**Proceedings of the ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference**
**IDETC/CIE 2018**
**August 26-29, 2018, Quebec City, Quebec, Canada**

**DETC2018-86245**



# MODELING SPATIOTEMPORAL HETEROGENEITY OF CUSTOMER PREFERENCES IN ENGINEERING DESIGN

**Youyi Bi**
Integrated Design Automation Laboratory
Northwestern University
Evanston, IL, USA

**Jian Xie**
Industrial Engineering Laboratory
Beijing Institute of Technology
Beijing, China

**Zhenghui Sha**
System Integration & Design Informatics Laboratory
University of Arkansas
Fayetteville, AR, USA

**Mingxian Wang**
Ford Analytics
Global Data Insight & Analytics
Ford Motor Company
Dearborn, MI, USA

**Yan Fu**
Ford Analytics
Global Data Insight & Analytics
Ford Motor Company
Dearborn, MI, USA

**Wei Chen***
Integrated Design Automation Laboratory
Northwestern University
Evanston, IL, USA

## ABSTRACT
Customer preferences are found to evolve over time and correlate with geographical locations. Studying spatiotemporal heterogeneity of customer preferences is crucial to engineering design as it provides a dynamic perspective for a thorough understanding of preference trend. However, existing analytical models for demand modeling do not take the spatiotemporal heterogeneity of customer preferences into consideration. To fill this research gap, a spatial panel modeling approach is developed in this study to investigate the spatiotemporal heterogeneity of customer preferences by introducing engineering attributes explicitly as model inputs in support of demand forecasting in engineering design. In addition, a step-by-step procedure is proposed to aid the implementation of the approach. To demonstrate this approach, a case study is conducted on small SUV in China's automotive market. Our results show that small SUVs with lower prices, higher power, and lower fuel consumption tend to have a positive impact on their sales in each region. In understanding the spatial patterns of China's small SUV market, we found that each province has a unique spatial specific effect influencing the small SUV demand, which suggests that even if changing the design attributes of a product to the same extent, the resulting effects on product demand might be different across different regions. In understanding the underlying social-economic factors that drive the regional differences, it is found that Gross Domestic Product (GDP) per capita, length of paved roads per capita and household consumption expenditure have significantly positive influence on small SUV sales. These results demonstrate the potential capability of our approach in handling spatial variations of customers for product design and marketing strategy development. The main contribution of this research is the development of an analytical approach integrating spatiotemporal heterogeneity into demand modeling to support engineering design.
**Keywords:** spatiotemporal heterogeneity, customer preference, spatial panel model, demand forecasting, engineering design

## 1. INTRODUCTION
Customer preference models support product design in many aspects [1–9] as they can quantitatively characterize the interrelationship between market demand, engineering design attributes, and customer demographics. In the past decade, Discrete Choice Analysis (DCA) [10] has been prevalent in modeling customers' choice behaviors in engineering design [11–17]. To overcome DCA's limitations in dealing with the dependency of alternatives and the collinearity of design attributes [18], recent studies explored the capability of leveraging complex network theory, such as the social network analysis [19, 20], stochastic network modeling [21, 22], multidimensional network analysis [18, 23], and two-stage bipartite network analysis [24] in modeling customer preferences.

*Corresponding author, Wilson-Cook Professor in Engineering Design, Northwestern University
Email: weichen@northwestern.edu

Modeling customer preferences is inherently difficult due to the challenges in modeling various forms of heterogeneity in customer behaviors [25–27]. Different approaches have been used to capture the heterogeneity within the DCA framework, for example, by introducing the error term in utility functions [6,12,14] to capture systematic heterogeneity, and using random effect model (e.g., the mixed logit model) to capture heterogeneous preferences among individual customers. In addition to these forms of heterogeneity, customer preferences may also evolve over time [28, 29] and correlate with geographical locations [30, 31]. Human behaviors at one location can diffuse to adjacent locations that share similar socioeconomic status [32]. Nevertheless, none of the aforementioned approaches take the spatiotemporal heterogeneity of customer preferences into consideration.

Understanding the spatiotemporal heterogeneity of customer preferences supports strategic decisions in product development (e.g. to determine the release of certain products in specific regions) in highly dynamic and diversified markets, such as China's automotive market. Along with the rapid trend of urbanization, China's automotive market has shown strong regional characteristics [33]. According to a McKinsey research [34], for example, customers in Hangzhou and Shandong care more about attractive external styling while their counterparts in Shanghai and Fujian are less concerned with exterior appearance and more sensitive to price. These differences exist even though these four regions are all located along China's eastern coast. In 2016, 2.2% private vehicles sold in Beijing are New Energy Cars (e.g. hybrid vehicles, electric plug-in vehicles, etc.), while this percentage in Shandong Province is only 0.4% based on a recognized survey representing China's automotive market [24]. Chinese customers now exhibit much more differentiated behavior than before, and automakers must prepare to serve those diverse needs [33]. Therefore, without a thorough understanding on the spatiotemporal heterogeneity of customer preferences, it is challenging for designers to create targeted and personalized products and for companies to develop localized marketing strategies.

Previous research in consumer studies and marketing on trend analysis of customer preferences has been mainly focused on demand forecasting [35–37], without explicitly considering the impact of engineering design attributes. In engineering design, researchers are able to utilize open data (e.g. online customer reviews and social media) to extract product features [38] and predict emerging product design trends [39, 40] based on data mining and machine learning techniques or time-series analysis [40]. These methods may achieve high accuracy, however, their interpretability is insufficient because the underlying reasons for the observed preference trends are unattainable from those methods.

In this research, we propose to employ spatial panel models [41] for understanding and analyzing the spatiotemporal heterogeneity in product demand and the impact of geographical, social, and economic factors from different regions. Rooted in spatial econometrics and regional science [42–44], spatial panel models are effective for modeling the lagged effect of dependent variables and independent variables in both space and time [45]. Our contribution in this work is to extend the spatial panel models to the engineering design field for modeling product demand as a function of engineering design attributes, customer attributes, and regional attributes as well as spatiotemporal effects. This enables quantitative assessments of the spatial patterns of customer preferences to support design for customization. Current research in this paper shows the spatiotemporal heterogeneity of product demand at the aggregate level (e.g. provinces in China). Once the heterogeneity is identified, disaggregated consumer preference models can be created for individual provinces to further examine the differences in consumer preferences.

The structure of the remaining paper is arranged as follows. Section 2 reviews the existing techniques of spatiotemporal data modeling, especially the spatial panel models. In Section 3, a step-by-step procedure for understanding spatiotemporal heterogeneity of customer preferences is introduced. The procedure is demonstrated using the China's automotive market as an example, covering dataset preparation, key attributes identification, and results from multiple forms of spatial panel models. Section 4 discusses the implications of results. Section 5 is the closure of the paper.

## 2. SPATIOTEMPORAL DATA AND SPATIAL PANEL MODELS

### 2.1. Spatial interaction effects and static spatial panel model

Spatiotemporal data is the data to which labels were added showing where and when the data was collected [46]. Depending on whether the spatial data is continuous or discrete, spatiotemporal models can be classified into two main categories: point-referenced models and lattice/panel models. Point-referenced models are appropriate in modeling data in which locations are points with coordinates (e.g. longitude-latitude), such as the daily ozone concentration observed at 28 monitoring sites in the state of New York from July to August in 2006 [47]. The lattice/panel models, as known as spatial panel models, are used for modeling repeated observations on the same set of regions over time, such as the public capital productivity in 48 US states observed over 17 years [48]. In these types of models, a spatial weights matrix $W$ is often used to describe the spatial adjacency structure of the geographical units.

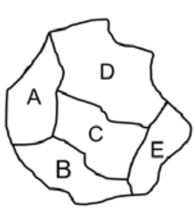
The theoretical foundation of spatial panel models originates from three basic spatial interaction effects [45]. Suppose $y$ and $x$ represent the dependent and independent variables observed in a spatial unit, respectively, **endogenous interaction effects** measure how the value of the dependent variable for one spatial unit is jointly determined with that of neighboring units ($y$ of unit A $\leftrightarrow$ $y$ of unit B). **Exogenous interaction effects** measure how the dependent variable of a particular unit depends on independent explanatory variables of other units ($y$ of unit A ← $x$ of unit B). **Interaction effects among the error terms** measure how the unobserved factors in neighboring spatial units influence each other (error term $u$ of unit A ↔ error term $u$ of unit B). A comprehensive *spatial dependence model* can be represented in Eqns. (1) and (2):

$$\boldsymbol{Y} = \lambda \boldsymbol{W}\boldsymbol{Y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{W}\boldsymbol{X}\boldsymbol{\theta} + \boldsymbol{u} \tag{1}$$

$$\boldsymbol{u} = \rho \boldsymbol{W}\boldsymbol{u} + \boldsymbol{\varepsilon} \tag{2}$$

where $\boldsymbol{Y}$ is an $N \times 1$ vector $(y_1, y_2, \dots, y_N)$ consisting of one observation on the dependent variable for every space units in the sample, $N$ is the number of space units, and $\boldsymbol{W}$ is a nonnegative $N \times N$ spatial weights matrix. $\boldsymbol{WY}$ is the endogenous interaction effect among the dependent variable $y$, $\boldsymbol{WX}$ is the exogenous interaction effects among the independent variables $x$, and $\boldsymbol{Wu}$ is the interaction effects among the disturbance term of different units. $\boldsymbol{X}$ is an $N \times K$ matrix of exogenous explanatory variables, $\boldsymbol{u}$ is error term, and $\boldsymbol{\beta}, \boldsymbol{\theta}$ are associated $K \times 1$ vectors with unknown parameters to be estimated. $\lambda$ is the spatial autoregressive coefficient, $\rho$ is the spatial autocorrelation coefficient, and $\boldsymbol{\varepsilon} = (\varepsilon_1, \ \dots, \ \varepsilon_N)^T$ is a vector of disturbance terms, where $\varepsilon_i$ is independent and identically distributed following standard Normal distribution, $(0, \ \sigma^2)$.

A typical spatial weights matrix $\boldsymbol{W}$ is a adjacency matrix of regions whose diagonal elements are set as 0 and for each observation (row) those regions (columns) that belong to its neighborhood are set as 1 (see Fig. 1). Two regions are neighbors if they share a side.




| | A | B | C | D | E |
|---|---|---|---|---|---|
| A | 0 | 1 | 1 | 1 | 0 |
| B | 1 | 0 | 1 | 0 | 1 |
| C | 1 | 1 | 0 | 1 | 1 |
| D | 1 | 0 | 1 | 0 | 1 |
| E | 0 | 1 | 1 | 1 | 0 |

**Figure 1. A typical spatial weights matrix. Note that A and E do not share a side, thus their corresponding cell in the matrix is set to zero.**

When modeling spatial panel data, i.e. data containing time series observations of a number of geographical units [45], the above spatial dependence model can be extended by adding the time index $(t)$. See the Eqns. (3) and (4) in Fig. 2, which illustrates the role of each term in the static spatial panel model. Note here two additional effects need to be considered: **spatial specific effects** $\boldsymbol{\mu}$, control for all time-invariant variables (e.g. norms and values regarding labor, crime and religion in a region) and **time-period specific effects** $\boldsymbol{\xi}$, control for all space-invariant variables (e.g. one year marked by economic recession, the other by a boom; changes in legislation or government

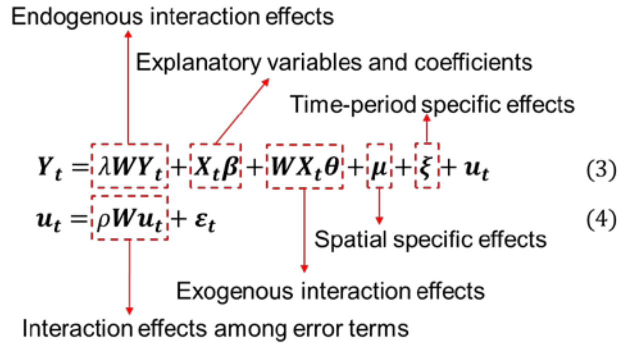



**Figure 2. The role of each term in the static spatial panel model.**

policy). Spatial specific effects can be treated as fixed effects or random effects. When the spatial regression analysis are constrained with specific geographical units such as the 48 contiguous states in the United States, usually a fixed effects model is a better choice [42].

### 2.2. Spatial Auto Regressive Model (SAR) and Spatial Error Model (SEM)

In practice, simplified spatial panel models are used since the full model is hard to explain and its parameters are difficult to estimate. For example, in spatial lag model, also known as the Spatial Auto Regressive Model (SAR), only the endogenous interaction effect among the dependent variable is considered $(\boldsymbol{\theta}, \rho = 0)$. See Eqn. (5).

$$\boldsymbol{Y_t} = \lambda \boldsymbol{W}\boldsymbol{Y_t} + \boldsymbol{X_t}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\xi} + \boldsymbol{\varepsilon_t} \tag{5}$$

While in Spatial Error Model (SEM**)**, only the interaction effect between error terms is considered $(\boldsymbol{\theta}, \lambda = 0)$. See Eqns. (6) and (7).

$$\boldsymbol{Y_t} = \boldsymbol{X_t}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\xi} + \boldsymbol{u_t} \tag{6}$$

$$\boldsymbol{u_t} = \rho \boldsymbol{W}\boldsymbol{u_t} + \boldsymbol{\varepsilon_t} \tag{7}$$

When none of the spatial dependence and space/time specific effects is considered (i.e. $\boldsymbol{\theta}, \lambda, \rho, \boldsymbol{\mu}, \boldsymbol{\xi} = 0$), a spatial panel model simply becomes a linear regression model:

$$\boldsymbol{Y_t} = \boldsymbol{X_t}\boldsymbol{\beta} + \boldsymbol{\varepsilon_t} \tag{8}$$

Eqns. (9) and (10) are used to make predictions on spatial panel data for fixed effects SAR model and SEM model, respectively, according to Baltagi et al. [49] and Elhorst's work [50], where $T + C$ is a future time period, $\hat{\lambda}$ is the estimated spatial auto regression coefficient, $\hat{\beta}_{GMM}$ is the estimated coefficients using Generalized Method of Moments (GMM), $\hat{\mu}$ is the estimated spatial specific effect.

$$\hat{y}_{T+C} = \left(I_N - \hat{\lambda} W\right)^{-1} (X_{T+C}\hat{\beta}_{GMM} + \hat{\mu}) \tag{9}$$

$$\hat{y}_{T+C} = X_{T+C}\hat{\beta}_{GMM} + \hat{\mu} \tag{10}$$

### 2.3. Dynamic spatial panel model

The models represented in Eqns. (3) and (4) are static spatial panel models as they just pool time-series cross-sectional data [45]. As an extension to static spatial panel models, dynamic spatial panel models (see Eqn. (11)) are able to deal with serial dependence between the observations on each spatial unit over time and independent variables lagged in space and/or time:

$$\boldsymbol{Y}_t = \tau \boldsymbol{Y}_{t-1} + \delta \boldsymbol{W}\boldsymbol{Y}_t + \eta \boldsymbol{W}\boldsymbol{Y}_{t-1} + \boldsymbol{X}_t \boldsymbol{\beta} + \boldsymbol{\varepsilon}_t \quad (11)$$

where $\boldsymbol{Y}_{t-1}$ is a dependent variable lagged in time, $\boldsymbol{W}\boldsymbol{Y}_t$ is a dependent variable lagged in space, $\boldsymbol{W}\boldsymbol{Y}_{t-1}$ is a dependent variable lagged in both space and time. Note that Eqn. (11) is a simplified version of the full dynamic spatial panel model. A full model will include the independent variables lagged in time $\boldsymbol{X}_{t-1}$, lagged in space $\boldsymbol{W}\boldsymbol{X}_t$, and lagged in both space and time $\boldsymbol{W}\boldsymbol{X}_{t-1}$. Elhorst stressed that the full model suffers from identification problems and is not useful for empirical research [45].

In summary, spatial panel models can characterize how a dependent variable is impacted by dependent or independent variables within different spatial units (e.g. data of adjacent regions) and/or time periods (e.g. data of previous years). They can complement existing techniques on demand modeling by explicitly considering spatiotemporal heterogeneity. In this research, we investigate the use of the spatial lag model (SAR) in Eqn. (5) and spatial error model (SEM) in Eqns. (6-7) to understand the spatial dependence of customer preferences between different regions.

## 3. EXTENDING SPATIAL PANEL MODELS FOR MODELING HETEROGENEITY OF CUSTOMER PREFERENCES

### 3.1. A step-by-step procedure for understanding spatiotemporal heterogeneity of customer preferences

Figure 3 illustrates a step-by-step procedure for understanding spatiotemporal heterogeneity of aggregate customer preferences

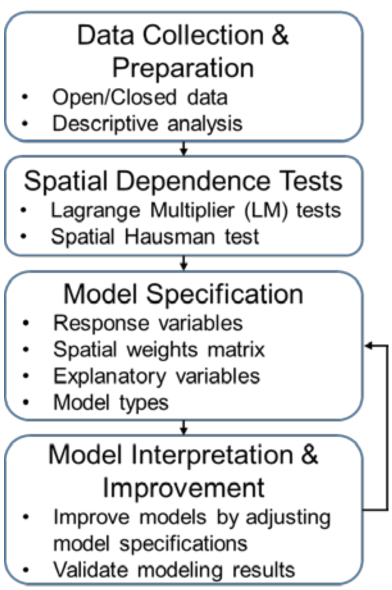



**Figure 3. A step-by-step procedure for understanding spatiotemporal heterogeneity of aggregate customer preferences in engineering design**

in engineering design. In the rest of this subsection, a detailed description of each step is given.

*Step 1: data collection and preparation*. Data sources for understanding spatiotemporal heterogeneity of customer preferences can be open data and closed data. Open data refers to data that can be freely used, re-used, and redistributed by anyone [51], including patents, scientific publications, product forums, social media networks, design repository databases, etc. Closed data limits the replicability of the research process and results, including surveys, experimental data, crowdsourced data, etc. For either source, it is important to prepare a clean dataset with clear space and time tags. Descriptive statistics and visualization techniques can provide an intuitive impression of the dataset, especially its spatial patterns. In our study, both open data and closed data are leveraged (see Sec. 3.2 for details).

*Step 2: spatial dependence tests*. Spatial dependence is the most fundamental assumption for spatial panel models. If no spatial dependence exists among the data collected, it is not necessary to build spatial panel models. Common spatial dependence tests include Lagrange multiplier (LM) Tests [52] and the Hausman Test [53], which can be used to examine the spatial dependence and provide clues for model specification (e.g. choice between a random effects model and a fixed effects model, SAR or SEM).

*Step 3: model specification*. After passing the spatial dependence tests, the modeling parameters and model types need to be specified. Response variables ($\boldsymbol{Y}$) could be the demand (sales) or subjective rating of a specific product (e.g. VW Jetta), product segment (e.g. small SUV) or product feature (e.g. turbo-charged engine). Spatial weights matrix ($\boldsymbol{W}$) as defined in Fig. 1 is used to describe the distance between spatial units. This distance can be measured by geographical proximity or demographic similarity, representing the geographical, social or economic relations between different regions. Explanatory variables ($\boldsymbol{X}$) can include customer demographics (e.g. age, income. etc.), regional characteristics (e.g. GDP, population, etc.), and product attributes. Common methods for selecting explanatory variables include multi-collinearity analysis, stepwise logistic regression and principal component analysis. These methods can help determine the most influential explanatory variables. Besides the SAR and SEM models mentioned in Sec. 2, there are some other spatial panel models, such as the Spatial Durbin Model, Spatial Autoregressive Confused (SAC) model and Spatial Lag of X (SLX) model. One can refer to [54] for a detailed description about the characteristics of these models and how to choose from them. In this study, we investigate the use of SAR and SEM models for modeling the heterogeneity of customer preference across different regions because they are the most fundamental models to characterize the spatial dependence between different regions.

*Step 4: model interpretation and improvement*. After specifying the model parameters and types, the associated coefficients can be estimated using methods such as maximum likelihood and generalized moments [55]. These coefficients need to be checked to see if they are consistent with the real space-time patterns of customer preferences. The model specifications may need to be adjusted iteratively until the models with good interpretability are obtained.

The ultimate application of this approach is to extract insights from spatiotemporal data — how customer preferences change with space and time. These insights can be used to support demand forecasting for engineering design and creating new or improving existing design, product features, or product segments for differentiated customer groups.

### 3.2. Dataset of China's automotive Market

To demonstrate the proposed approach, we conduct a case study employing the data from a recognized, reputable survey representing China's automotive market [24]. This survey data consists of about 30,000 to 50,000 new car buyers' responses and purchase history covering about 400 different vehicle models in China's market from 2012 to 2016. Respondents were asked to list the cars they purchased with their residential information and purchase time recorded. The vehicle's attributes, such as engine power and fuel consumption, are reported by customers in the survey and verified by the data company.

Our focus in this study is on the small SUV segment, as the demand for small SUVs has been rising rapidly and increasingly affluent Chinese buyers opt for more spacious vehicles [56]. According to the survey data, 14.8% of the respondents purchased a small or mini SUV in 2012 and this percentage increased to 21.6% in 2016. The survey data was collected every four months in each year (called one wave), thus we have three-wave data for each year and 15-wave data for five years. We only considered 27 provinces in mainland China in our modeling, as these provinces have complete sales data of small SUV for 15 waves. We also collected open data about the regional statistics (demography and socioeconomics) of these provinces from the National Bureau of Statistics of China [57].

### 3.3. Descriptive analysis of the key variables

Based on our previous research of customer preference in vehicle consideration and choice [24], we identified three vehicle attributes — *price, power, and fuel consumption*, one customer attribute — *monthly household income*, and chose three regional attributes that have been broadly studied in automotive market research [58,59]— *GDP per capita, household consumption expenditure and length of paved roads per capita* to capture the engineering, demographic and regional effects on small SUV sales. Note the vehicle attributes and customer attributes are taken from the survey [24], while the regional attributes are obtained from [57]. Table 1 provides the descriptive statistics of these attributes at the province level. It can be seen that the mean sales of small SUVs, GDP per capita, household consumption expenditure and length of paved roads per capita of each province in 2016 is larger than those in 2012, while the mean price, power, fuel consumption, and household income in 2016 is smaller than those in 2012. These results may imply that small SUVs are increasingly popular as they are becoming more affordable. To meet these customers' needs, car manufacturers may tend to build smaller SUVs with lower prices, power, and fuel consumption.

**Table 1. Descriptive statistics of the key variables (province-level) between 2012 and 2016**

| | Mean (SD) | Min | Max |
|---|---|---|---|
| **2012** | | | |
| Sales (units) | 102.9 (85.3) | 9 | 351 |
| Price (10,000 RMB) | 23.6 (1.2) | 20.6 | 26.6 |
| Power (BHP) | 165 (5.8) | 152.2 | 180.5 |
| Fuel consumption (liter/100 km) | 11.1 (0.5) | 10.3 | 12.5 |
| Household income (1,000 RMB) | 15.7 (4.2) | 9.2 | 26.8 |
| GDP per capita (10,000 RMB) | 4.3 (1.9) | 1.9 | 9.1 |
| Household Expend. (1000 RMB) | 14.3 (6.7) | 5.3 | 36.9 |
| Length of paved roads per capita (km/10, 000 residents) | 2.4 (1.1) | 0.7 | 4.6 |
| **2016** | | | |
| Sales (units) | 421.8 (350.2) | 15 | 1119 |
| Price (10,000 RMB) | 15.2 (1.5) | 12.9 | 18.1 |
| Power (BHP) | 150.4 (4.2) | 141.9 | 158.9 |
| Fuel consumption (liter/100 km) | 9.1 (0.6) | 8.4 | 10.8 |
| Household income (1,000 RMB) | 13.7 (3.9) | 8.0 | 23.3 |
| GDP per capita (10,000 RMB) | 5.7 (2.6) | 2.8 | 11.8 |
| Household Expend. (1000 RMB) | 21.3 (9.8) | 9.7 | 49.6 |
| Length of paved roads per capita (km/10, 000 residents) | 2.8 (1.1) | 1.1 | 5.6 |

*Note*: Standard deviations are in parenthesis, and BHP stands for Brake Horsepower.

Figure 4 presents the spatial distribution of a few selected variables in 2016. These graphs provide an intuitive impression of the relationship between the sales of small SUVs and some explanatory variables. For example, the sales of small SUVs seems to positively correlate with income and household consumption expenditure, and negatively correlate with price. We expect to obtain consistent results but in a quantitative way from the spatial panel models.

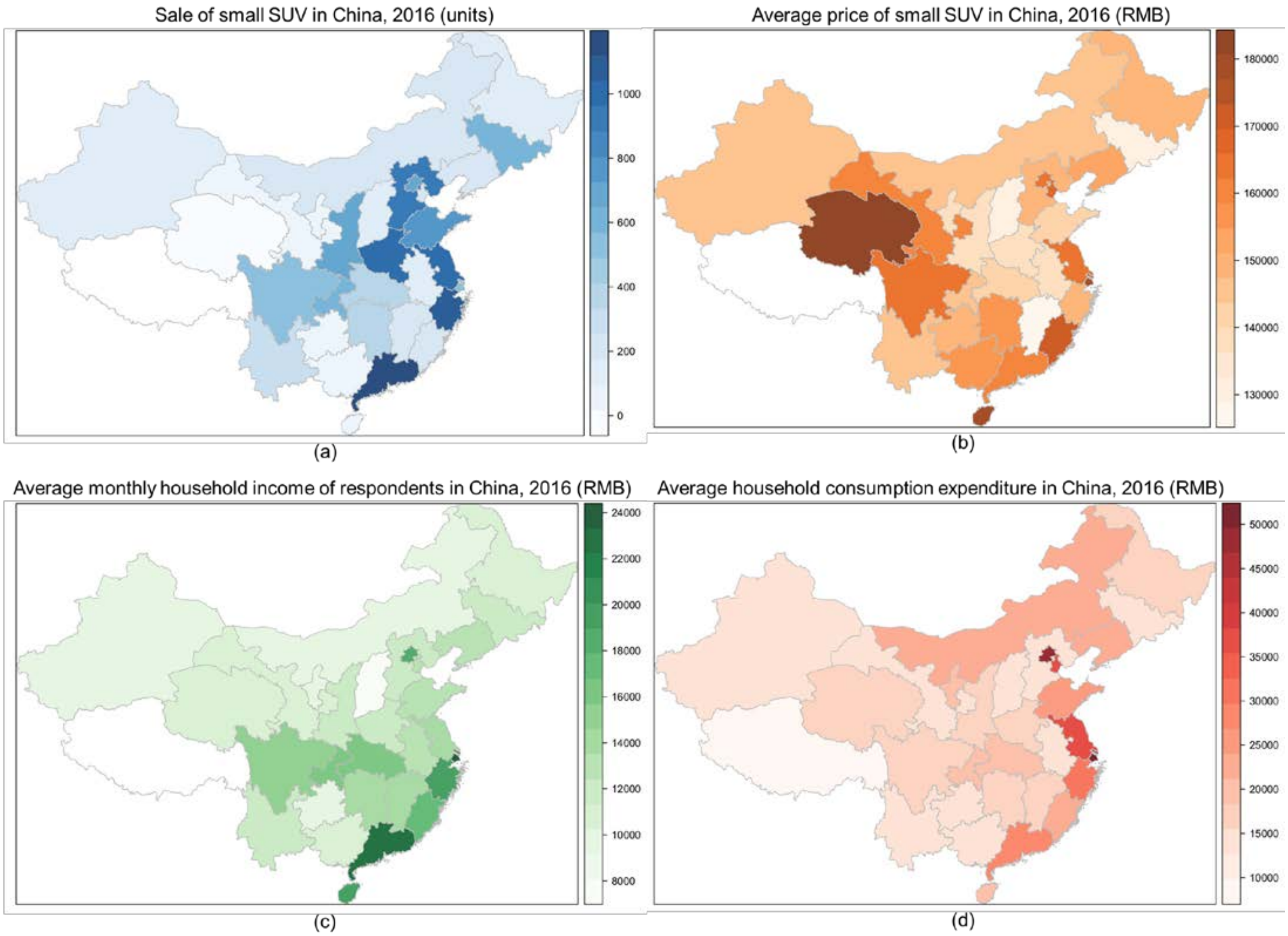



**Figure 4. The spatial distribution of certain model variables in 2016: a) sale of small SUVs, b) price, c) monthly household income, d) household consumption expenditure. White areas represent missing data.**

### 3.4. Spatial dependence tests

Spatial dependence tests can examine the existence of spatial effects and provide clues for model specifications. Table 2 presents the results of two tests: $LM_H$ and Hausman Test with their respective null hypotheses ($H_0$). Detailed procedures for running these tests can be found in [52]. As the p-values shown in Table 2, the null hypotheses of both tests could be rejected at the 5% significance level. The result of the $LM_H$ test suggests that at least one of the spatial autoregressive coefficient ($\lambda$) and the variance of spatial specific effects ($\sigma_\mu^2$) is not zero. This result implies the existence of the spatial effect of the dependent variable — small SUV sales in China. The result of Hausman Test implies that the assumption of random effects is not supported by the data, and fixed effects models may perform better than random effects models.

**Table 2. Results of spatial dependence tests**

| Test | $H_0$ | statistic | p-value |
|---|---|---|---|
| $LM_H$ | Spatial autoregressive coefficient ($\lambda$) and variance of spatial specific effects ($\sigma_\mu^2$) are both zero | 783.84 | <0.001 |
| Hausman Test | Random effects assumption is supported by the data | 44.83 | <0.001 |

### 3.5. Results and interpretations of static spatial panel models

#### 3.5.1. Model specifications

After passing the spatial dependence tests, we build representative static spatial panel models to quantify the impacts of the identified explanatory variables associated with spatial specific effects and time-period specific effects on the sales of small SUVs. The modeling parameters are chosen as follows:

- Response variable ($\boldsymbol{Y}$): small SUV sales in 27 provinces of China in each wave from 2012 to 2016. The number of the surveyed respondents is highly correlated with the number of new vehicle registrations over multiple years in each province of China ($r = 0.931$), thus we assume the survey data can represent the actual sales numbers.
- Spatial weights matrix ($\boldsymbol{W}$): binary geographical adjacency matrix adjusted with *per capita disposable income* of each province. This adjustment is based on the assumption that wealthier regions produce larger spatial influence on their neighbor regions. Eqn. (12) represents the calculation of this adjusted spatial weights matrix:

$$W = w * diag\left(\frac{\overline{y_1}}{\bar{y}}, \frac{\overline{y_2}}{\bar{y}}, \dots, \frac{\overline{y_n}}{\bar{y}}\right),$$

$$\bar{y}_i = \frac{1}{t_T - t_1 + 1} \sum_{t_1}^{t_T} y_{it},\ \ \bar{y} = \frac{1}{t_T - t_1 + 1} \sum_{i=1}^{n} \sum_{t_1}^{t_T} y_{it} \qquad (12)$$

  where $w$ is the geographical adjacency matrix, $diag()$ is a diagonal matrix with all non-diagonal elements set to zero, $y_{it}$ is the per capita disposable income of province $i$ in time $t$. $\bar{y}_i$ is the average per capita disposable income of province $i$ from $t_1$ to $t_T$ ($T$ time periods totally), and $\bar{y}$ is the average per capita disposable income of all provinces from $t_1$ to $t_T$. For the ease of interpretation, $W$ is normalized such that the elements of each row sum to unity [45]. Note that this binary adjacency matrix is the most frequently used form of spatial weights matrix, and the economic interactions between non-adjacent regions are not considered here. This assumption may not be always true in practice, and other forms of spatial weights matrix reflecting the socioeconomic or cultural interactions between both adjacent and non-adjacent regions can be employed in future research.
- Explanatory variables ($\boldsymbol{X}$): three vehicle attributes, one customer attribute and three regional attributes as presented in Sec. 3.3.

### 3.5.2. Estimated spatial parameters and coefficients of explanatory variables

Table 3 presents a summary of the estimated coefficients of four different spatial panel models and one linear regression model as a benchmark. The response variable and explanatory variables are transformed to natural logarithms when estimating these models. The difference between the random effects model and the fixed effect model depends on whether the spatial specific effects ($\mu$) are treated as random effects or fixed effects in the estimation process as describe in Sec. 2. In SAR models (Eqn.5), the dependence of the response variables in different regions is estimated and denoted by the spatial autoregressive coefficient ($\lambda$). In SEM models (Eqns. 6-7), the dependence of the error terms in different regions is estimated instead, and denoted by the spatial autocorrelation coefficient ($\rho$). In both SAR and SEM models, in addition to the results shown in Table 3, spatial specific effects ($\boldsymbol{\mu}$) and time-period specific effects ($\boldsymbol{\xi}$) are also estimated (see details in Sec. 3.5.3). Linear regression model is the simplest model with no spatial dependence nor space/time specific effects considered. The $R^2$ of these models suggest that fixed effects models perform better than random effects models. This result is consistent with the implication of model performance obtained from the spatial dependence test in Sec. 3.4. In addition, the $R^2$ of the linear regression model is low, which suggests that the spatial and temporal effects (e.g. $\lambda, \rho$) not included in the linear model, are indeed important. When

**Table 3. Estimated coefficients of five different models. Refer to Eqn. (5) for the SAR model and Eqns. (6-7) for the SEM model.**

| | | Linear Model | Random Effects Models | | Fixed Effects Models | |
|---|---|---|---|---|---|---|
| | | | SAR | SEM | SAR | SEM |
| Spatial Parameters | $\lambda$ | | 0.17** | | 0.14* | |
| | $\rho$ | | | 0.07 | | 0.05 |
| Product Attributes | Price ($\beta_1$) | -2.37*** | -0.85** | -1.04*** | -0.53. | -0.62* |
| | Power ($\beta_2$) | -0.64 | 0.95 | 1.13. | 0.77 | 0.87 |
| | Fuel Consump. ($\beta_3$) | -0.05 | -1.45** | -1.64** | -1.09* | -1.17* |
| Customer Attribute | House. Income ($\beta_4$) | 1.06*** | 0.22. | 0.23. | 0.17 | 0.19 |
| | Length_Roads ($\beta_5$) | -0.29* | 0.46 | 0.38 | 1.78*** | 1.81*** |
| | House Expend. ($\beta_6$) | 0.75* | 0.92* | 1.29** | 0.73 | 1.07* |
| | GDP per capita ($\beta_7$) | 0.64* | 0.52 | 0.47 | 1.06* | 1.09* |
| Goodness of Fit | $R^2$ | 0.53 | -0.02 | 0.39 | 0.84 | 0.84 |

*Note*: $\lambda$, Spatial autoregressive coefficient; $\rho$, Spatial autocorrelation coefficient
. p<.10; * p < .05; ** p < .01; *** p < 0.001

using the first 14-wave data to train the model and the last wave data to test the model, the Root Mean Squared Errors (RSME) of fixed effects SAR model and SEM model are 0.47 and 0.48 respectively, while the RSME of linear regression model is 4.69. Thus, spatial panel models are more accurate than linear regression models in explaining the response variable, as they include the impact of the attributes of products, customers and regions, as well as the associated spatial interactions.

The positive spatial autoregressive coefficient ($\lambda = 0.14, p < .01$) obtained from the fixed effects SAR model (see Eqn. (5)) suggests that a region with higher small SUV sales is likely adjacent to several regions with high small SUV sales. This result makes sense that neighbors influence each other. The spatial autocorrelation coefficients ($\rho = 0.05, p > .05$) estimated from the fixed effect SEM model (see Eqns. (6-7)) implies that unobservable factors (e.g. other explanatory variables not included in our models) have insignificant influence with those in adjacent regions.

When examining the estimated coefficients ($\boldsymbol{\beta}$) of explanatory variables in the fixed effects SEM model, we find that at an aggregated market level small SUVs with a lower price ($\beta_1 = -0.62, p < .05$) and lower fuel consumption ($\beta_3 = -1.17, p < .05$) tend to have a positive impact on sales. The effects of power ($\beta_2 = 0.87, p > .05$) and monthly household income of customers ($\beta_4 = 0.19, p > .05$) are not significant. Among the three regional attributes, length of paved roads per capita ($\beta_5 = 1.81, p < 0.001$), household consumption expenditure ($\beta_6 = 1.07, p < 0.05$) and GDP per capita ($\beta_7 = 1.09, p < 0.05$) all have significant positive influences on the sales of small SUVs. The estimated coefficients of price and household consumption expenditure in spatial panel models are consistent with the visualized distribution of these two attributes shown in Fig. 4. In addition, the estimated coefficients in the fixed effects SAR model are similar to those in the fixed effects SEM model. Since the $R^2$ of the linear regression model is low, the estimated coefficient of length of paved roads ($\beta_5$) is not trust-worthy.

### 3.5.3. Estimated spatial specific effects and time-period specific effects

Spatial specific effect controls for all time-invariant variables contributing to the response variable, which reflects the inherent characteristics of one region. Figure 5 presents the estimated spatial specific effects obtained from the fixed effects SEM model (see Table 4 for the values). As shown in Figure 5, the red regions exhibit negative spatial specific effects. This suggests that these regions have some inner drivers that negatively influence their small SUV sales. This may be due to relatively lower-level of socioeconomic development in these provinces (Xinjiang ($\mu = -1.62$) and Ningxia ($\mu = -2.09$)). By contrast, economically developed areas such as Beijing ($\mu = 0.53$) and Shanghai ($\mu = 1.24$) have positive spatial specific effects (blue regions), which suggest that these cities have certain inner drivers that positively influence their respective small SUV sales. In addition, the results imply that when changing the attributes of a vehicle to the same extent, the influence on vehicle


**Figure 5. Spatial specific effects obtained from the fixed effects SEM model (gray color represents missing of data).**

sales is different from region to region. For example, when decreasing the average price ($x$) of small SUVs by the same amount, the growth rate of sales ($y$) in Shanghai (with larger $\mu$) may be higher than in Shandong (with smaller $\mu$) as shown in Eqn. (6). Assume the average price of small SUV decreased by 20,000 RMB in 2016 (wave 3), the sales of small SUV would increase 4% in Shanghai and 3% in Shandong when using the prediction formula in Eqn. (10). This is consistent with the McKinsey research observation [34] that Shanghai customers are more sensitive to price than Shandong customers. Traditional pricing strategy may not be that effective in the regions with negative space specific effects, and car companies may try to offer more customized vehicles to attract customers with different tastes in these regions.

**Table 4. Estimated spatial specific effects ($\boldsymbol{\mu}$)**

| **Province** ($i$) | $\mu_i$ | **Province** ($i$) | $\mu_i$ |
|---|---|---|---|
| Anhui | -0.55 | Jilin | -0.12 |
| Beijing | 0.53 | Liaoning | -0.08 |
| Chongqing | 0.40 | Inner Mongolia | -1.40 |
| Fujian | -0.09 | Ningxia | -2.09 |
| Gansu | -0.59 | Shandong | 0.17 |
| Guangdong | 0.95 | Shanghai | 1.24 |
| Guangxi | -0.34 | Shaanxi | 0.94 |
| Hebei | 0.92 | Shanxi | -0.18 |
| Heilongjiang | -0.80 | Sichuan | 1.31 |
| Henan | 1.49 | Tianjin | -0.96 |
| Hubei | -0.20 | Xinjiang | -1.62 |
| Hunan | 0.48 | Yunnan | 0.64 |
| Jiangsu | 0.11 | Zhejiang | 0.41 |
| Jiangxi | -0.58 | | |

Time-period specific effects control for all space-invariant variables contributing to the response variable (i.e., the sales), which reflect the regional characteristic of one time-period. Table 5 presents the estimated time-period specific effects obtained from the fixed effects SEM model (see Eqns. (6-7)). It is observed that most time-period specific effects ($\boldsymbol{\xi}$) are *negative* before 2015, and remain *all positive* afterwards. This result suggests that maybe certain nation-wide factors greatly stimulated the sales of small SUVs in China starting in 2015. One possible reason is that since 2015 many domestic car producers began to offer small SUVs with lower prices compared to foreign brands, which stimulated sales to broader populations. Another potential contributing factor is China has reduced the vehicle purchase tax from 10% to 5 % for small passenger cars (engine displacement ≤ 1.6 liters) in 2015 [60]. Social influence might have played a role.

**Table 5. Estimated time-period specific effects ($\xi$)**

| **Time-period** ($t$) | $\xi_t$ | **Time-period** ($t$) | $\xi_t$ |
|---|---|---|---|
| 2012-1 | -0.49 | 2014-3 | -0.03 |
| 2012-2 | -0.41 | 2015-1 | 0.24 |
| 2012-3 | -0.93 | 2015-2 | 0.36 |
| 2013-1 | -0.48 | 2015-3 | 0.18 |
| 2013-2 | -0.19 | 2016-1 | 0.69 |
| 2013-3 | -0.09 | 2016-2 | 0.58 |
| 2014-1 | 0.02 | 2016-3 | 0.62 |
| 2014-2 | -0.07 | | |

## 4. DISCUSSION

A major contribution of this research is the development of an analytical approach integrating spatiotemporal heterogeneity into the modeling of customer preferences. Although spatial effects are widely investigated in marketing studies, these studies do not support engineering design directly (e.g. by including design attributes in the model). Our holistic approach enables the assessment of various driving factors in the small SUV market, such as engineering design attributes, customer demographics, and space and time effects. Static spatial panel models are effective at modeling spatial interaction effects, including endogenous/exogenous interaction effects and interaction effects among the unobservable factors (error terms). Dynamic spatial panel models further consider the time-lag effect of both dependent and independent variables, which are able to model sophisticated space-time interaction effects comprehensively. The proposed procedure for understanding spatiotemporal heterogeneity of customer preferences provides guidance for systematically implementing the modeling work.

We utilize China's automotive market as an example to show the modeling process, including data collection and descriptive analysis, spatial dependence tests, and model specifications. At the beginning of this process, it is important to collect reliable and comprehensive data on regional characteristics (e.g., GDP, population, etc.). Not only can these data be employed as explanatory variables, but they will also assist with creating spatial weights matrix, which reflects the geographic, social, economic, demographic, and cultural interactions among spatial units, separately or synthetically. Descriptive analysis on one hand helps check the validity and consistency of the collected data. On the other hand, it provides us with a preliminary understanding about the spatial patterns of the collected data. In our case study, in those regions with lower small SUV sales, the average price tends to be higher than those with higher sales. Based on this observation, we expect to see a negative effect of vehicle price on the sale of small SUVs, which is later confirmed in the spatial panel models. Descriptive analysis can also examine whether results obtained from spatial panel models are trustworthy.

The motivation of integrating spatiotemporal heterogeneity into customer preference modeling originates from the hypothesis that customer behaviors are not spatially independent. LM Tests and Hausman Test can assess the existence of spatial dependency and guide the construction of spatial panel models, including choosing between the random effects model and fixed effects model. In our case study, the implications of the spatial dependence tests are consistent with the $R^2$ values obtained in different spatial panel models built afterwards.

Our study shows that spatial panel models are able to quantify the influence of product attributes, customer demographics, and regional characteristics on aggregate customer choices (in our case study, the sales or demands of small SUVs). Among all explanatory variables, GDP capita and the length of paved roads per capita seems to have the most significant positive influence on the sales of small SUV sales. This result indicates that strong economic growth and solid infrastructure in one region are usually associated with higher demand for emerging product segments in China's market. Although the linear regression model provides similar results on selected explanatory variables, one critical advantage of spatial panel models is that they can model and profoundly reveal the spatial dependence between dependent/independent variables in various regions. This advantage allows us to capture how certain customer preferences diffuse spatially. As shown in Figure 5, although Beijing and Tianjin are both economically developed regions in China, their spatial specific effects are completely different. Social, cultural, and government policy factors need to be taken into account to explain these differences in spatial influence. Furthermore, spatial panel models enable the assessment of the time-period specific effects, which reveals the influence of space-invariant factors on the temporal change of customer preferences.

Apart from the methodological contribution to customer preference modeling, the knowledge and insights gained from our work also have implications to engineering design and industry practice. These insights allow vehicle manufacturers to develop customized products and marketing strategies for different regions to improve the market share in a specific region. For example, we find fuel consumption has significantly negative influence on the sales of small SUVs while the influence of power is insignificant, thus car companies may pay more attention to fuel economy in the development and marketing of small SUVs, especially in the regions with higher GDP growth rates as more sales are expected in these regions. The product development team can make region-specific adjustments to the design attributes when offering products in the regions with different spatial characteristics, as the same changes to the design attributes may have weaker influence in those regions with negative spatial specific effects.

## 5. CONCLUSION

An approach based on the spatial panel model for modeling spatiotemporal heterogeneity of customer preferences is developed in this paper. A step-by-step procedure for implementing this approach is proposed, including a descriptive analysis of space-time data, spatial dependence tests, specification of model parameters, and interpretation of the modeling results. Our approach models the influence of explanatory variables (attributes of products, customers, and regions) and associated spatial and temporal effects on aggregated customer preferences, i.e., demand or sales in this paper. Using China's small SUV market as an example, we find a spatial dependence effect wherein one region with higher small SUV sales is likely adjacent to several regions with high sales. Each province may have a unique spatial effect influencing its sales of small SUVs, which suggests that when changing the design attributes of a product to the same extent, the impact on product demand in each region is different. In understanding the impact of design attributes, we find that price and fuel consumption have negative effects on the sales of small SUVs based on the aggregated data over multiple years and multiple regions. Among the underlying social-economic factors, GDP per capita, household consumption expenditure and length of paved roads per capita have positive effects on small SUV sales. In addition, we find that beginning in 2015, some nation-wide factors greatly stimulated the sale of small SUVs in China. These results demonstrate the potential use of our approach in supporting product design and strategic decision-making considering the spatiotemporal variations of customers. We also show how descriptive analysis and spatial dependence tests help specify the modeling parameters and verify the modeling results.

At this exploratory stage, only static spatial panel models are studied. We plan to build dynamic spatial panel models and eventually show the application of dynamic models in predicting customer preferences in both space and time dimensions. Our approach's capability of predicting demand or market share across different spatial regions can directly assist with the specification of design parameters for localized product development. Another limitation of this work is that we only model the spatiotemporal heterogeneity of customer preferences at an aggregate level (demand in this case). Once the regional difference is identified through the method shown in this research, disaggregated consumer preference models can be created for individual provinces using methods like discrete choice analysis or network modeling to further examine the heterogeneity in consumer preferences.

## ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from NSF-CMMI-1436658, the Ford-Northwestern Alliance Project, and a grant from the China Scholarship Council (No. 201706030108).

## REFERENCES

[1] He, L., Chen, W., Hoyle, C., and Yannou, B., 2012, "Choice Modeling for Usage Context-Based Design," J. Mech. Des., **134**(3), p. 31007.

[2] Kumar, D., Hoyle, C., Chen, W., Wang, N., Gomez-Levi, G., and Koppelman, F., 2009, "A Hierarchical Choice Modelling Approach for Incorporating Customer Preferences in Vehicle Package Design," Int. J. Prod. Dev., **8**(3), pp. 228–251.

[3] Hoyle, C. J., and Chen, W., 2009, "Product Attribute Function Deployment (PAFD) for Decision-Based Conceptual Design," IEEE Trans. Eng. Manag., **56**(2), pp. 271–284.

[4] MacDonald, E. F., Gonzalez, R., and Papalambros, P. Y., 2009, "Preference Inconsistency in Multidisciplinary Design Decision Making," J. Mech. Des., **131**(3), p. 31009.

[5] Sha, Z., Saeger, V., Wang, M., Fu, Y., and Chen, W., 2017, "Analyzing Customer Preference to Product Optional Features in Supporting Product Configuration," SAE Int. J. Mater. Manuf., **10**(2017-01–0243).

[6] Kumar, D., Chen, W., and Simpson, T. W., 2009, "A Market-Driven Approach to Product Family Design," Int. J. Prod. Res., **47**(1), pp. 71–104.

[7] Michalek, J. J., Ceryan, O., Papalambros, P. Y., and Koren, Y., 2006, "Balancing Marketing and Manufacturing Objectives in Product Line Design," J. Mech. Des., **128**(6), pp. 1196–1204.

[8] Sha, Z., and Panchal, J. H., 2014, "Estimating Local Decision-Making Behavior in Complex Evolutionary Systems," J. Mech. Des., **136**(6), p. 61003.

[9] Sha, Z., Moolchandani, K., Panchal, J. H., and DeLaurentis, D. A., 2016, "Modeling Airlines' Decisions on City-Pair Route Selection Using Discrete Choice Models," J. Air Transp.

[10] Train, K. E., 2009, *Discrete Choice Methods with Simulation*, Cambridge university press.

[11] He, L., Wang, M., Chen, W., and Conzelmann, G., 2014, "Incorporating Social Impact on New Product Adoption in Choice Modeling: A Case Study in Green Vehicles," Transp. Res. Part D Transp. Environ., **32**, pp. 421–434.

[12] Hoyle, C., Chen, W., Wang, N., and Koppelman, F. S., 2010, "Integrated Bayesian Hierarchical Choice Modeling to Capture Heterogeneous Consumer Preferences in Engineering Design," J. Mech. Des., **132**(12), p. 121010.

[13] Chen, W., Hoyle, C., and Wassenaar, H. J., 2013, *Decision-Based Design: Integrating Consumer Preferences into Engineering Design*, Springer Science and Business Media.

[14] Wassenaar, H. J., and Chen, W., 2003, "An Approach to Decision-Based Design with Discrete Choice Analysis for Demand Modeling," J. Mech. Des., **125**(3), pp. 490–497.

[15] Kumar, D. K. D., Chen, W., and Kim, H. M., 2006,

"Multilevel Optimization for Enterprise-Driven Decision-Based Product Design," Decis. Mak. Eng. Des., pp. 203–215.
[16] Michalek, J. J., Feinberg, F. M., and Papalambros, P. Y., 2005, "Linking Marketing and Engineering Product Design Decisions via Analytical Target Cascading," J. Prod. Innov. Manag., **22**(1), pp. 42–62.
[17] Frischknecht, B. D., Whitefoot, K., and Papalambros, P. Y., 2010, "On the Suitability of Econometric Demand Models in Design for Market Systems," J. Mech. Des., **132**(12), p. 121007.
[18] Wang, M., Chen, W., Huang, Y., Contractor, N. S., and Fu, Y., 2016, "Modeling Customer Preferences Using Multidimensional Network Analysis in Engineering Design," Des. Sci., **2**.
[19] Wichmann, B., Chen, M., and Adamowicz, W., 2016, "Social Networks and Choice Set Formation in Discrete Choice Models," Econometrics, **4**(4), p. 42.
[20] Richards, T. J., Hamilton, S. F., and Allender, W. J., 2014, "Social Networks and New Product Choice," Am. J. Agric. Econ., **96**(2), pp. 489–516.
[21] Wang, M., and Chen, W., 2015, "A Data-Driven Network Analysis Approach to Predicting Customer Choice Sets for Choice Modeling in Engineering Design," J. Mech. Des., **137**(7), p. 71410.
[22] Lee, L., Li, J., and Lin, X., 2014, "Binary Choice Models with Social Network under Heterogeneous Rational Expectations," Rev. Econ. Stat., **96**(3), pp. 402–417.
[23] Wang, M., Chen, W., Huang, Y., Contractor, N. S., and Fu, Y., 2015, "A Multidimensional Network Approach for Modeling Customer-Product Relations in Engineering Design," *ASME 2015 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Boston, MA.
[24] Fu, J. S., Sha, Z., Huang, Y., Wang, M., Fu, Y., and and Chen, W., 2017, "Modeling Customer Choice Preferences in Engineering Design Using Bipartite Network Analysis," *Proceedings of the ASME 2017 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Cleveland, Ohio.
[25] Shocker, A. D., Ben-Akiva, M., Boccara, B., and Nedungadi, P., 1991, "Consideration Set Influences on Consumer Decision-Making and Choice: Issues, Models, and Suggestions," Mark. Lett., **2**(3), pp. 181–197.
[26] Shao, W., 2007, *Consumer Decision-Making: An Empirical Exploration of Multi-Phased Decision Processes*, Griffith University.
[27] Carpenter, G. S., and Nakamoto, K., 1989, "Consumer Preference Formation and Pioneering Advantage," J. Mark. Res., pp. 285–298.
[28] Min, S.-H., and Han, I., 2005, "Detection of the Customer Time-Variant Pattern for Improving Recommender Systems," Expert Syst. Appl., **28**(2), pp. 189–199.
[29] Bi, Y., Li, S., Wagner, D., and Reid, T., 2017, "The Impact of Vehicle Silhouettes on Perceptions of Car Environmental Friendliness and Safety in 2009 and 2016: A Comparative Study," Des. Sci., **3**.
[30] Levitt, T., 1993, "The Globalization of Markets," Readings Int. Bus. a Decis. approach, **249**.
[31] Frank, B., Abulaiti, G., and Enkawa, T., 2014, "Regional Differences in Consumer Preference Structures within China," J. Retail. Consum. Serv., **21**(2), pp. 203–210.
[32] Mosleh, M., and Heydari, B., 2016, "Spatial Diffusion of Risk: The Case of Risky Teenage Drivers," *5th International Engineering Systems Symposium*, George Washington University, Washington DC.
[33] Wang, A., Liao, W., and Hein, A.-P., 2012, "Bigger, Better, Broader: A Perspective on China's Auto Market in 2020," McKinsey Company, New York.
[34] Atsmon, Y., Dixit, V., Magni, M., and St-Maurice, I., 2010, "2010 Annual Chinese Consumer Study," McKinsey Company, New York.
[35] Woodruff, R. B., 1997, "Customer Value: The next Source for Competitive Advantage," J. Acad. Mark. Sci., **25**(2), p. 139.
[36] Flint, D. J., and Woodruff, R. B., 2001, "The Initiators of Changes in Customers' Desired Value: Results from a Theory Building Study," Ind. Mark. Manag., **30**(4), pp. 321–337.
[37] Smith, K. T., and Brower, T. R., 2012, "Longitudinal Study of Green Marketing Strategies That Influence Millennials," J. Strateg. Mark., **20**(6), pp. 535–551.
[38] Lim, S., and Tucker, C. S., 2016, "A Bayesian Sampling Method for Product Feature Extraction From Large-Scale Textual Data," J. Mech. Des., **138**(6), p. 61403.
[39] Tucker, C. S., and Kim, H. M., 2011, "Trend Mining for Predictive Product Design," J. Mech. Des., **133**(11), p. 111008.
[40] Tucker, C., and Kim, H., 2011, "Predicting Emerging Product Design Trend by Mining Publicly Available Customer Review Data," *DS 68-6: Proceedings of the 18th International Conference on Engineering Design (ICED 11), Impacting Society through Engineering Design, Vol. 6: Design Information and Knowledge, Lyngby/Copenhagen, Denmark, 15.-19.08. 2011*.
[41] Anselin, L., 2013, *Spatial Econometrics: Methods and Models*, Springer Science & Business Media.
[42] Baltagi, B., 2008, *Econometric Analysis of Panel Data*, John Wiley & Sons.
[43] Rey, S. J., and Montouri, B. D., 1999, "US Regional Income Convergence: A Spatial Econometric Perspective," Reg. Stud., **33**(2), pp. 143–156.
[44] Badinger, H., Müller, W., and Tondl, G., 2004, "Regional Convergence in the European Union, 1985-1999: A Spatial Dynamic Panel Analysis," Reg. Stud., **38**(3), pp. 241–253.
[45] Elhorst, J. P., 2014, *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer: Berlin New York Dordrecht London.
[46] Cressie, N., and Wikle, C. K., 2015, *Statistics for Spatio-Temporal Data*, John Wiley & Sons.
[47] Sahu, S. K., and Bakar, K. S., 2012, "A Comparison of Bayesian Models for Daily Ozone Concentration Levels,"

Stat. Methodol., **9**(1–2), pp. 144–157.

[48] Munnell, A. H., and others, 1990, “Why Has Productivity Growth Declined? Productivity and Public Investment,” New Engl. Econ. Rev., (Jan), pp. 3–22.

[49] Baltagi, B. H., Bresson, G., and Pirotte, A., 2012, “Forecasting with Spatial Panel Data,” Comput. Stat. Data Anal., **56**(11), pp. 3381–3397.

[50] Elhorst, J. P., 2014, “Spatial Panel Data Models,” *Spatial Econometrics*, Springer, pp. 37–93.

[51] Parraguez, P., Maier, A., and Parraguez Ruiz, Pedro; maier, A., 2017, “Data-Driven Engineering Design Research: Opportunities Using Open Data,” *The 21st International Conference on Engineering Design (ICED 17) Vol 7: Design Theory and Research Methodology, Vancouver, Canada, 21-25.08. 2017*, pp. 41–50.

[52] Baltagi, B. H., Song, S. H., and Koh, W., 2003, “Testing Panel Data Regression Models with Spatial Error Correlation,” J. Econom., **117**(1), pp. 123–150.

[53] Hausman, J. A., 1978, “Specification Tests in Econometrics,” Econom. J. Econom. Soc., pp. 1251–1271.

[54] Elhorst, P., and Vega, S. H., 2013, “On Spatial Econometric Models, Spillover Effects, and W.”

[55] Millo, G., Piras, G., and others, 2012, “Splm: Spatial Panel Data Models in R,” J. Stat. Softw., **47**(1), pp. 1–38.

[56] Bloomberg, 2016, “China Auto Sales Growth Accelerates on Rising SUV Demand” [Online]. Available: https://www.bloomberg.com/news/articles/2016-07-08/china-auto-sales-grow-at-faster-pace-on-suv-electric-car-demand.

[57] “National Bureau of Statistics of China” [Online]. Available: http://www.stats.gov.cn/english/.

[58] Huo, H., Wang, M., Johnson, L., and He, D., 2007, “Projection of Chinese Motor Vehicle Growth, Oil Demand, and CO2 Emissions through 2050,” Transp. Res. Rec. J. Transp. Res. Board, (2038), pp. 69–77.

[59] Huo, H., and Wang, M., 2012, “Modeling Future Vehicle Sales and Stock in China,” Energy Policy, **43**, pp. 17–29.

[60] Bloomberg News, “Chinese Automakers Surge After Government Cuts Purchase Tax” [Online]. Available: https://www.bloomberg.com/news/articles/2015-09-29/china-reduces-small-car-purchase-tax-by-half-as-deliveries-slow.