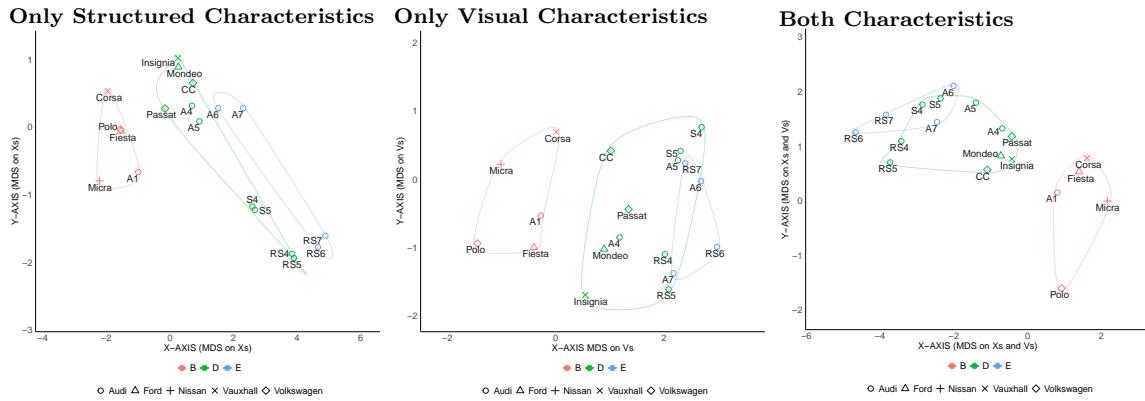
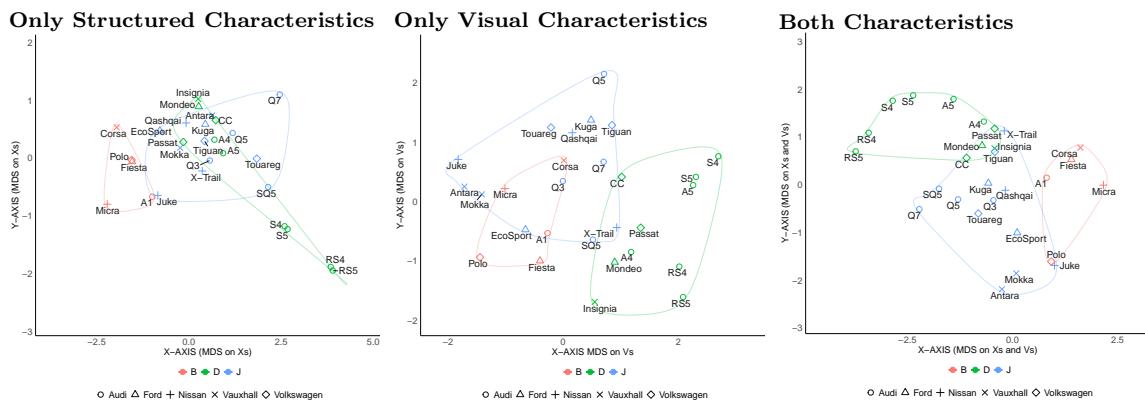
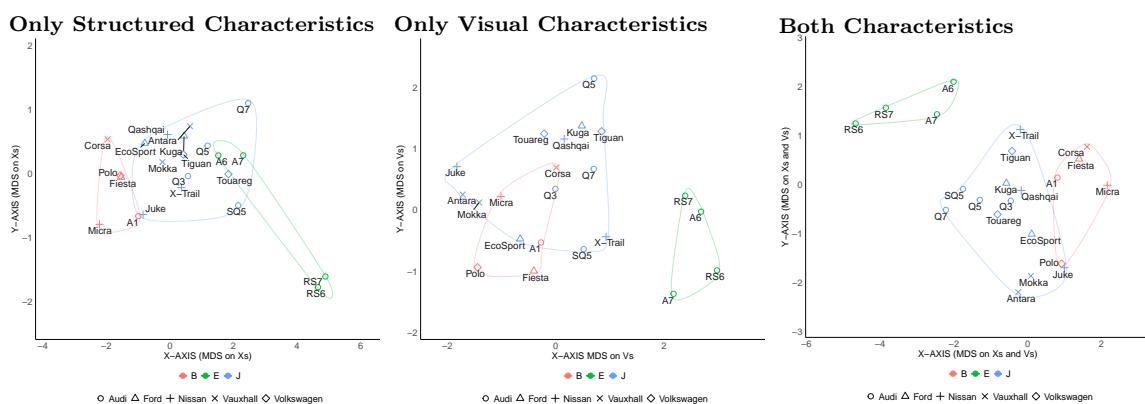
**Figure EC.21 (Color Online) Segment A, B & M: Market Structure Map for 2013****Figure EC.22 (Color Online) Segment A, C & D: Market Structure Map for 2013****Figure EC.23 (Color Online) Segment A, C & E: Market Structure Map for 2013**

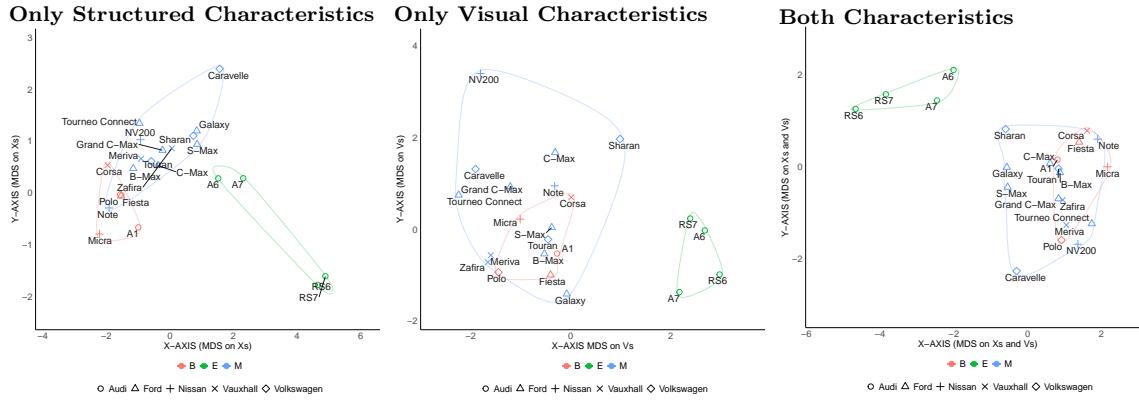
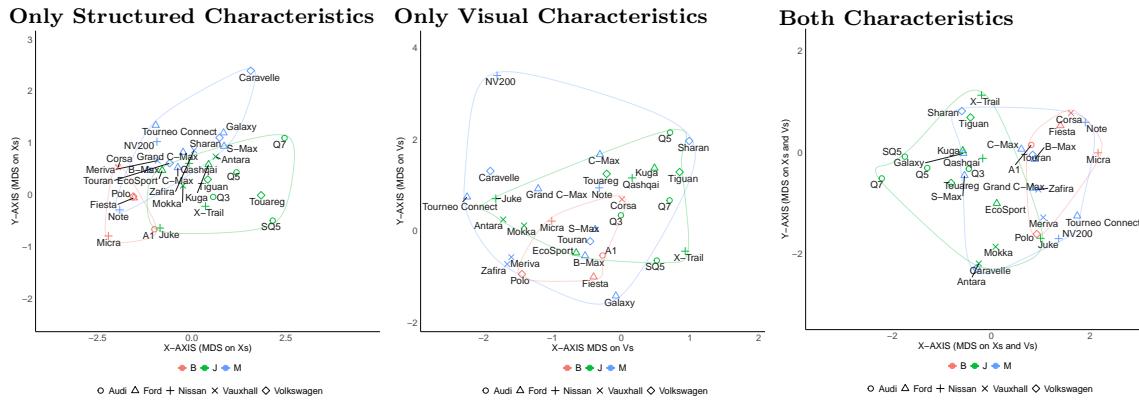
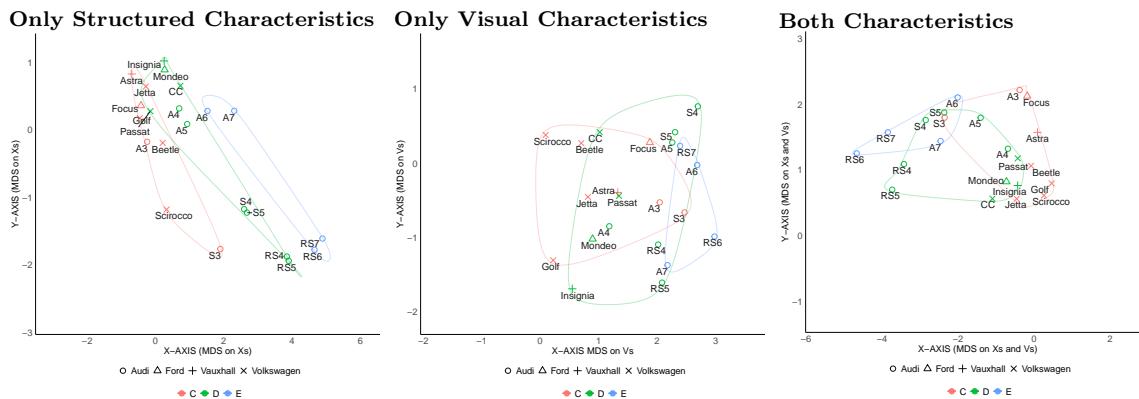
**Figure EC.24 (Color Online) Segment A, C & J: Market Structure Map for 2013****Figure EC.25 (Color Online) Segment A, C & M: Market Structure Map for 2013****Figure EC.26 (Color Online) Segment A, D & E: Market Structure Map for 2013**

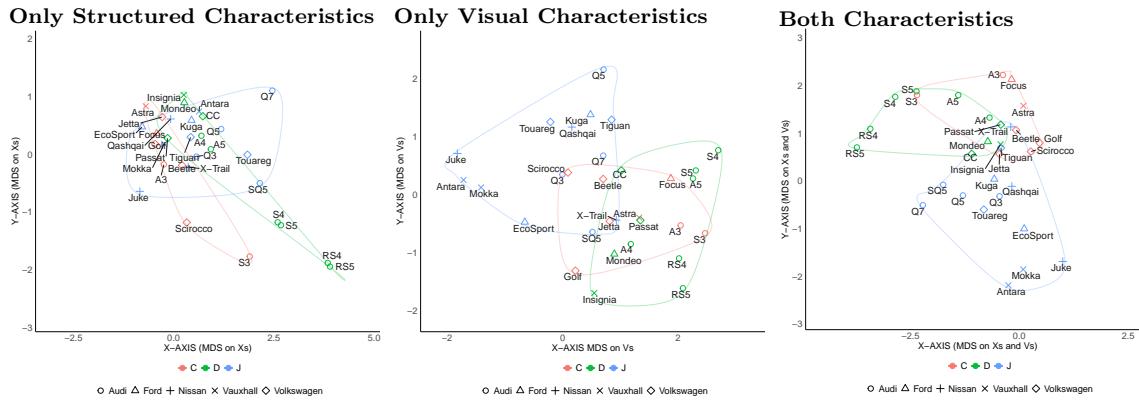
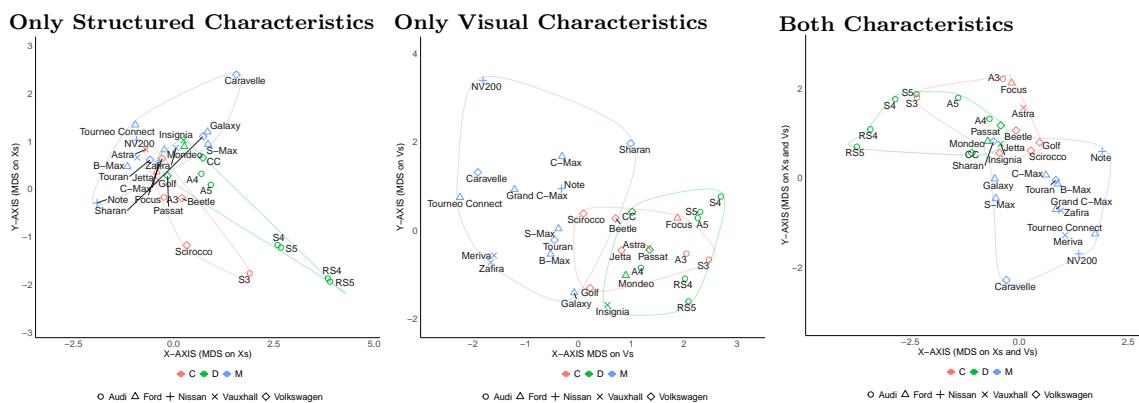
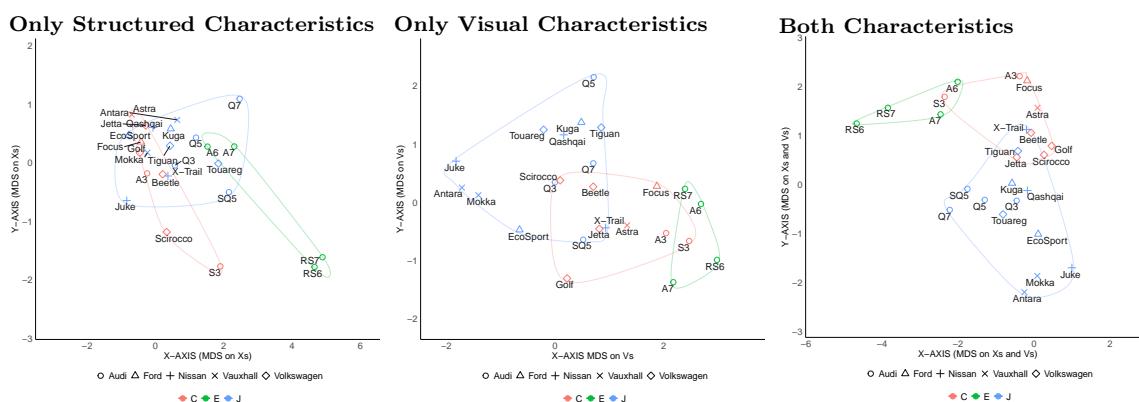
**Figure EC.27 (Color Online) Segment A, D & J: Market Structure Map for 2013****Figure EC.28 (Color Online) Segment A, D & M: Market Structure Map for 2013****Figure EC.29 (Color Online) Segment A, E & J: Market Structure Map for 2013**

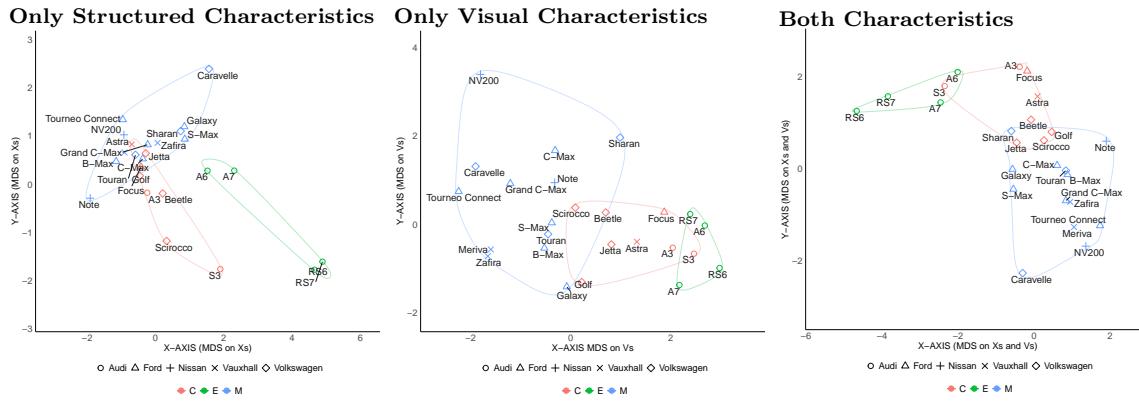
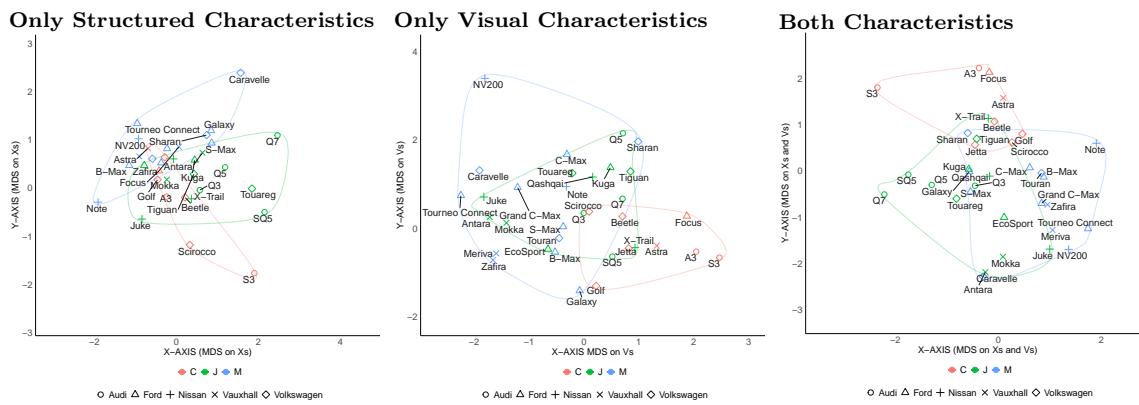
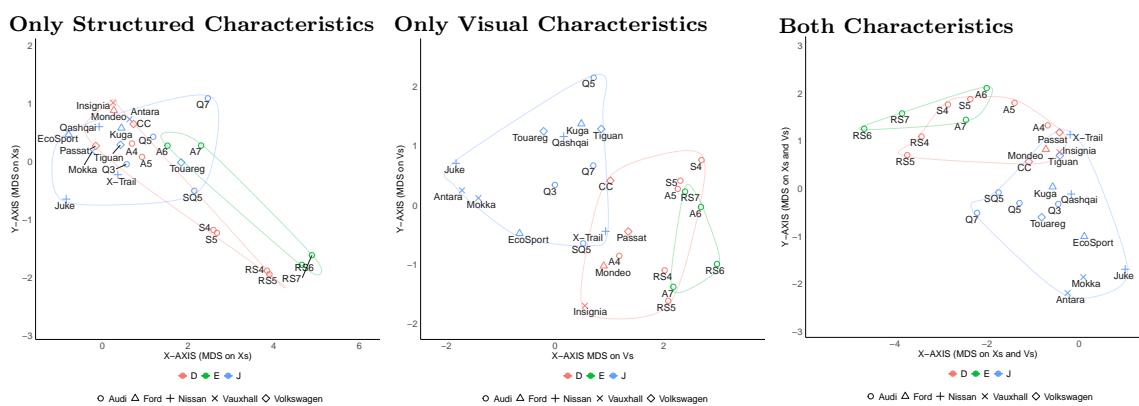
**Figure EC.30 (Color Online) Segment A, E & M: Market Structure Map for 2013****Figure EC.31 (Color Online) Segment A, J & M: Market Structure Map for 2013****Figure EC.32 (Color Online) Segment B, C & D: Market Structure Map for 2013**

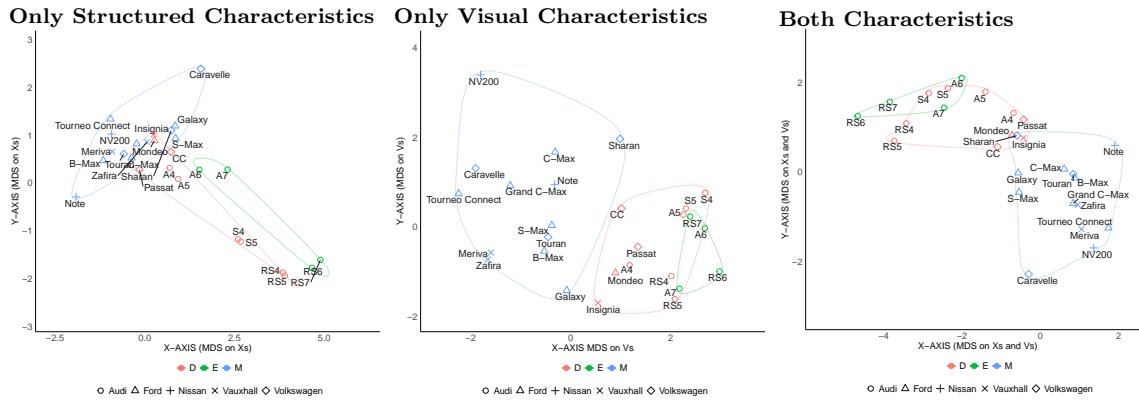
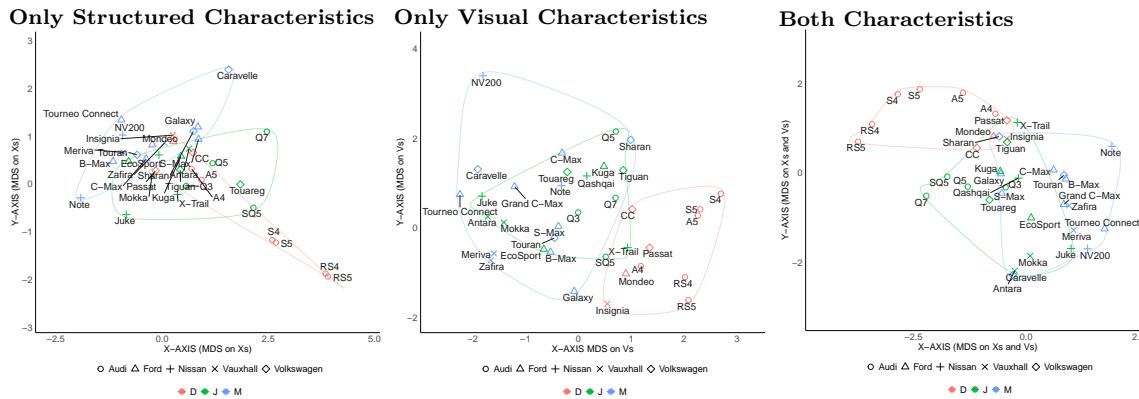
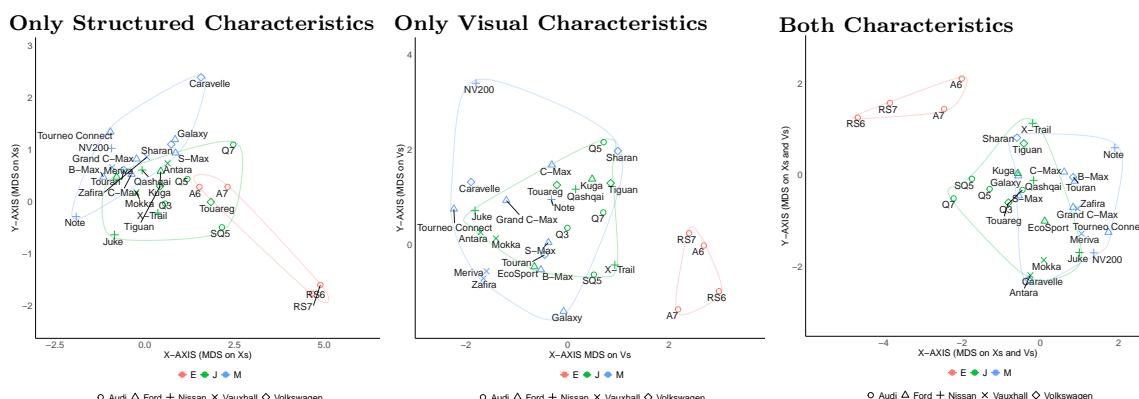
**Figure EC.33 (Color Online) Segment B, C & E: Market Structure Map for 2013****Figure EC.34 (Color Online) Segment B, C & J: Market Structure Map for 2013****Figure EC.35 (Color Online) Segment B, C & M: Market Structure Map for 2013**

**Figure EC.36 (Color Online) Segment B, D & E: Market Structure Map for 2013****Figure EC.37 (Color Online) Segment B, D & J: Market Structure Map for 2013****Figure EC.38 (Color Online) Segment B, E & J: Market Structure Map for 2013**

**Figure EC.39 (Color Online) Segment B, E & M: Market Structure Map for 2013****Figure EC.40 (Color Online) Segment B, J & M: Market Structure Map for 2013****Figure EC.41 (Color Online) Segment C, D & E: Market Structure Map for 2013**

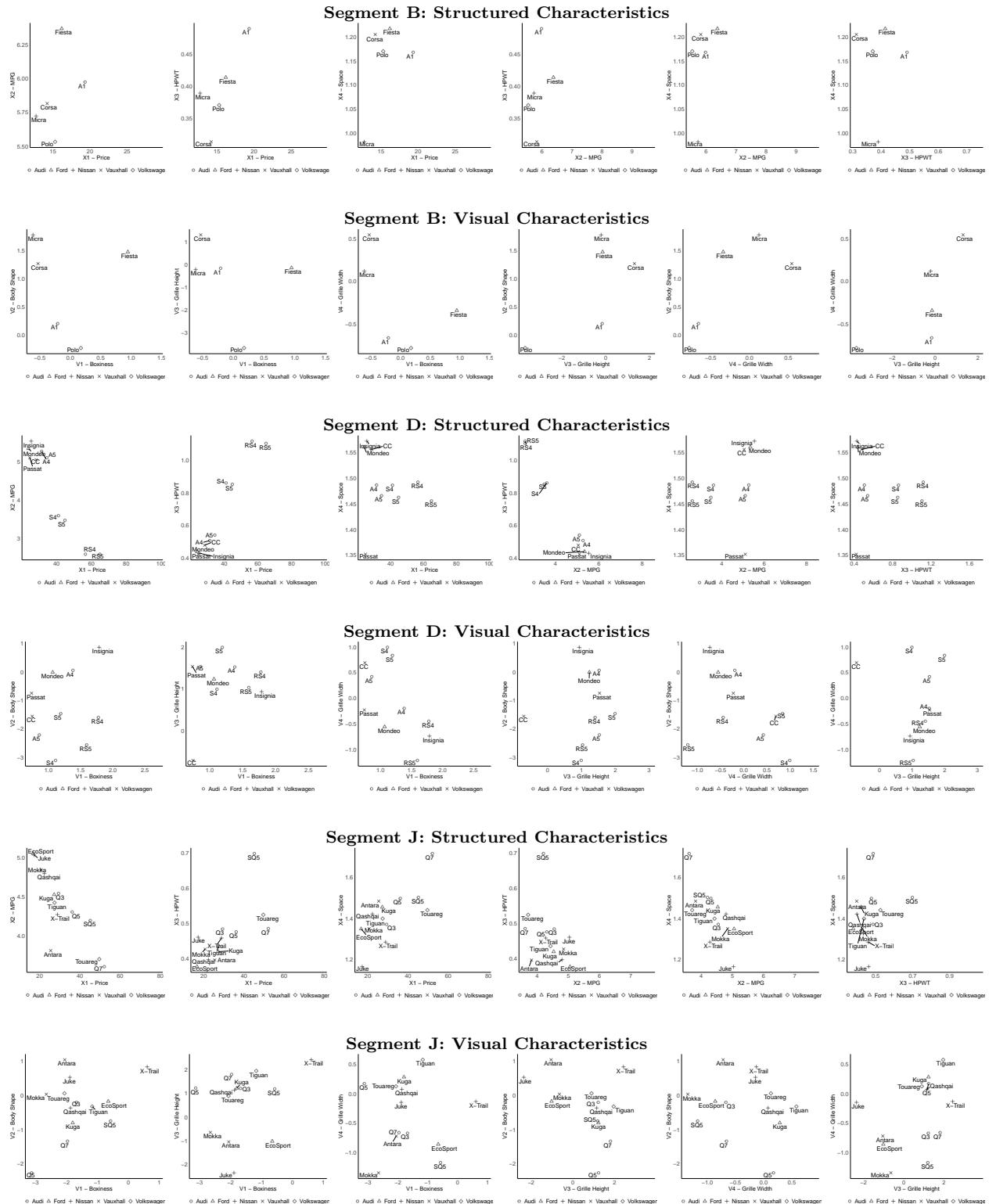
**Figure EC.42 (Color Online) Segment C, D & J: Market Structure Map for 2013****Figure EC.43 (Color Online) Segment C, D & M: Market Structure Map for 2013****Figure EC.44 (Color Online) Segment C, E & J: Market Structure Map for 2013**

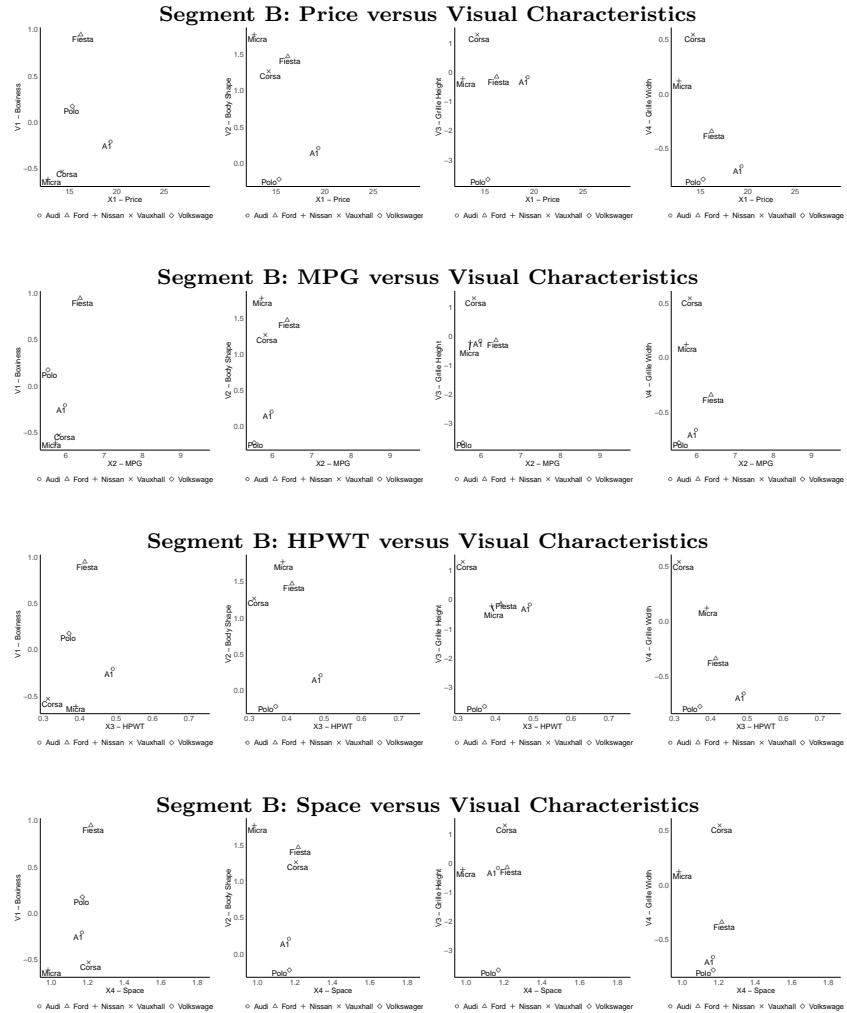
**Figure EC.45 (Color Online) Segment C, E & M: Market Structure Map for 2013****Figure EC.46 (Color Online) Segment C, J & M: Market Structure Map for 2013****Figure EC.47 (Color Online) Segment D, E & J: Market Structure Map for 2013**

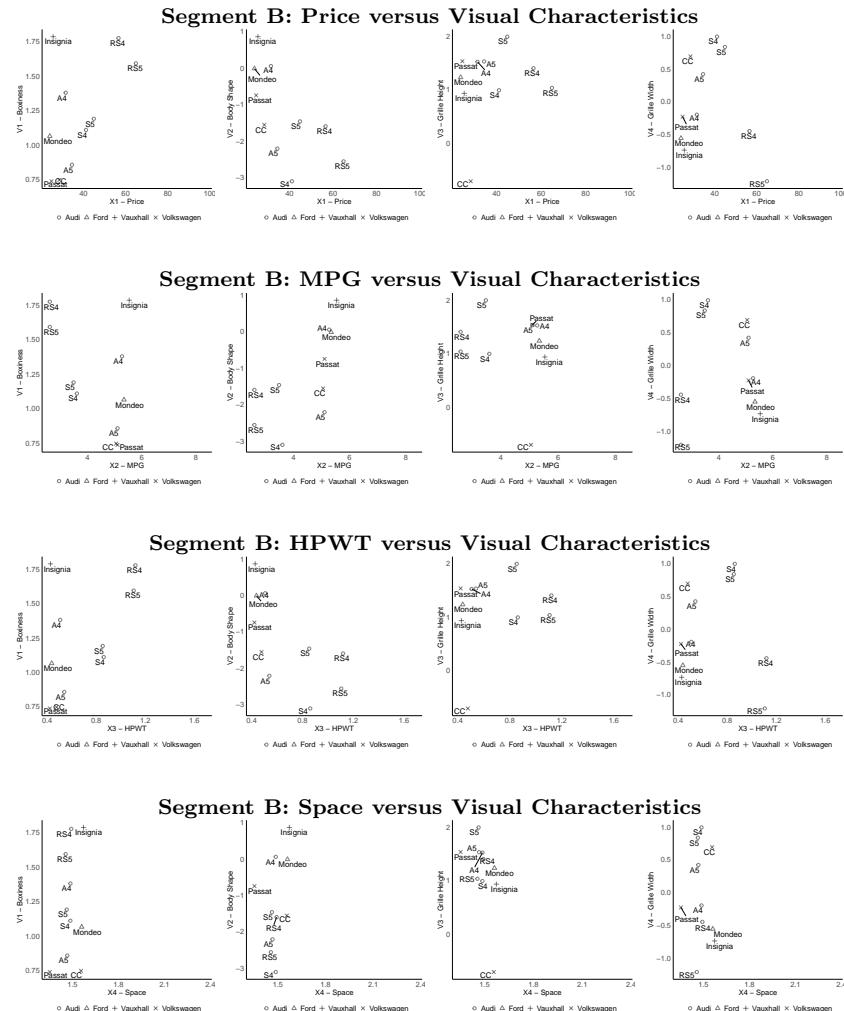
**Figure EC.48 (Color Online) Segment D, E & M: Market Structure Map for 2013****Figure EC.49 (Color Online) Segment D, J & M: Market Structure Map for 2013****Figure EC.50 (Color Online) Segment E, J & M: Market Structure Map for 2013**

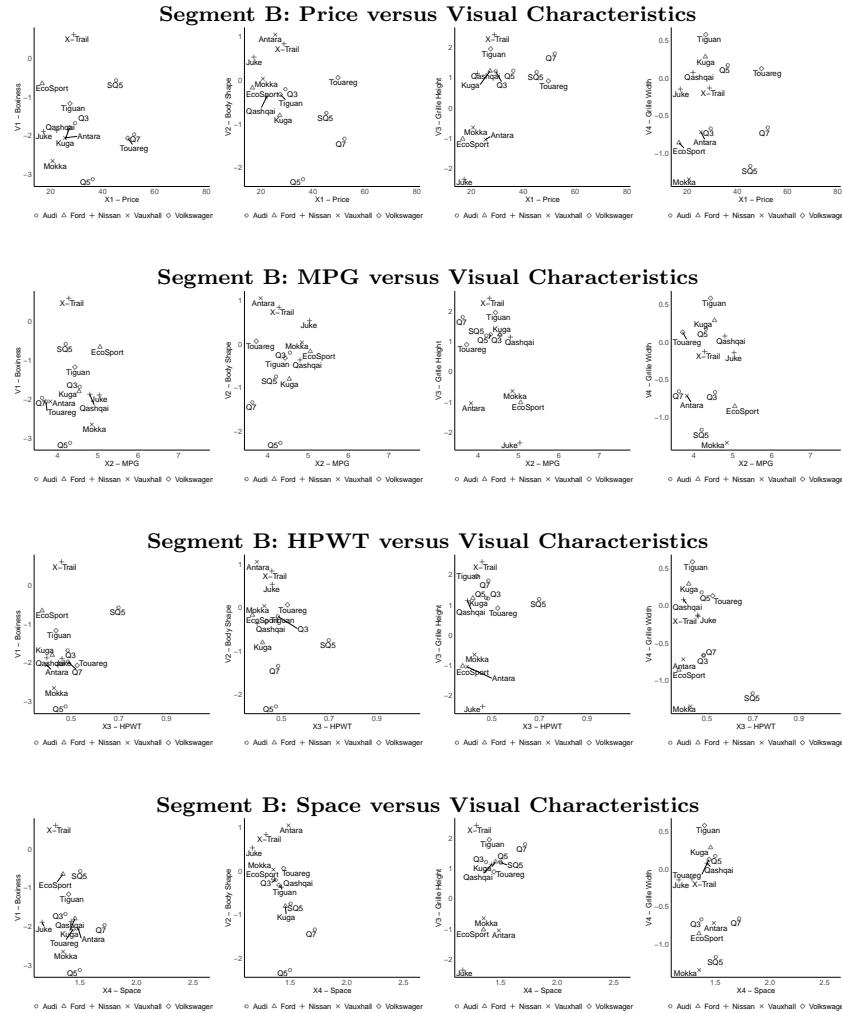
### **Appendix I: Maps of Two Characteristics By a Time in 2013**

Finally, we aim to understand why make-models are located close together or further apart in market structure maps. The use of interpretable product characteristics (both structured and visual) allows us to do so. A survey based approach to understand visual similarity between a pair of make-models would not allow us to understand the reason for why make-models are located close together or further apart in market structure maps. Figure EC.54 plots 6 characteristic by characteristic for both structured as well as visual product characteristics for Segment B, D and J.

**Figure EC.51 Segment-Wise Same-Characteristic Map for 2013**

**Figure EC.52 Segment B Cross-Characteristic Map for 2013**

**Figure EC.53 Segment D Cross-Characteristic Map for 2013**

**Figure EC.54 Segment J Cross-Characteristic Map for 2013**

## Appendix J: 32x32 Resolution

In this section, we explore the impact of using a lower image resolution (32x32 pixels) on our methodology. Table EC.3 compares the performance of different supervisory approaches. Figure EC.55 shows that we only discover two (ignoring color) visual characteristics at this resolution

**Table EC.3 Comparison of Different Supervisory Approaches at 32x32**

Number of Signals	Supervisory Signals	$\lambda_1$	$\lambda_2$	UDR
1	Make-Model FE	30	40	0.694
0	Unsupervised $\beta$ -TCVAE	40	0	0.743
3	HP/Weight, MPG, Space	50	10	0.587
1	Price	30	20	0.709

**Figure EC.55 Discovered Visual Characteristics @ 32x32**



Left to Right: Vary one visual characteristic, keeping all others fixed

## Appendix K: 64x64 Resolution

In this section, we explore the impact of using a lower image resolution (64x64 pixels) on our methodology.

Table EC.4 compares the performance of different supervisory approaches. Figure EC.56 shows that we only discover three (ignoring color) visual characteristics at this resolution

**Table EC.4 Comparison of Different Supervisory Approaches at 64x64**

Number of Signals	Supervisory Signals	$\lambda_1$	$\lambda_2$	UDR
1	Make-Model FE	50	40	0.762
0	Unsupervised $\beta$ -TCVAE	50	0	0.800
3	HP/Weight, MPG, Space	40	40	0.772
1	Price	50	30	0.764

**Figure EC.56 Discovered Visual Characteristics @ 64x64**



Left to Right: Vary one visual characteristic, keeping all others fixed