ASME Journal of Computing and Information Science in Engineering
Online journal at:
https://asmedigitalcollection.asme.org/computingengineering

Modeling Spatiotemporal Heterogeneity of Customer Preferences With Small-Scale Aggregated Data: A Spatial Panel Modeling Approach

Yuyang Chen
University of Michigan—Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University,
800 Dongchuan Road, Minhang District,
Shanghai 200240, China
e-mail: chen-yu-yang@sjtu.edu.cn

Youyi Bi[1]
University of Michigan—Shanghai Jiao Tong University Joint Institute,
Shanghai Jiao Tong University,
800 Dongchuan Road, Minhang District,
Shanghai 200240, China
e-mail: youyi.bi@sjtu.edu.cn

Jian Xie
School of Mechanical Engineering,
Beijing Institute of Technology,
5 South Zhonguancun St., Haidian District,
Beijing 100081, China
e-mail: xiejian@bit.edu.cn

Zhenghui Sha
Walker Department of Mechanical Engineering,
The University of Texas at Austin,
204 E. Dean Keeton Street,
Austin, TX 78712
e-mail: zsha@austin.utexas.edu

Mingxian Wang
Integrated Design Automation Laboratory,
Northwestern University,
2145 Sheridan Road, Tech A216,
Evanston, IL 60208
e-mail: mingxianwang2016@u.northwestern.edu

Yan Fu
Global Data Insight & Analytics,
Ford Motor Company,
1 American Road,
Dearborn, MI 48126
e-mail: yfu4@ford.com

Wei Chen
Integrated Design Automation Laboratory,
Northwestern University,
2145 Sheridan Road, Tech A216,
Evanston, IL 60208
e-mail: weichen@northwestern.edu

Customer preferences are found to evolve over time and correlate with geographical locations. Studying the spatiotemporal heterogeneity of customer preferences is crucial to engineering design as it provides a dynamic perspective for understanding the trend of customer preferences. However, existing choice models for demand modeling do not take the spatiotemporal heterogeneity of customer preferences into consideration. Learning-based spatiotemporal data modeling methods usually require large-scale datasets for model training, which are not applicable to small aggregated data, such as the sale records of a product in several regions and years. To fill this research gap, we propose a spatial panel modeling approach to investigate the spatiotemporal heterogeneity of customer preferences. Product and regional attributes varying in time are included as model inputs to support demand forecasting in engineering design. With case studies using the dataset of small SUVs and compact sedans in China's automotive market, we demonstrate that the spatial panel modeling approach outperforms other statistical spatiotemporal data models and non-parametric regression methods in goodness of fit and prediction accuracy. We also illustrate a potential design application of the proposed approach in a portfolio optimization of two vehicles from the same producer. While the spatial panel modeling approach exists in econometrics, applying this approach to support engineering decisions by considering spatiotemporal heterogeneity and introducing engineering attributes in demand forecasting is the contribution of this work. Our paper is focused on presenting the approach rather than the results per se.
[DOI: 10.1115/1.4065211]

Keywords: spatiotemporal heterogeneity, customer preference, small dataset, spatial panel model, demand forecasting, data-driven engineering

1 Introduction

Customer preference models support product design in many aspects [1] as they can quantitatively characterize the interrelationship between market demand, engineering design attributes, and customer demographics. However, modeling customer preferences is inherently difficult due to the challenges in modeling various forms of heterogeneity in customer behaviors [2]. Different approaches have been used to capture the heterogeneity under the

[1]Corresponding author.
Manuscript received August 14, 2023; final manuscript received January 15, 2024; published online April 16, 2024. Assoc. Editor: Seung-Kyum Choi.

discrete choice analysis (DCA) framework [3]. Nevertheless, none of them takes the spatiotemporal[2] heterogeneity of customer preferences into consideration. Customer preferences are found to evolve over time and correlate with geographical locations [4]. Human behaviors at one location can diffuse to adjacent locations that share similar socioeconomic status [5]. A thorough understanding of the spatiotemporal heterogeneity of customer preferences can help designers create customized products and support companies to develop localized marketing strategies.

In engineering design, researchers expect to investigate the spatiotemporal heterogeneity of customer preferences to guide product design with various methods. For example, time-series analysis methods [6] have been used to extract product features and predict emerging product design trends from longitudinal online customer reviews and social media content. These methods can achieve high accuracy of prediction, but they may not provide enough insights into what factors influence these trends to support design decision making. Learning-based spatiotemporal data modeling methods such as long short-term memory networks [7], gated recurrent unit networks [8], and graph neural networks [9] are powerful in capturing non-linear features, but they usually require large-scale datasets for model training (e.g., more than 10,000 samples). Network-based methods [10] can analyze and model complex relationships based on a networked graph, where nodes represent individual customers or products and links represent relationships between them. They are suitable for modeling individual customer preferences when a large number of preference records are available and the interrelationships between customers are explicit. Thus, if only small-scale aggregated customer preference data are available, the above-mentioned methods are not applicable. Here, small-scale aggregated data refer to those customer preference data collected in an aggregated way (i.e., a group of customers' preferences are collected, such as customers from a residential community rather than individuals) with limited size (often less than a thousand samples) [11]. For example, the sale record of a product in several regions and years is a typical kind of small-scale aggregated data, and it is difficult to find explicit interrelationships between regions like individual customers. Therefore, an approach for modeling spatiotemporal heterogeneity of customer preferences with small-scale aggregated data is needed.

In this study, we propose to employ spatial panel models [12] for analyzing and understanding the spatiotemporal heterogeneity of customer preferences in support of engineering design by considering the impact of geographical, social, and economic factors from different regions in addition to those traditionally considered product design attributes. Rooted in spatial econometrics and regional science [13], spatial panel models are effective for modeling correlations between dependent variables and independent variables in both space and time [14]. Although this method has been applied in research on transportation mode choice modeling [15], this is the first attempt to employ spatial panel models to support engineering design.

The main contribution of our work is the introduction of spatial panel models into 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. In addition to our previous work [16], we examined the effectiveness of this approach by thoroughly comparing it with other statistical spatiotemporal data models and non-parametric regression methods in case studies. Significant spatiotemporal effects of customer preferences are captured in the case study, and the better prediction accuracy of our approach also supports the solving of a portfolio design optimization problem. These results and insights can provide other researchers in the engineering design community with more thoughts or inspirations about how to model the influence of spatial factors, temporal factors, and their interactions on customer choice/product demand. Note that as regression models, although spatial panel models can explicitly model the correlations between dependent variables and independent variables when the size of the dataset is too small, these models are likely to encounter the problem of underfitting. Also, non-linear relationships between the variables cannot be captured by linear regression models.

2 Review of Related Methods

Spatiotemporal data are data that relate to both space and time and can be categorized into spatial panel data and point-referenced data. The former refers to a cross section of observations on a set of spatial units (e.g., cities, states, provinces) repeated over several time periods, such as the recent 5-year population of the 48 contiguous states in the United States. The latter is characterized by a specific spatial position (e.g., latitude and longitude) and a timestamp, such as the meteorological data recorded at different monitoring stations.

Researchers have developed modeling approaches for different types of spatiotemporal data. A typical class of models for spatial panel data is the spatial panel model. Its theoretical foundation originates from three basic spatial interaction effects [14]. Endogenous interaction effects measure how the dependent variable of one spatial unit is jointly determined with that of neighboring units. Exogenous interaction effects measure how the dependent variable of a particular unit depends on the independent variables of other units. Interaction effects among the error terms measure how the unobserved factors in neighboring spatial units influence each other. The spatial panel model has been used in the analysis of economic activities, such as bicycle sharing demand modeling [17]. With limited parameters to estimate, the spatial panel model is not difficult to implement. It typically requires a medium amount of data (e.g., a thousand samples) for model training, and the interpretability of its results is good.

Panel vector autoregression (PVAR) is a multivariate time-series model used in panel data analysis. This model combines the vector autoregression (VAR) model [18] with panel data. PVAR has been utilized in capturing the spatiotemporal patterns of panel data. For example, Xing and Ye [19] employed PVAR to model low-carbon green transition, consumption upgrading, and industrial structure change. Compared to spatial panel models, PVAR is slightly more difficult to implement. PVAR needs more data for model training since it includes lagged terms in regression. The results of PVAR models are interpretable.

Unlike spatial panel models and PVAR, geographically weighted regression (GWR) is a method for point-referenced spatial data, which has been used in spatial analysis of transportation, such as modeling urban travel demand [20]. GWR adopts a local strategy to fit regression models at each geographic location based on its neighbors within a specific bandwidth. GWR can work with medium- to large-scale datasets (e.g., a thousand samples or more), and its implementation is relatively easy. However, the results of the GWR model are more difficult to interpret because they involve a large number of coefficients to describe the localized relationships, which requires the comparison and interpretation of results in each spatial unit.

Geographically and temporally weighted regression (GTWR) is an extension of GWR by incorporating temporal effects, which can model both spatial and temporal heterogeneity. Shen et al. [21] used GTWR to investigate the spatiotemporal influence of land use and household properties on the demand for travel with cars. Compared to GWR, GTWR is more complex as it accounts for local effects in both space and time. GTWR can work with medium- to large-scale datasets.

Compared to the parametric regression methods reviewed above, Gaussian process regression (GPR) is a typical non-parametric regression method, as it does not rely on a fixed mathematical

[2]The word "spatial" is a terminology extensively used in spatial economics and regional science. It has a broader meaning than "geographic" as it can also describe regions based on their social or economic relations beyond geographic borders.

expression. Based on the Bayesian probability theory, GPR can model and predict the relationship between independent and dependent variables by using Gaussian processes to establish the probability distribution function between these variables. In the field of customer analysis, GPR has been applied to predict the needs of customers. For instance, Sun and Lu [22] applied GPR to model bike-sharing demand, considering the influence of different land use types. Compared to the parametric methods, GPR has fewer requirements on the dataset, and it can provide estimates of the uncertainty of predictions. However, GPR is difficult to implement since it requires finding the optimal combination of hyperparameters. The results of GPR are also less interpretable.

Based on the above reviews, we expect that the spatial panel model will work better with small-scale data, and the obtained results possess better interpretability.

3 A Spatial Panel Modeling Approach for Modeling Heterogeneity of Customer Preferences

The proposed approach aims to extract insights from small-scale spatiotemporal data—how aggregated customer preferences change with space and time. Figure 1 illustrates a step-by-step procedure for implementing the approach. In Step 1, data reflecting the spatiotemporal heterogeneity of customer preferences (e.g., sale records of a product in multiple regions and years) and potential influencing factors (e.g., product design attributes, customer demographics) are collected and preprocessed. Descriptive analysis and visualization of these data can help with selecting appropriate explanatory variables in the subsequent modeling process.

In Step 2, spatial dependence tests (e.g., Lagrange multiplier (LM) test [23], Hausman test [24]) are used to examine the spatial dependence of collected data and provide clues for model specification (e.g., choose between a spatial panel model or a linear regression model, choose between a random effects model and a fixed effects model). Specifically, the LM test is utilized to examine the presence of spatial dependence within the data. If the null hypothesis can be rejected, the spatial dependence exists, and the spatial panel model can be chosen. Otherwise, the spatial panel model does not need to be considered since no spatial dependence is detected. The Hausman test is employed to determine whether a fixed effects model or a random effects model is more suitable for the panel data. If the null hypothesis is rejected, the fixed effects model is expected to perform better.

In Step 3, after obtaining the spatial dependence test results, the modeling variables and model types can be specified. Typical response variables ($\boldsymbol{Y}$) could be the demand, sales, or subjective rating of a specific product (e.g., VW Jetta) or product segment (e.g., small SUV). The spatial weights matrix ($\boldsymbol{W}$) is used to describe the geographic, demographic, or socioeconomic distance between spatial units. Explanatory variables ($\boldsymbol{X}$) can include customer demographics, regional characteristics, and product attributes. Commonly used statistical analysis techniques for identifying key attributes in customer preference modeling include multicollinearity analysis, stepwise logistic regression, and principal component analysis, as shown in Ref. [25]. Usually, multicollinearity analysis can be first used to identify the correlations between explanatory variables. By excluding those variables with strong correlations, the regression results can be more reliable and accurate. Then, principal component analysis is useful for extracting the most influencing variables when a large number of explanatory variables are available. Stepwise logistic regression can be used to support the fine adjustment of explanatory variables by iteratively adding or removing certain variables to the regression model based on specific criteria, such as the Akaike information criterion (AIC) or Bayesian information criterion (BIC).

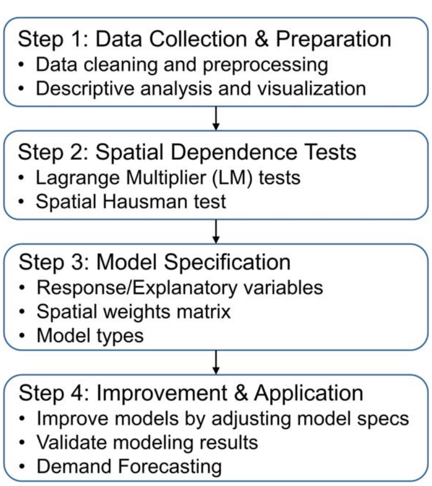

Fig. 1 A step-by-step procedure for understanding spatiotemporal heterogeneity of customer preferences in engineering design

$$\boldsymbol{Y_t} = \overbrace{\boxed{\lambda \boldsymbol{W} \boldsymbol{Y_t}}}^{\text{Endogenous interaction effects}} + \overbrace{\boxed{\boldsymbol{X_t}\boldsymbol{\beta}}}^{\text{Explanatory variables and coefficients}} + \underbrace{\boxed{\boldsymbol{W}\boldsymbol{X_t}\boldsymbol{\theta}}}_{\text{Exogenous interaction effects}} + \underbrace{\boxed{\boldsymbol{\mu}}}_{\text{Spatial specific effects}} + \overbrace{\boxed{\boldsymbol{\xi}}}^{\text{Time-period specific effects}} + \boldsymbol{u_t} \tag{1}$$

$$\boldsymbol{u_t} = \underbrace{\boxed{\rho \boldsymbol{W} \boldsymbol{u_t}}}_{\text{Interaction effects among error terms}} + \boldsymbol{\varepsilon_t} \tag{2}$$

Fig. 2 The full static spatial panel model

A comprehensive static spatial panel model can be represented in Eqs. (1) and (2) in Fig. 2. In practice, depending on the application context, simplified versions of this model are more often used. For example, the spatial autoregressive model (SAR, see Eq. (3)) only considers the endogenous interaction effect among the dependent variable by setting θ and ρ to zero, while the spatial error model (SEM, see Eqs. (4) and (5)) only considers the interaction effect between error terms by setting $\boldsymbol{\theta}$ and λ to zero. When none of the spatial dependence and space/time-specific effects are considered according to the testing results in Step 2, $\boldsymbol{\theta}$, λ, ρ, $\boldsymbol{\mu}$, and $\boldsymbol{\xi}$ will all be set to zero, and then a spatial panel model simply degenerates into a linear regression model. After specifying the model types and variables, the associated coefficients can be estimated by the maximum likelihood or generalized moments [26], in which the former method is more often used due to its better applicability for different forms of spatial panel models.

$$\boldsymbol{Y_t} = \lambda \boldsymbol{W} \boldsymbol{Y_t} + \boldsymbol{X_t}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\xi} + \boldsymbol{\varepsilon_t} \tag{3}$$

$$\boldsymbol{Y_t} = \boldsymbol{X_t}\boldsymbol{\beta} + \boldsymbol{\mu} + \boldsymbol{\xi} + \boldsymbol{u_t} \tag{4}$$

$$\boldsymbol{u_t} = \rho \boldsymbol{W} \boldsymbol{u_t} + \boldsymbol{\varepsilon_t} \tag{5}$$

There are some other variations of spatial panel models, such as the spatial Durbin model, spatial autoregressive confused (SAC) model, and spatial lag of X (SLX) model [27]. In this study, we investigate the use of SAR and SEM since they are the most fundamental models to characterize spatial dependence, which have been used to investigate the spatial effects of economic and social activities. For example, Zhang et al. [28] employed SEM for the spatial analysis in mass appraisal of commercial real estate. Qu and Lee [29] demonstrated the use of SAR in the spatial dependence analysis of adjacent school districts. However, few previous studies introduce these models in modeling the spatiotemporal heterogeneity of customer preferences. Thus, our work is a good demonstration of the proposed approach in this field.

In Step 4, the model specifications may need to be adjusted iteratively until the models obtain sufficient explanatory power. The

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

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

Note: Standard deviations are in parentheses, and BHP stands for Brake Horsepower. * represents the significance of difference in comparing the mean values of 2012 and 2016 using a Welch two-sample t test ($\alpha = 0.05$): $p < .10$; $*p < .05$; $**p < .01$; $***p < 0.001$.

explanatory power is commonly measured by R^2, which quantifies the percentage of variance in the dependent variable that is predictable from the independent variables. Based on this metric, we can refine the model by adding or removing explanatory variables and choosing between different models (e.g., SAR or SEM) until a model with better explanatory power is obtained. A typical application of the built models is demand forecasting in engineering design, especially for investigating how product design change can influence its market demand. Equations (6) and (7) are used to make predictions on spatial panel data for fixed effects SAR model and SEM model, respectively, according to Baltagi et al. [30] and Elhorst's work [31]. Here, $T + C$ is a future time-period point, $\hat{\lambda}$ is the estimated spatial autoregression coefficient, $\hat{\beta}_{GMM}$ is the estimated coefficients using the generalized method of moments (GMM), and $\hat{\mu}$ is the estimated spatial-specific effect.

$$\hat{y}_{T+C} = (I_N - \hat{\lambda} W)^{-1}(X_{T+C}\hat{\beta}_{GMM} + \hat{\mu}) \tag{6}$$

$$\hat{y}_{T+C} = X_{T+C}\hat{\beta}_{GMM} + \hat{\mu} \tag{7}$$

Although the above-mentioned methods are not new, as they originate from various areas such as spatial econometrics and regional science, our work contributes to introducing, combining, and tailoring these methods into the engineering design field in a systematic way. It offers a comprehensive workflow consisting of data preparation, spatial effect detection, and model selection, training, and prediction for modeling spatiotemporal heterogeneity of customer preferences with small-scale aggregated data.

4 A Case Study of Passenger Vehicles

To demonstrate the proposed approach, we present a case study employing the data from a recognized, reputable survey representing China's automotive market [32]. This survey data consists of about 50,000 new car buyers' responses and purchase history covering about 400 different vehicle models in China's market each year from 2012 to 2016. Respondents were asked to list the cars they purchased with their residential information and the purchase time. 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 the small SUV segment (i.e., mini and compact SUVs), as the demand for small SUVs has been rising rapidly and increasingly, and affluent Chinese buyers opt for more spacious vehicles [33]. According to the survey data, 14.8% of the respondents purchased a small SUV in 2012, and this percentage increased to 21.6% in 2016. Our analysis is performed by grouping data samples in a given time interval, called *wave*. In each year, the survey was collected every four months; thus, we have three-wave data for each year and 15-wave data for 5 years. We considered 27 provinces in mainland China as the basic spatial units (i.e., regions) in this study, as these provinces have complete 15-wave data of small SUV sales. In addition to the survey data, we also collected regional statistics (demographics and socioeconomics) of these provinces from the National Bureau of Statistics of China [34].

4.1 Descriptive Analysis of the Key Variables.

Based on our prior research on customer preferences in vehicle consideration and choice [32], we identified three vehicle attributes (*price*, *power*, and *fuel consumption*), one customer attribute (*monthly household income*) and chose two regional attributes that have been broadly studied in automotive market research [35] (*GDP per capita* and *length of paved roads per capita*) to study the engineering, demographic and regional effects on small SUV sales. The vehicle attributes and customer attributes are taken from the survey [32], while the regional attributes are obtained from Ref. [34]. Note that our modeling results could be biased by not including the omitted variables that are critical for modeling customer preferences, which is a research topic in itself [36]. The main purpose of this work is to demonstrate the approach of integrating spatiotemporal heterogeneity into demand modeling rather than the results per se.

Table 1 provides the descriptive statistics of these attributes at the province level. The results indicate that the mean values on *sales of small SUVs*, *GDP per capita*, and *length of paved roads per capita* of each province in 2016 are larger than those in 2012, while the mean values on *price*, *power*, *fuel consumption*, and *household income* in 2016 are smaller than those in 2012. These results may imply that small SUVs are increasingly popular as they become more affordable, and the small SUVs offered in China's auto market tend to have lower power and fuel consumption to match the decreased prices. The decreased monthly household income of customer profiles implies that more customers with relatively lower incomes entered the market of small SUVs. The increased GDP per capita and the length of paved roads per capita indicate the growing economy and improved infrastructures in China in those 5 years.

Figure 3 presents the spatial distributions of selected variables in 2016. These graphs provide an intuitive reflection of the relationship between the sales of small SUVs and the selected explanatory variables. For example, the sales of small SUVs seem to positively correlate with income and GDP per capita and negatively correlate with price. We expect to obtain consistent results but in a quantitative way from the spatial panel models.

4.2 Spatial Dependence Tests.

Spatial dependence tests are used to examine the existence of spatial effects and provide clues for identifying 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 Ref. [23]. As shown in Table 2, the null hypotheses of both tests are rejected at the 5% level of significance. The result of the LM_H test suggests that at least one of the spatial autoregressive coefficient (λ) and the variance of spatial-specific effects (σ_μ^2) is not zero. It implies the existence of the spatial effect of the dependent variable, i.e., the sales of small SUVs in China. The result of the Hausman test indicates that the assumption of

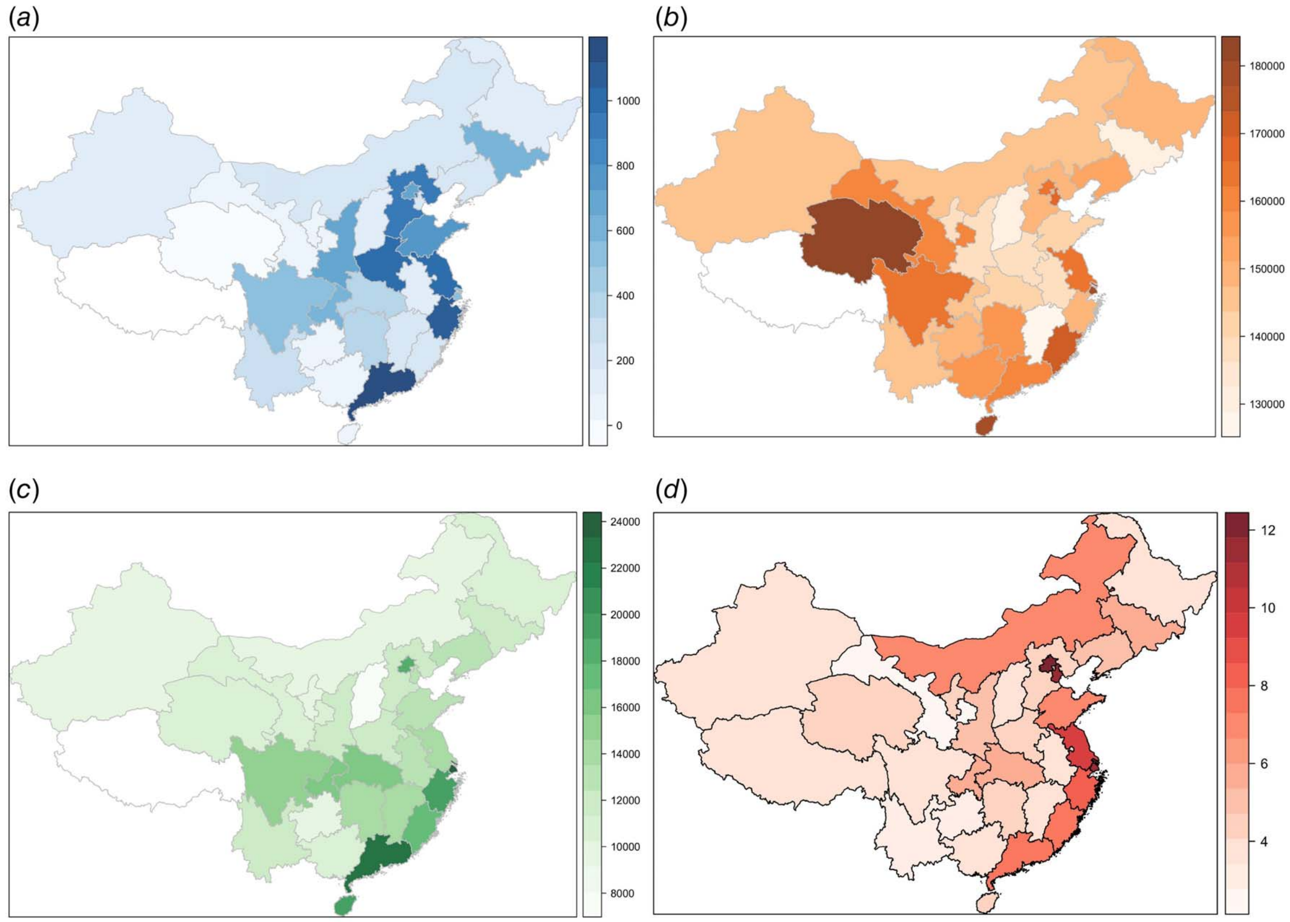

Fig. 3 The spatial distribution of certain model variables in 2016: (*a*) sale of small SUVs, (*b*) price, (*c*) monthly household income, and (*d*) GDP per capita (white areas represent missing data)

random effects is not supported by the data, and fixed effects models should perform better than random effects models.

4.3 Model Specifications. Upon the completion of the spatial dependence tests, spatial panel models can be built with detailed specifications as follows:

- Response variable (***Y***): small SUV sales in 27 provinces of China in each wave from 2012 to 2016. Due to the high correlation ($r = 0.931$) between the number of the surveyed respondents and the number of new vehicle registrations over multiple years in each province of China, the survey data were used as the surrogate of the actual sales numbers in regression analysis.
- Spatial weights matrix (***W***): binary geographical adjacency matrix based on purely geographical considerations. For ease of interpretation, W is normalized such that the elements of each row sum to unity [14].
- Explanatory variables (***X***): the three vehicle attributes, one customer attribute, and two regional attributes presented in Sec. 4.1.

4.4 Settings of Other Comparative Models. To demonstrate the effectiveness of the proposed approach, we build four other spatiotemporal data models as a comparison. Equation (8) shows the PVAR model:

$$\boldsymbol{Y}_{i,t} = \boldsymbol{\Phi}_{i,t-1}\boldsymbol{Y}_{i,t-1} + \boldsymbol{\Psi}_{i,t-1}\boldsymbol{X}_{i,t-1} + \epsilon_{i,t} \tag{8}$$

here, $\boldsymbol{Y}_{i,t}$ and $\boldsymbol{Y}_{i,t-1}$ are the ith observation on the dependent variable at t/ t -1 moment. $\boldsymbol{X}_{i,t-1}$ is the ith observation on the independent variable at $t-1$ moment. The coefficient matrices to be estimated, denoted as $\boldsymbol{\Phi}_{i,t-1}$ and $\boldsymbol{\Psi}_{i,t-1}$, are both $N \times N$ matrices. Based on the limited size of the available dataset and the obtained AIC results, we only consider the impact of the past period's data on the current period's data (i.e., the lag order is set to 1), ignoring the fixed effects of individual spatial units. Unit root tests [37] are then performed to examine the data stationarity, and Granger causality tests [38] are utilized to assess whether explanatory variables have a causal impact on response variables. Subsequently, after fitting the model, stability and co-integration tests are carried out to evaluate the effectiveness of the model.

Table 2 Results of spatial dependence tests

Test	H_0	Statistic	p value
LM_H	Spatial autoregressive coefficient (λ) and variance of spatial-specific effects (σ_μ^2) are both zero	747.80	<0.001
Hausman test	Random effects assumption is supported by the data	20.61	<0.001

Table 3 Estimated coefficients, goodness of fitness, and prediction accuracy of 10 different models in small SUV study

	Linear model	Linear model with fixed effects	Random effects spatial panel models		Fixed effects spatial panel models		PVAR	GWR	GTWR	GPR
			SAR	SEM	SAR	SEM				
λ			0.21***		0.18**					
ρ				0.14*		0.11.				
Price (β_1)	−2.53***	−0.86**	−0.91**	−1.18***	−0.59*	−0.75*	−1.01	−0.49	−1.59	\
Power (β_2)	−0.22	1.21.	1.12.	1.39*	0.89	1.08.	1.29	0.78	2.01	\
Fuel Consump. (β_3)	−0.41	−1.48**	−1.69***	−2.05***	−1.26*	−1.49**	−1.18	0.11	−2.51	\
House. Income (β_4)	1.20***	0.19	0.21	0.2	0.16	0.16	0.27	0.31	0.67	\
Length_Roads (β_5)	−0.33*	2.14***	0.55	0.41	1.91***	1.94***	0.21	1.22	0.59	\
GDP per capita (β_6)	1.31*	1.89***	1.25***	1.63***	1.57***	2.03***	0.33	0.73	0.94	\
R^2	0.52	0.56	−0.30	0.35	0.84	0.83	0.79	0.84	0.71	0.62
RMSE	5.60	0.47	4.07	0.85	0.47	0.48	2.16	2.33	2.52	2.41

Note: λ: spatial autoregressive coefficient; ρ, spatial autocorrelation coefficient; RSME, root-mean-square error. Since GPR is a non-parametric regression model, there are no estimated coefficients. $p < .10$; $*p < .05$; $**p < .01$; $***p < 0.001$.

Equation (9) shows the GWR model:

$$y_i = \beta_{i_0}(u_i, v_i) + \Sigma_{k=1}^{p-1} \beta_{i_k}(u_i, v_i) x_{ik} + \varepsilon_i \tag{9}$$

here, y_i is the dependent variable, $\beta_{i_0}(u_i, v_i)$ is the constant term, and $\beta_{i_k}(u_i, v_i)$ is the regression coefficient. x_{ik} is the independent variable, and p is the total number of coefficients to be estimated. When applying this model, we first obtain the latitude and longitude of the capital city of each province as their spatial coordinates. We include time as an additional explanatory variable in the model since GWR itself does not consider the time effects. We then select the optimal bandwidth based on the AIC and use the Gaussian function to determine the weight. For the GTWR model (see Eq. (10)), its modeling process is similar to GWR since it is an extension of GWR by adding time effects.

$$y_i = \beta_{i_0}(u_i, v_i, t_i) + \Sigma_{k=1}^{p-1} \beta_{i_k}(u_i, v_i, t_i) x_{ik} + \varepsilon_i \tag{10}$$

The GPR model assumes that the output y follows a Gaussian distribution, i.e., $y \sim \mathcal{N}(\mu(\boldsymbol{x}), \; k(\boldsymbol{x}, \boldsymbol{x}'))$, where $\mu(\boldsymbol{x})$ is the mean function, and $k(\boldsymbol{x}, \boldsymbol{x}')$ is the covariance function. We assume a prior mean of 0, and the prior variance is specified by the kernel function. In this study, we evaluate the impact of various kernels on the model fit (R-squared values) and prediction performance (root-mean-square error (RMSE)) of GPR. Our results show that the RMSE and R-squared of the GPR model with the radial basis function (RBF) kernel are 24.87 and −27.25, 11.48 and −5.02 with the Matérn kernel, and 3.06 and 0.57 with the rational quadratic kernel. The White kernel, on its own, exhibited an RMSE of 24.86 and an R-squared of −27.25. When combined with the DotProduct kernel, the performance was significantly improved, showing a reduced RMSE of 2.41 and an increased R-squared of 0.62. Thus, we ultimately select the best-performing kernel function, i.e., the sum of DotProduct and WhiteKernel. We then optimize the hyperparameters, such as the σ_0 (i.e., Gaussian noise variance, used to describe the noise level in GPR) of Dot product and noise_level of WhiteKernel, by maximizing the log marginal likelihood (LML).

5 Results

5.1 Estimated Spatial Parameters and Coefficients of Explanatory Variables. Table 3 presents the summary of the estimated coefficients, goodness of fit, and prediction accuracy of four different spatial panel models, two linear regression models, and PVAR, GWR, GTWR, and GPR models as comparisons. The experiment has been repeated multiple times, and the averaged results are reported. Here, the response variable and explanatory variables are transformed with natural logarithms, which is a common treatment in spatial regression analysis for smoothing data and getting results with practical meanings [39]. Depending on whether the spatial-specific effects (μ) are treated as random effects or fixed effects in the estimation, the random effects model and the fixed effect model are achieved, respectively. In SAR models (Eq. (3)), the dependence of the response variables in different regions is estimated and denoted by the spatial autoregressive coefficient (λ). In SEM models (Eqs. (4) and (5)), the dependence of the error terms in different regions is estimated instead and denoted by the spatial autocorrelation coefficient (ρ). 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. 5.2). The linear regression model is the simplest model with neither spatial dependence nor space/time-specific effects. A linear regression model with fixed effects is an extension to the linear model by adding spatial ($\boldsymbol{\mu}$) or temporal ($\boldsymbol{\xi}$) specific effects but no spatial dependence effect (here, we report the results of the linear model with spatial-specific effects, which has the best goodness of fit).

From Table 3, we can observe that the R^2 of these models suggests that fixed effects models ($R^2 = 0.84$, 0.83 for SAR and SEM, respectively) perform better than random effects models ($R^2 = -0.30$, 0.35 for SAR and SEM, respectively).[3] This result is consistent with the implication of model performance obtained from the spatial dependence test in Sec. 4.2. In addition, the R^2 of the linear regression model and the linear regression model with fixed effects are lower ($R^2 = 0.52$, 0.56), which indicates that the lack of the spatial dependence effects (e.g., λ, ρ) seems to weaken the model's goodness of fit. Furthermore, when using the first 14-wave data to train the model and the last wave data for testing, the RMSEs of fixed effects SAR model, SEM model, and linear model with fixed effects are 0.47, 0.48, and 0.47, respectively, while the RMSE of linear regression model without any spatiotemporal effects is 5.60. Thus, among the models that exhibited a positive R^2, the linear regression model is the least accurate one in predicting the response variable due to its lack of consideration of spatial dependence effects and spatial/temporal-specific effects.

The four columns on the right side of Table 3 show the estimation results of other three parametric spatiotemporal data models (PVAR, GWR, and GTWR) and one non-parametric model (GPR). As we can see, the GWR and PVAR models have better performance in the goodness of fit ($R^2 = 0.84$, 0.79 for GWR and PVAR, respectively) compared to the GTWR ($R^2 = 0.71$) and GPR ($R^2 = 0.62$) models. When using the first 14-wave data to train the models and the last wave data for testing, the RMSE of

[3] R^2 can be negative if the regression result is even worse than using the mean value of the data samples as the predictions.

the PVAR, GWR, GTWR, and GPR models are 2.16, 2.33, 2.52, and 2.41, respectively. These values are relatively close to each other, but all of them are much larger than the RMSE of spatial panel models, reflecting lower prediction accuracies. In addition, none of the estimated coefficients in these models are significant, which provides limited interpretability to the influencing factors on the spatiotemporal heterogeneity of customer preferences. One possible reason is that since PVAR, GWR, and GTWR involve too many coefficients for estimation, such as the lagged terms in PVAR and the varying coefficients for every geographic location in GWR and GTWR, they may not work well in our context with a small training dataset. There is a risk of underfitting when the coefficients to estimate are relatively large compared to the available data samples, which may also lead to poor prediction performance. For GPR, its hyperparameters may be not sufficiently optimized due to the limited size of the dataset and noise or outliers in the data, which results in poor performance in prediction. Therefore, we only present and discuss the results and applications of spatial panel models in the following sections.

The positive spatial autoregressive coefficient ($\lambda = 0.18$, $p < .01$) obtained from the fixed effects SAR model (see Eq. (3)) suggests that a region with higher small SUV sales is likely adjacent to several regions with high small SUV sales. This result implies the effect of geographical proximity on product sales. The spatial autocorrelation coefficient ($\rho = 0.11$, $p > .05$) estimated from the fixed effect SEM model (see Eqs. (4) and (5)) implies that the unobservable factors (i.e., other explanatory variables not included in our models) in one region have insignificant correlations 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, the increases of *price* ($\beta_1 = -0.75$, $p < .05$) and *fuel consumption* ($\beta_3 = -1.49$, $p < .01$) tend to have a negative impact on sales of small SUVs. The effects of *power* ($\beta_2 = 1.08$, $p > .05$) and *monthly household income* of customers ($\beta_4 = 0.16$, $p > .05$) are not significant. Among the two regional attributes, *length of paved roads per capita* ($\beta_5 = 1.94$, $p < 0.001$) and *GDP per capita* ($\beta_6 = 2.03$, $p < 0.001$) both have significant positive influences on the sales of small SUVs. The estimated coefficients of *price* and *GDP per capita* in spatial panel models are consistent with the observations from the choropleth map shown in Fig. 3. In addition, the estimated coefficients in the fixed effects SAR model are similar to those in the fixed effects SEM model.

5.2 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 a particular region. Figure 4 presents the estimated spatial-specific effects obtained from the fixed effects SEM model (see Table 4 for the values). As shown in Fig. 4, the blue regions exhibit negative spatial-specific effects. This suggests that these regions have some unobserved factors that weakly influence their small SUV sales. This may be due to relatively lower levels of socioeconomic development in these provinces (Xinjiang ($\mu = -1.52$) and Ningxia ($\mu = -2.05$)). By contrast, economically developed areas such as Beijing ($\mu = 0.33$) and Shanghai ($\mu = 0.95$) have positive spatial-specific effects (red regions), which suggest that these regions have certain factors that strongly influence their respective small SUV sales. It is also interesting to see that Henan ($\mu = 1.55$) and Sichuan ($\mu = 1.35$) have the highest estimated spatial-specific effects, even higher than Beijing and Shanghai, although these two provinces are less economically developed. Possible factors contributing to this result include market saturation, economic growth rate, and market capacity. Compared to Beijing and Shanghai, the market of small SUVs is far away from saturation in provinces like Henan and Sichuan. The higher economic growth rates and larger populations in Henan and Sichuan may lead to a much greater demand for small SUVs. In addition, Henan and Sichuan contain a large number of rural areas that were experiencing increased wealth. For many customers there, a small SUV might be a practical choice for their first vehicle due to its large space, good passability, and affordable price. These results imply that when changing the attributes of a vehicle to the same extent, the influence on vehicle sales in different regions can be different. Traditional pricing strategies may not be effective in regions with negative space-specific effects, and car companies may want to devise more customized marketing strategies to attract customers with unique preferences in those regions.

Fig. 4 Spatial-specific effects obtained from the fixed effects SEM model (dark gray color represents missing of data)

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

Province (i)	μ_i	Province (i)	μ_i
Anhui	−0.49	Jilin	−0.01
Beijing	0.33	Liaoning	−0.14
Chongqing	0.39	Inner Mongolia	−1.36
Fujian	−0.10	Ningxia	−2.05
Gansu	−0.50	Shandong	0.16
Guangdong	0.83	Shanghai	0.95
Guangxi	−0.29	Shaanxi	1.00
Hebei	1.03	Shanxi	−0.11
Heilongjiang	−0.77	Sichuan	1.35
Henan	1.55	Tianjin	−1.04
Hubei	−0.16	Xinjiang	−1.52
Hunan	0.51	Yunnan	0.67
Jiangsu	0.01	Zhejiang	0.29
Jiangxi	−0.52	—	—

Time-period-specific effects control for all space-invariant variables contributing to the response variable (i.e., the sales), which reflects the regional characteristic in a particular time period. Table 5 presents the estimated time-period-specific effects obtained from the fixed effects SEM model (see Eqs. (4) and (5)). It is observed that most time-period-specific effects ($\boldsymbol{\xi}$) are *negative* before 2015 but become *all positive* afterward. This is probably due to the implementation of certain nationwide incentives starting from 2015, which greatly stimulated the sales of small SUVs in China. For example, China reduced the vehicle purchase tax from 10% to 5% for small passenger cars (engine displacement ≤ 1.6 liters) in 2015 [40].

5.3 Validation Test.

To further validate our approach, we conduct another case study by employing the survey data mentioned in Sec. 4 but focusing on a different market segment, compact sedan. The spatial panel modeling approach and other comparative methods are tested with the same data analysis and modeling procedures as described in Sec. 4. The tests have been repeated multiple times, and averaged results are summarized in Table 6. From Table 6, we can find that the fixed effect model of SAR still performs the best ($R^2 = 0.86$, $\text{RMSE} = 0.40$). Although

Table 5 Estimated time-period-specific effects (ξ)

Time period (t)	ξ_t	Time period (t)	ξ_t
2012-1	−0.43	2014-3	−0.03
2012-2	−0.34	2015-1	0.21
2012-3	−0.86	2015-2	0.33
2013-1	−0.45	2015-3	0.14
2013-2	−0.16	2016-1	0.62
2013-3	−0.06	2016-2	0.52
2014-1	0.02	2016-3	0.56
2014-2	−0.07	—	—

PVAR and GWR achieve similar goodness of fit ($R^2 = 0.86$), their prediction accuracy (RMSE = 2.03, 1.82) is much poorer. These results provide more evidence for the effectiveness of our approach.

6 Application in Vehicle Portfolio Design Optimization Problem

In this section, we demonstrate the application of the proposed approach to a hypothetical vehicle portfolio design optimization problem. Suppose an auto company has two car models from the same segment, how can the designer(s) optimally adjust a powertrain design attribute (e.g., 0–60 MPH acceleration time) of the two models such that their overall profit is maximized across all regions? Fig. 5 shows the formulation of this enterprise-driven (i.e., profit maximization) portfolio design problem. In the simulated scenario, and for the purpose of demonstration, the interaction between the sales of the two car models is ignored for simplicity. Here, we take VW FAW Sagitar and Bora as examples of two car models in the optimization because they have complete sales records in the 21 provinces of China from 2012 to 2016. By fitting a spatial error model (SEM) for the sales of Sagitar and Bora, respectively, the influence of three car attributes, one customer attribute, and two regional attributes aforementioned in Secs. 3 and 4, along with their associated spatial-specific effects, can be estimated. The demand functions are then constructed

Table 6 Estimated coefficients, goodness of fitness, and prediction accuracy of 10 different models in compact sedan study

	Linear model	Linear model with fixed effects	Random effects spatial panel models		Fixed effects spatial panel models		PVAR	GWR	GTWR	GPR
			SAR	SEM	SAR	SEM				
λ			0.37***		0.36**					
ρ				0.41***		0.39***				
Price (β_1)	−0.67	0.12	0.40	0.66.	0.41	0.67.	−0.18	−0.27	−0.28	\
Power (β_2)	−2.45	−2.43*	−2.41*	−1.84.	−2.68*	−2.15*	1.19	−2.40	−3.68	\
Fuel Consump. (β_3)	−3.23***	−1.33*	−1.39**	−1.91***	−1.07*	−1.55**	−0.23	−1.61	−2.71	\
House. Income (β_4)	1.36***	−0.06	0.07	0.11	−0.03	0.01	0.26	0.48	1.14	\
Length_Roads (β_5)	−0.25*	1.32**	0.28	−0.09	0.79.	0.44	−0.11	2.45	−0.32	\
GDP per capita (β_6)	1.15***	1.23**	0.87**	1.46***	0.81*	1.53***	0.33*	−1.32	1.24	\
R^2	0.43	0.26	−3.41	0.34	0.86	0.83	0.86	0.86	0.66	0.85
RMSE	2.96	0.42	4.82	0.87	0.40	0.43	2.03	1.82	3.18	2.11

Note: λ: spatial autoregressive coefficient; ρ: spatial autocorrelation coefficient; RSME: root-mean-square error. Since GPR is a non-parametric regression model, there are no estimated coefficients. $p < .10$; *$p < .05$; **$p < .01$; ***$p < 0.001$.

Given

1) Market data

Sales of VW FAW Sagitar (A) and Bora (B) in 21 provinces of China (2012 wave 2 to 2016 wave 2)

2) Demand function (Spatial Panel Model, SEM)

$$Q_i(x) = \left[P\ FC\ \ldots X_{GDP/capita}\right] \cdot \begin{bmatrix}\beta_1\\ \beta_2\\ \ldots\\ \beta_6\end{bmatrix} + \mu_i \qquad (11)$$

$$FC = \left(0.035 + \frac{53.5 + 69.5e^{-x} - 1.8x^{1.4} + \frac{106.9}{x}}{1000}\right) \cdot a \qquad (12)$$

Q is demand, i is the index of region, P is price, FC is fuel consumption, x is 0-60 MPH acceleration time (in seconds), μ_i is the spatial specific effect, a is a constant different between Sagitar and Bora.

3) Cost function

$$\mathrm{C}(x) = Q(x) \cdot v(x) + C_{fixed} \qquad (13)$$

$$\mathrm{v}(x) = \left(e^{\frac{x}{12}}\left(1.5 + 1.97e^{-x} - 0.04x + \frac{1}{x-1.5}\right)\right).b \qquad (14)$$

$Q(x) \cdot v(x)$ is the variable cost depending on 0-60 MPH acceleration time, b is a constant different between Sagitar and Bora. C_{fixed} is the fixed cost

Find

The optimal 0-60 MPH acceleration (x_A, x_B) for two cars

Maximize

Overall profit $\Pi = \Pi(x_A) + \Pi(x_B)$ (15)

where $\Pi(x) = \sum_{i=1}^{N}(Q_i(x) \times P_i(x)) - C(x)$

Subject to

Design constraints $x_A \in [5, 12.5]$, $x_B \in [7, 12.5]$

Fig. 5 Formulation for profit maximization-based design problem

Table 7 Optimization results

	0–60 MPH acceleration time (s)		Overall profit (million RMB)	
	Optimized result	Market data	Optimized result	Market data
Sagitar	9.18	9.70	31.51	31.42
Bora	10.68	11.70	109.35	109.28

using these estimated coefficients ($\beta_1, \beta_2, \ldots$) and the formula in Eq. (7). Note in Eq. (7), time fixed effects are not included, which is a common treatment in current spatial economics literature [31]. To make any assumptions on time fixed effects in predictions, one needs to forecast factors such as the national macro-economic situation, change of government regulations, market sentiment, etc., in a future period, which is beyond the scope of this tech brief. As shown in Eq. (11) of Fig. 5, the demand (Q) now is a linear function of price, fuel consumption, and other modeled attributes and is unique to each province due to the variances in regional attributes and spatial-specific effects (μ). Fuel consumption is a function of the design variable x (0–60 MPH acceleration time) based on the empirical engineering model used in Ref. [41]. To match with the real vehicle performance [42] as much as possible, we added a constant a to this function ($a = 110$ for Sagitar and 150 for Bora).

Equation (12) in Fig. 5 is a simulated cost function, including the fixed cost C_{fixed} (constant), and the variable cost $Q(x) \cdot v(x)$ depending on both demand and the design variable. We assume the fixed cost is 18% of the average revenue per wave for Sagitar (8.46 million RMB) and Bora (5.04 million RMB). As for the variable cost coefficient function $v(x)$, we adopted the cost model used in Ref. [41] and added a constant b to differentiate the cost of Sagitar ($b = 2.43$) and Bora ($b = 0.76$) and match with their real cost [42] as much as possible.

In this optimization problem, we consider price as a constant in predictions, which is equal to the average price in the predicted time period (i.e., 2016 wave 3). The purpose of this treatment is to examine how a lower-level engineering design variable (0–60 MPH acceleration time) can influence the sales and profit of two vehicles from the same segment considering regional differences by excluding the impact of price (the pricing strategy could be very complicated in a real market). In addition, to quantify the relationship between power and acceleration time, we fitted a linear model using the vehicle website data [42], and the resulting model is $\text{power} = 310 - 16.77 * \text{accerlation time}$ for Sagitar with a $R^2 = 0.877$. The same approach was applied to Bora.

Under these settings and assumptions, Table 7 provides the optimization results. It can be observed that the optimized 0–60 MPH acceleration time of Sagitar (9.18 s) and Bora (10.68 s) are smaller than the market data (9.70 s, 11.70 s) in 2016 wave 3, and the overall profit using the optimized design variable is larger than the profit calculated by using the market data [42]. These results demonstrate the need for considering spatiotemporal heterogeneity of customer preferences in product family design and the capability of our approach.

7 Conclusion

We introduce the spatial panel modeling approach into the engineering design field for capturing the spatiotemporal heterogeneity of customer preferences. Our approach provides a systematic and comprehensive workflow consisting of data preparation, spatial effect detection, and model selection, training, and prediction for modeling spatiotemporal heterogeneity of customer preferences with small-scale aggregated data. Our study shows that spatial panel models can quantify the influence of product attributes, customer demographics, and regional characteristics on aggregate customer choices. Their model fitting performance and prediction accuracy are found to outperform other parametric or non-parametric spatiotemporal data models, such as PVAR, GWR, GTWR, and GPR. The effectiveness of our approach is also demonstrated by its implementation in a portfolio design optimization problem.

Although the linear regression models may provide similar results on selected explanatory variables when modeling spatiotemporal heterogeneity of customer preferences, one critical advantage of spatial panel models is that they can model and reveal the spatial dependence between dependent/independent variables in various regions, which allows us to capture how certain customer preferences diffuse spatially. 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.

The knowledge and insights gained from our work also have implications for 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 high fuel consumption may reduce the sales of small SUVs; 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. In addition, the capability of predicting demand or market share across different spatial regions can directly assist localized product development and product family design over multiple market regions. It can also support the study of the spatial diffusion patterns of product designs, e.g., how the launch of a new design or a design improvement to existing products in one region influences customers' choice behaviors in neighboring regions.

Acknowledgment

The authors gratefully acknowledge the financial support from National Natural Science Foundation of China (52005328), NSF-CMMI-1436658, the Ford-Northwestern Alliance Project, and a grant from the China Scholarship Council (No. 201706030108).

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

Nomenclature

y = dependent variables observed in a spatial unit
x = independent variables observed in a spatial unit
K = number of explanatory variables
N = number of spatial units
T = number of time periods

W = A nonnegative $N \times N$ spatial weights matrix
X_t = $NT \times K$ matrix of exogenous explanatory variables
u_t = error term
Y_t = $NT \times 1$ vector consisting of T observations on the dependent variable for every space units in the sample
Wu_t = interaction effects among the disturbance terms of different units
WX_t = exogenous interaction effects term
WY_t = endogenous interaction effect term
λ = spatial autoregressive coefficient
ρ = spatial autocorrelation coefficient
ε = vector of disturbance term
μ = spatial-specific effects
ξ = time-period-specific effects
γ_i = individual-specific unobserved fixed effects matrix
β, θ = associated $K \times 1$ vectors with unknown parameters to be estimated
$\Phi_{i,k}$, $\Psi_{i,j}$ = $N \times N$ coefficient matrices
(u_i, v_i) = geographic center coordinates of the sample spatial unit
$\beta_{i_0}(u_i, v_i)$ = constant term estimate for the i th sample
$\beta_{i_k}(u_i, v_i)$ = regression coefficient for the k th independent variable of the i th sample
$Q(x)$ = demand, a linear function of price, fuel consumption and other modeled attributes
$v(x)$ = variable cost coefficient function
C_{fixed} = fixed cost
a, b = constants different between Sagitar and Bora

References

[1] Donndelinger, J. A., and Ferguson, S. M., 2020, "Design for the Marketing Mix: The Past, Present, and Future of Market-Driven Engineering Design," ASME J. Mech. Des., **142**(6), p. 060801.
[2] Jin, J., Liu, Y., Ji, P., and Kwong, C. K., 2019, "Review on Recent Advances in Information Mining From Big Consumer Opinion Data for Product Design," ASME J. Comput. Inf. Sci. Eng., **19**(1), p. 010801.
[3] Hoyle, C., Chen, W., Wang, N., and Koppelman, F. S., 2010, "Integrated Bayesian Hierarchical Choice Modeling to Capture Heterogeneous Consumer Preferences in Engineering Design," ASME J. Mech. Des., **132**(12), p. 121010.
[4] 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**, p. e23.
[5] Mosleh, M., and Heydari, B., 2016, "Spatial Diffusion of Risk: The Case of Risky Teenage Drivers," 5th International Engineering Systems Symposium, Washington, DC, June 27–29.
[6] Lim, S., and Tucker, C. S., 2016, "A Bayesian Sampling Method for Product Feature Extraction From Large-Scale Textual Data," ASME J. Mech. Des., **138**(6), p. 061403.
[7] Lee, H., Puranik, T. G., and Mavris, D. N., 2021, "Deep Spatio-Temporal Neural Networks for Risk Prediction and Decision Support in Aviation Operations," ASME J. Comput. Inf. Sci. Eng., **21**(4), p. 041013.
[8] Yu, T., and Wang, J., 2021, "A Spatiotemporal Convolutional Gated Recurrent Unit Network for Mean Wave Period Field Forecasting," J. Mar. Sci. Eng., **9**(4), p. 383.
[9] Lira, H., Martí, L., and Sanchez-Pi, N., 2022, "A Graph Neural Network With Spatio-Temporal Attention for Multi-Sources Time Series Data: An Application to Frost Forecast," Sensors, **22**(4), p. 1486.
[10] Bi, Y., Qiu, Y., Sha, Z., Wang, M., Fu, Y., Contractor, N., and Chen, W., 2021, "Modeling Multi-Year Customers' Considerations and Choices in China's Auto Market Using Two-Stage Bipartite Network Analysis," Networks Spat. Econ., **21**(2), pp. 365–385.
[11] Bonnell, T. R., Dutilleul, P., Chapman, C. A., Reyna-Hurtado, R., Hernández-Sarabia, R. U., and Sengupta, R., 2013, "Analysing Small-Scale Aggregation in Animal Visits in Space and Time: The ST-BBD Method," Anim. Behav., **85**(2), pp. 483–492.
[12] Anselin, L., 2013, *Spatial Econometrics: Methods and Models*, Springer Science & Business Media, Dordrecht, Netherlands.
[13] Baltagi, B., 2008, *Econometric Analysis of Panel Data*, John Wiley & Sons, Hoboken, NJ.
[14] Elhorst, J. P., 2014, *Spatial Econometrics: From Cross-Sectional Data to Spatial Panels*, Springer, Berlin, New York, Dordrecht, London.
[15] Paleti, R., Bhat, C. R., Pendyala, R. M., and Goulias, K. G., 2013, "Modeling of Household Vehicle Type Choice Accommodating Spatial Dependence Effects," Transp. Res. Rec., **2343**(1), pp. 86–94.
[16] Bi, Y., Xie, J., Sha, Z., Wang, M., Fu, Y., and Chen, W., 2018, "Modeling Spatiotemporal Heterogeneity of Customer Preferences in Engineering Design," Volume 2A: 44th Design Automation Conference, Quebec City, Quebec, Canada, Aug. 26–29.
[17] Faghih-Imani, A., and Eluru, N., 2016, "Incorporating the Impact of Spatio-Temporal Interactions on Bicycle Sharing System Demand: A Case Study of New York CitiBike System," J. Transp. Geogr., **54**, pp. 218–227.
[18] Christiano, L. J., 2012, "Christopher A. Sims and Vector Autoregressions," Scand. J. Econ., **114**(4), pp. 1082–1104.
[19] Xing, X., and Ye, A., 2022, "Consumption Upgrading and Industrial Structural Change: A General Equilibrium Analysis and Empirical Test With Low-Carbon Green Transition Constraints," Sustainability, **14**(20), p. 13645.
[20] Tang, J., Gao, F., Liu, F., Zhang, W., and Qi, Y., 2019, "Understanding Spatio-Temporal Characteristics of Urban Travel Demand Based on the Combination of GWR and GLM," Sustainability, **11**(19), p. 5525.
[21] Shen, X., Zhou, Y., Jin, S., and Wang, D., 2020, "Spatiotemporal Influence of Land Use and Household Properties on Automobile Travel Demand," Transp. Res. Part D: Transp. Environ., **84**, p. 102359.
[22] Sun, C., and Lu, J., 2023, "The Relative Roles of Different Land-Use Types in Bike-Sharing Demand: A Machine Learning-Based Multiple Interpolation Fusion Method," Inf. Fusion, **95**, pp. 384–400.
[23] 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.
[24] Hausman, J. A., 1978, "Specification Tests in Econometrics," Econom. J. Econom. Soc., **46**(6), pp. 1251–1271.
[25] Ma, X., Zhang, J., Ding, C., and Wang, Y., 2018, "A Geographically and Temporally Weighted Regression Model to Explore the Spatiotemporal Influence of Built Environment on Transit Ridership," Comput. Environ. Urban Syst., **70**, pp. 113–124.
[26] Millo, G., and Piras, G., 2012, "Splm: Spatial Panel Data Models in R," J. Stat. Softw., **47**(1), pp. 1–38.
[27] Elhorst, P., and Vega, S. H., 2013, "On Spatial Econometric Models, Spillover Effects, and W," 53rd Congress of the European Regional Science Association: "Regional Integration: Europe, the Mediterranean and the World Economy," Palermo, Italy, Aug. 27–31.
[28] Zhang, R., Du, Q., Geng, J., Liu, B., and Huang, Y., 2015, "An Improved Spatial Error Model for the Mass Appraisal of Commercial Real Estate Based on Spatial Analysis: Shenzhen as a Case Study," Habitat Int., **46**, pp. 196–205.
[29] Qu, X., and Lee, L., 2015, "Estimating a Spatial Autoregressive Model With an Endogenous Spatial Weight Matrix," J. Econom., **184**(2), pp. 209–232.
[30] Baltagi, B. H., Bresson, G., and Pirotte, A., 2012, "Forecasting With Spatial Panel Data," Comput. Stat. Data Anal., **56**(11), pp. 3381–3397.
[31] Elhorst, J. P., 2014, "Spatial Panel Data Models," *Spatial Econometrics*, Springer, New York, pp. 37–93.
[32] Fu, J. S., Sha, Z., Huang, Y., Wang, M., Fu, Y., and Chen, W., 2017, "Two-Stage Modeling of Customer Choice Preferences in Engineering Design Using Bipartite Network Analysis," Volume 2A: 43rd Design Automation Conference, Cleveland, OH, Aug. 6–9.
[33] Bloomberg, 2016, "China Auto Sales Growth Accelerates on Rising SUV Demand," https://www.bloomberg.com/news/articles/2016-07-08/china-auto-sales-grow-at-faster-pace-on-suv-electric-car-demand.
[34] National Bureau of Statistics of China, 2023, "Regional Data: Annual by Province," https://data.stats.gov.cn/english/easyquery.htm?cn=E0103.
[35] Huo, H., and Wang, M., 2012, "Modeling Future Vehicle Sales and Stock in China," Energy Policy, **43**, pp. 17–29.
[36] Heckman, J. J., 1979, "Sample Selection Bias as a Specification Error," Econom. J. Econom. Soc., **47**(1), pp. 153–161.
[37] Choi, I., 2001, "Unit Root Tests for Panel Data," J. Int. Money Financ., **20**(2), pp. 249–272.
[38] Lopez, L., and Weber, S., 2017, "Testing for Granger Causality in Panel Data," Stata J. Promot. Commun. Stat. Stata, **17**(4), pp. 972–984.
[39] 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.
[40] Bloomberg News, "Chinese Automakers Surge After Government Cuts Purchase Tax." https://www.bloomberg.com/news/articles/2015-09-29/china-reduces-small-car-purchase-tax-by-half-as-deliveries-slow.
[41] Ross Morrow, W., Long, W., and MacDonald, E. F., 2014, "Market-System Design Optimization With Consider-Then-Choose Models," ASME J. Mech. Des., **136**(3), p. 031003.
[42] Autohome China, 2016, "VW FAW Sagitar Specifications—2016 Model," https://car.autohome.com.cn/config/series/633-8349.html%0A.