ISSN: 1648-715X / eISSN: 1648-9179

VILNIUS TECH
Vilnius Gediminas Technical University

INTERNATIONAL JOURNAL OF STRATEGIC PROPERTY MANAGEMENT

2025
Volume 29
Issue 2
Pages 102–113
https://doi.org/10.3846/ijspm.2025.23639

SPATIOTEMPORAL PATTERNS AND PREDICTION OF MULTI-REGION HOUSE PRICES VIA FUNCTIONAL MIXED EFFECTS MODEL

Yilin CHEN, Haitao ZHENG*

Department of Statistics, School of Mathematics, Southwest Jiaotong University, Chengdu, China

Article History:
- received 22 July 2024
- accepted 4 March 2025

Abstract. House prices have always been a popular indicator for real estate market monitoring. This study explores the spatiotemporal patterns of house prices at the community level in San Francisco from January 2009 to April 2024. A functional spatiotemporal semiparametric mixed effects (FST-SM) model was proposed to analyze the Zillow Home Value Index (ZHVI), considering spatiotemporal variations. This response is associated with known influences and unknown latent random effects. The random-effects component was expanded using functional principal components. The conditional autoregressive (CAR) structure of the principal component scores was adopted to analyze nonparametric time trends and spatiotemporal correlations. The proposed model was compared with other time-series models in terms of spatiotemporal prediction. The results show that the prediction accuracy of the proposed model is higher than that of other regular models. In summary, a functional mixed effects model was proposed to describe spatiotemporal patterns and forecast house prices. This study can provide valuable references for decision-making by local governments, real estate suppliers, and house buyers.

Keywords: house price prediction, spatiotemporal data, spatial dependence, functional principal component analysis, conditional autoregressive model, ZHVI.

Online supplementary material: Supporting information for this paper is available as online supplementary material at https://doi.org/10.3846/ijspm.2025.23639

*Corresponding author. E-mail: *htzheng@swjtu.edu.cn*

1. Introduction

Housing prices have always been a key concern for the public and government. Ansell (2014) pointed out that housing prices are important for residents' welfare and macroeconomic health. The Housing Price Index (HPI) is a statistical measure of house prices. This analysis will help various stakeholders make informed decisions (Nunna et al., 2023). The HPI trend can help buyers plan purchases more efficiently. Real estate investors rely on HPI to identify potential investment opportunities. The government utilizes HPI to improve infrastructure development decisions and ensure housing demand (Zhang et al., 2024). The real estate market in the United States is one of the main sources of spillovers from global financial markets (Syriopoulos et al., 2015). Exploring the spatiotemporal patterns of HPI is the key to the health of real estate in the United States. Therefore, we aim to provide an effective HPI spatiotemporal modeling approach for the real estate market in the United States, which will serve as a model for other countries and regions.

There are several housing price indices, including the S&P CoreLogic Case-Shiller (S&P/C-S) index, the Federal Housing Finance Agency (FHFA) index, and the Zillow Home Value Index (ZHVI). In recent years, many scholars have used the ZHVI to study real estate activity, providing insights relevant to economic policies (Holt & Borsuk, 2020; Howard et al., 2023; Kim, 2024). This study proposes a functional spatiotemporal semiparametric mixed effects (FST-SM) model to model the multiregional ZHVI. We used the ZHVI owing to its advantages.

First, the ZHVI is characterized by a fine-grained construction. The ZHVI allows users to observe dynamic changes in considerably small areas or specific subsets of houses (Kim, 2024). Second, the ZHVI is comprehensive. The ZHVI is calculated for all houses, including newly constructed houses and those not traded on the market for many years (Glynn, 2022). The ZHVI can provide more comprehensive information for government decision-making than an index that relies only on houses sold during a specific period. Third, the ZHVI is accessible and timely. Major real estate websites have replaced traditional agents and newspaper transactions (Guo et al., 2020). Contrary to data that must be obtained from local municipalities, the ZHVI is easily available from the Zillow website. The ZHVI

usually publishes data for a given month in the third week of the following month, allowing users to access the latest data in time (Holt & Borsuk, 2020).

The FST-SM models the ZHVI as spatiotemporally correlated functional data. Our aim is not limited to predicting future trends but is interested in exploring spatiotemporal patterns between different regions. We consider house prices at the ZHVI community scale, allowing the functional principal component scores to be spatially correlated across communities. We add spatial random effects to the map to highlight hotspots with abnormally high or rising house price trends. This fine-grained monitoring of spatial information eliminates bias caused by the aggregation of large areas. Moreover, compared to other forecasting models, the FST-SM has an advantage in terms of forecasting accuracy. The ZHVI data cover all metropolitan areas in the United States, allowing us to easily extend the application of the FST-SM to a national scale (Holt & Borsuk, 2020). This study provides a reference point for urban development and housing planning.

The remainder of this paper is organized as follows. Section 2 describes previous studies on spatiotemporal modeling and forecasting of house prices. Section 3 describes the study area and provides a detailed description of the FST-SM model. Section 4 applies the proposed model to the San Francisco community-level ZHVI and tests its predictive capability. Section 5 summarizes this study.

2. Literature review

Traditional house price studies have focused on the time-series framework of econometrics (Anselin, 2013). Statisticians mostly use autoregressive models for forecasting, such as the autoregressive integrated moving average (ARIMA) and vector autoregression (VAR) models (Crawford & Fratantoni, 2003; Farhi & Young, 2010; Iacoviello, 2002). Ren et al. (2017) proposed a Bayesian nonparametric method to predict the ZHVI in Seattle and infer the clustering of census tracts. Economists favor a hedonistic approach to predicting house prices (Can, 1992; Selim, 2008; Straszheim, 1974). Holt and Borsuk (2020) implemented a hedonic approach to Zillow's publicly available data to measure the economic value of green infrastructure nationally.

However, traditional models are usually more demanding in terms of their premises and assumptions, and their use is limited. Subsequently, machine-learning algorithms have become popular. They overcome the shortcomings of traditional models and potentially improve the accuracy of real-estate price predictions (Yazdani, 2021). For example, Chen et al. (2017) predicted the house prices in four major cities in China. They found that recurrent neural networks (RNN) and the long-short term memory (LSTM) achieved better prediction than ARIMA. Bhakta et al. (2021) used two recurrent neural networks to predict the ZHVI for 1584 counties in the United States. They found that a deep learning model produced more accurate results than a traditional regression model.

With the development of spatial econometrics, scholars have begun to recognize real estate activity as a spatial phenomenon. Ignoring spatial effects may lead to large errors (Anselin, 2013; Krause & Bitter, 2012; Pijnenburg, 2017). They attempt to incorporate spatial dependence and heterogeneity into their models. For example, Howard et al. (2023) apply a spatial model to the ZHVI with heterogeneous housing elasticities to help governments identify changes in housing demand during a pandemic. Lee and Park (2018) explore the housing rent-sale ratio in Seoul at a micro-spatial scale regarding the importance of considering spatial variations.

Many multi-region neural network models that consider spatial correlation have emerged recently. For example, Lee (2022) implemented a multi-output LSTM method. By exploiting spatial correlation, he predicted house prices and transaction volumes in four regions of Seoul. He verified that the prediction accuracy of his method was better than that of a single-output LSTM model. Ge (2019) combined a Convolutional Neural Network (CNN) and LSTM networks to capture spatiotemporal and community features. This method effectively improved the accuracy of house price predictions. This evidence suggests that incorporating spatial dependence and heterogeneity into the model improves prediction accuracy.

These machine learning models already exhibit good prediction performance. However, Mullainathan and Spiess (2017) point out that machine-learning models cannot estimate or infer parameters from probability distributions. This black-box model is less transparent and does not provide a good indication of economic significance (Kim et al., 2020; Lee, 2022; Shi, 2023). We aim to overcome the limitations of machine learning and develop a spatiotemporal model with strong interpretability.

Benefiting from the development of acquisition techniques, functional data analysis (FDA) has been widely used in many fields, including environment science, geology, and healthcare (Collazos et al., 2023; Hael, 2023; Li & Guan, 2014). The functional principal component analysis (FPCA) is an important FDA technique. It is concerned with the efficient and interpretable representation of functional variability. Specifically, spatiotemporal dependence is induced between spatiotemporally correlated functional curves (Ramsay & Silverman, 2002). In this study, we model the ZHVI as spatiotemporally correlated functional data. Contrary to other multi-region forecasting models, we set up the realization of only one spatiotemporal process rather than a collection of multiple time series. This approach focuses on the relationships between variables within a function's domain and is not merely interested in extrapolating real estate activity in the future (Zhang et al., 2016).

When modeling spatiotemporally correlated functional data, spatial correlation is usually contained in random effects as a nonparametric function. The spatial correlation

functions are approximated by spline or Matérn families (Li et al., 2022). For example, Li and Guan (2014) proposed a semiparametric model for functional processes. Their study explained the spatial correlation between curves by constructing a spatial correlation function for the Matérn family. Scheipl et al. (2015) proposed an additive mixed model and approximated the functional spatial correlation via a separable covariance structure. Nevertheless, the separable covariance structure exhibits considerable local heterogeneity and does not apply to functional data with complex spatial correlations. Zhang et al. (2016) induced spatial correlation using a CAR model (Besag, 1974). They improved the regression performance of the spatiotemporal correlation function by extending the nonseparable covariance structure.

In this study, the FST-SM assumes that arbitrary fixed effects can be included within the framework of semiparametric mixed models. Moreover, the CAR model has an inseparable covariance structure and spatially correlated random effects. Therefore, the proposed model should be more flexible and accurate for estimation and inference purposes.

3. Data and model specification

3.1. Study area and dataset

San Francisco, California, was selected for analysis in this study. According to the United States Census Bureau, the city has a population density of 7,194 people/km^2. This makes it one of the most densely populated cities in the United States. Consequently, this region has become an area of close interest to the government and real estate investors (Kong & Kepili, 2023). Figure 1 shows the geographic location of San Francisco, as well as the administrative divisions of San Francisco at the community level (see Supplementary Material for more details).

Table 1. Housing stocks in San Francisco (source: United States Census Bureau, 2022)

Housing structure				
Category	Single-family residences	Condos	Mobile home or other	Total
Units	116,739	244,502	671	361,912
Proportion	32.26%	67.55%	18.01%	100%
Number of bedrooms				
Category	No bedroom or 1 bedroom	2 or 3 bedrooms	4 or more bedrooms	Total
Units	139,637	184,722	37,553	361,912
Proportion	38.59%	51.04%	10.37%	100%

As of 2022, the total housing stock in San Francisco was 361,912 units. Regarding the type of residential structure, single-family residences accounted for approximately 32%, and condos accounted for approximately 68%. Regarding the number of bedrooms, houses with two and three in number predominated at 51.04%. In Table 1, single-family residences and condos with two or three bedrooms are considered representative housing types in San Francisco.

The San Francisco community-level ZHVI data for January 2009 to April 2024 were downloaded from the Zillow website. The dataset contained missing values for 16 communities, which were excluded from our analysis (see Supplementary Material for more details). We used the data before January 2023 as the training set and the data after that as the test set. We also downloaded data for a subset of San Francisco single-family residences, condos, and houses containing two, three, and four bedrooms, respectively. The units of the ZHVI are 10,000 US dollars.

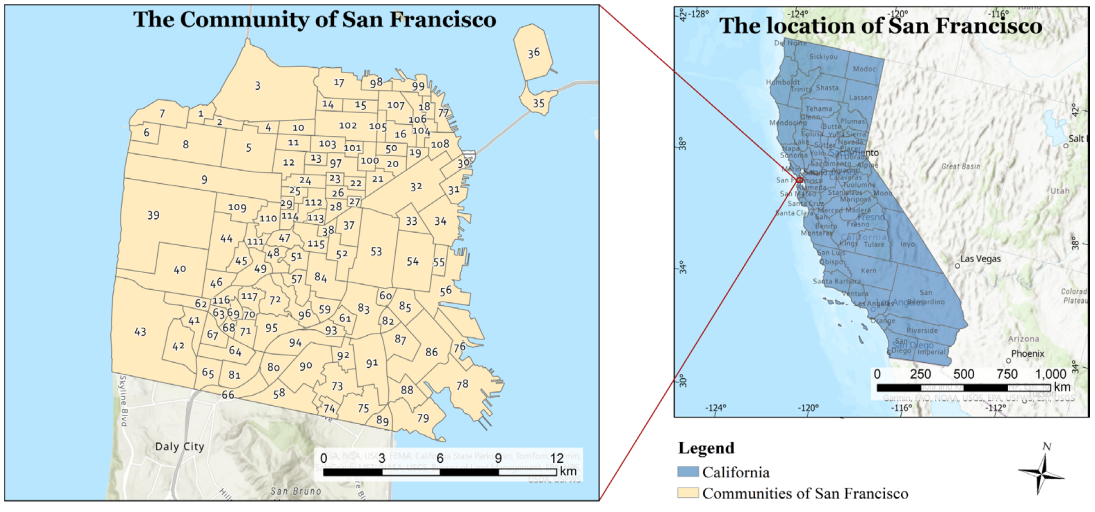


Figure 1. Research area (source: San Francisco Mayor's Office of Neighborhood Services, 2023)

3.2. Characteristics of data

Figure 2 shows the time-series curves of the ZHVI for all communities. The ZHVI was collected for 186 consecutive months, which is typical of time-series data. In each community, the ZHVI for the current month was similar to that of the previous month. Overall, the trend increased to varying degrees. Therefore, we consider the temporal correlation between the series when specifying a model. In particular, there is a clear upward trend after November 2011 and August 2020. We focus on these two periods in our subsequent analyses. On the one hand, this is related to policy regulations after being hit by the 2008 subprime crisis (Goetzmann et al., 2012). On the other hand, after the coronavirus disease pandemic in 2019, prices tended to rise rapidly as people resumed their productive lives (Zhao, 2020).

The Moran index (Moran's I) is a statistic commonly used to describe spatial correlations. It is usually categorized into global and local Moran's I (Moran, 1950). The global Moran's I for the ZHVI varies between 0.4 and 0.6 throughout the study time interval. This indicates a positive spatial autocorrelation in the studied area. We further observed the local clustering phenomenon using local Moran's I. The local Moran scatterplot shows the scatter relationship between the normalized z-value and spatial lag (Anselin, 1995). We computed the local Moran's I for all months and selected three months, as shown in Figure 3. The ZHVI of each community was concentrated in the first and third quadrants. This indicates that housing prices in San Francisco are spatially clustered, and the neighboring regions show a similar pattern. In summary, time series and spatial interdependencies are important components that deserve careful treatment when modeling ZHVI.

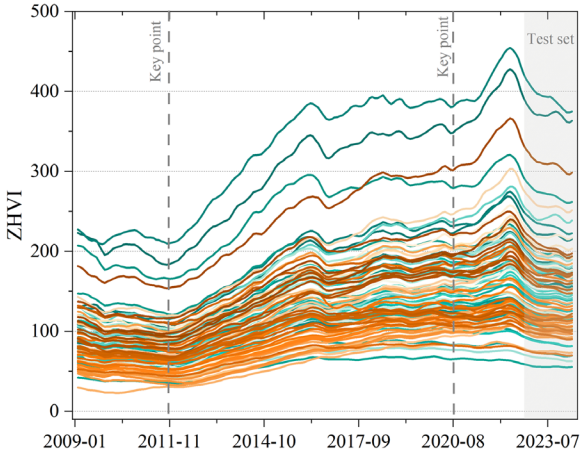


Figure 2. ZHVI trajectories for all communities

3.3. Model specification

From a previous analysis of the spatiotemporal correlation of housing prices in San Francisco, we propose a FST-SM model that considers spatiotemporal correlations as follows:

$$Y_s(t_k) = \boldsymbol{X}_s^T(t_k)\boldsymbol{\beta} + \delta_s(t_k) + \varepsilon_s(t_k), \tag{1}$$

where: $Y_s(t_k)$ is the k-th month ZHVI measured at a particular community s; $s = 1,\ldots,101$ denotes community, and $k = 1,\ldots,168$ denotes the month. In the framework of semiparametric mixed effects model, the fixed effects $\boldsymbol{X}_s(t)$ are theoretically allowed to include arbitrary variables of interest. Let $\boldsymbol{X}_s^T(t_k) = (Y_s(t_{k-1}), Y_s(t_{k-2}),\ldots, Y_s(t_{k-n}))^T$, and n denotes time lag. In our study, we focus on the relationship between the current month's ZHVI and its time-lagged series over a six-month period. Let n = 6, $\boldsymbol{\beta} = (\beta_1, \beta_2,\ldots,\beta_6)^T$ be the corresponding vectors of regression coefficients. $\varepsilon_s(t_k) \sim_{i.i.d.} N(0,\sigma^2)$ denote measurement errors.

$\delta_s(t_k)$ is a random effect that can be decomposed using the Karhunen-Loève expansion (Li & Guan, 2014):

$$\delta_s(t_k) = \mu(t_k) + \sum_{j=1}^{p} \xi_{sj}\psi_j(t_k), \tag{2}$$

where: $\mu(t) = E\{\delta_s(t)\}$ is the expectation taken over all locations; $\psi_j(t)$ is the j-th orthonormal eigenfunction of the covariance function $C(t_1,t_2) = \mathrm{cov}\{\delta_s(t_1), \delta_s(t_2)\}$, satisfying $\int \psi_j(t)\psi_{j'}(t) = 1$ if $j = j'$, and 0 otherwise; $\boldsymbol{\xi}_j = (\xi_{1j}, \xi_{2j},\ldots,\xi_{101j})^T$ are the corresponding principal component scores with zero mean. In practice, the Karhunen-Loève expansion is truncated by the number of principal components p; p can be ∞ in theory; p is usually selected by the fraction of variance

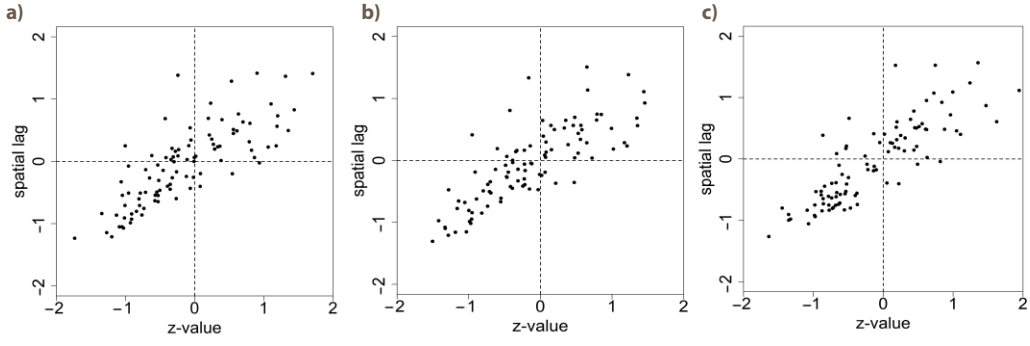


Figure 3. Local Moran's I scatter plot

explained (FVE) (Li et al., 2022). Furthermore, $\varepsilon_s(t_k)$ and the principal component scores $\boldsymbol{\xi}_j$ are independent.

The spatial correlation across communities can be modeled via a CAR model to the community-specific principal component scores, ξ_{sj}. The CAR model describes the spatial dependence between regions by specifying a full conditional distribution in autoregressive form. Determining the community system is crucial for determining the CAR model's dependence structure (Besag, 1974). A typical example is whether two communities have adjoining boundaries as the criterion for determination (Banerjee et al., 2014). Let $W = \{w_{ss'}\}$ denote the adjacency matrix; $w_{ss'} = 1$ if s and s' are neighbors; otherwise, $w_{ss'} = 0$.

According to the theoretical approach of the CAR model, the full conditional distribution of the j-th principal component score ξ_{sj} of region s is specified by:

$$\xi_{sj} \mid \{\xi_{s'j}\}_{s \neq s'} \sim N(\rho \sum_{s \neq s'} w_{ss'} \xi_{sj} / d_s, \alpha_j / d_s), \tag{3}$$

where: $d_s = \sum_{s \neq s'} w_{ss'}$ indicate the number of neighbors of area s; ρ is the spatial correlation parameter; α_j is the variance of the j-th feature component. The factorization theorem proposed by Besag (1974) can be applied to derive the joint distribution of principal component scores $\boldsymbol{\xi}_j$. The spatial correlation parameter ρ is restricted to lie between the boundaries given by the inverse of the minimum and maximum eigenvalues of the matrix $D_w^{-1/2} W D_w^{-1/2}$. This ensures that the matrix $(D_w - \rho W)$ is positive definite (Banerjee et al., 2014).

3.4. Estimation procedure

Owing to the desired numerical stability and computational efficiency of the B-spline, $\mu(t)$ can be approximated using a regression B-spline (Shen et al., 1998). Let $\hat{\mu}(t) = \boldsymbol{B}^T(t)\hat{\boldsymbol{\nu}}$ and $\boldsymbol{B}(t) = \{B_1(t), B_2(t), \ldots, B_N(t)\}^T$ be a set of spline basis functions, and N denotes the number of basis functions. Conditional on the spline order $\kappa = 4$, the optimal number of B-splines with internal knots L can be selected by minimizing the Bayes information criterion (BIC) (Zhang & Li, 2022); $N = \kappa + L$. Let $\boldsymbol{\nu} = \{\nu_1^T, \nu_2^T, \ldots, \nu_\kappa^T\}^T$ be the corresponding vector of the basis-function coefficients.

First we obtain the mean function $\hat{E}\{Y_s(t)\}$ by least squares, and the parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\nu}}$ can be obtained by minimizing the least-squares function $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\nu}}) = \underset{(\boldsymbol{\beta}, \boldsymbol{\nu}) \in \mathbb{R}^{\kappa \times N}}{\arg\min} L(\boldsymbol{\beta}, \boldsymbol{\nu})$:

$$L(\boldsymbol{\beta}, \boldsymbol{\nu}) = \sum_{s=1}^{n} \sum_{k=1}^{K} \left\{ Y_s(t_k) - \boldsymbol{X}_s(t_k)\boldsymbol{\beta} - \boldsymbol{B}^T(t_k)\boldsymbol{\nu} \right\}^2. \tag{4}$$

The estimated mean function is used to centralize the observed data, $\hat{Y}_s(t) = Y_s(t) - \hat{E}\{Y_s(t)\}$. Subsequently, the empirical covariance $\hat{G}(t, t') = \sum_{s=1}^{n} \hat{Y}_s(t)\hat{Y}_s(t') / n$ is obtained. Once the covariance has been obtained, estimates of the eigenfunctions $\hat{\psi}_j(t)$ and principal component scores $\hat{\xi}_{sj}$ can be obtained by FPCA. To retain sufficient information in the algorithm's initial step, we set FVE ≥ 99% as the criterion for calculating the number of feature components retained in the truncated expansion equation.

By estimating the mean function and eigenfunctions, the community-specific principal component scores (ξ_{sj}), spatial parameters (α_j and ρ), and measurement error variance (σ^2) of the FST-SM model can be estimated. Next, the posterior distributions of each parameter can be obtained by sampling via the Markov Chain Monte Carlo (MCMC) method (Gelfand, 2000). The prior distributions of the model parameters are given by:

$$\xi_j \sim N(0, \alpha_j (D_w - \rho W)^{-1}),\ \varepsilon_{sk} \sim N(0, \sigma^2),\ \sigma^2 \sim IG(a_{\sigma^2}, b_{\sigma^2}),$$
$$\alpha_j \sim IG(a_{\alpha_j}, b_{\alpha_j}),\ \rho \sim Unif(a_\rho, b_\rho). \tag{5}$$

The joint posterior distribution of the parameters is more complex when the parameter dimensions are high. Therefore, the Gibbs sampling algorithm can be applied to multidimensional parameter sampling problems. A combined Gibbs and Metropolis sampler is used to sample the posterior distributions (Gelfand, 2000). Thus, all the parameters for the FST-SM model are estimated. With the estimated parameters, the FST-SM model can be used to construct the trajectories of the region-specific prediction curves as follows:

$$\hat{Y}_s(t_k) = \boldsymbol{X}_s^T(t_k)\hat{\boldsymbol{\beta}} + \hat{\mu}(t_k) + \sum_{j=1}^{p} \xi_{sj} \hat{\psi}_j(t_k). \tag{6}$$

4. Results and discussion

4.1. Model fitting results and fixed effects

The FST-SM model was used to fit the ZHVI for 101 communities in San Francisco. Figure 4 displays the fitted curves and 95% confidence intervals for the three communities (the fitting results for the other communities are detailed in the Supplementary Material). Our model exhibited a good fit. Figure 5 shows a histogram of the residual density. The normality of the residuals shows good performance–the model applies to the actual ZHVI data.

As Table 2 shows, the fixed effects of $\hat{\beta}_1$, $\hat{\beta}_4$, and $\hat{\beta}_6$ are positive. This indicates that the ZHVI series with a time lag of 1, 4, or 6 positively correlates with the current month series. Numerically, the effect of the time-lagged ZHVI series fluctuates on a three-month cycle in the short term. The ZHVI series has the most significant impact with a time lag of 1. Conversely, the ZHVI series with a time lag of 2, 3, or 5 was negatively correlated with the current month series. The $\hat{\beta}_3$ mean and median values were –0.0374 and –0.0366, respectively. The 5th and 95th percentile ranges included zero and fluctuated around it. This suggests that a ZHVI series with a time lag of 3 has little effect on the current month's ZHVI.

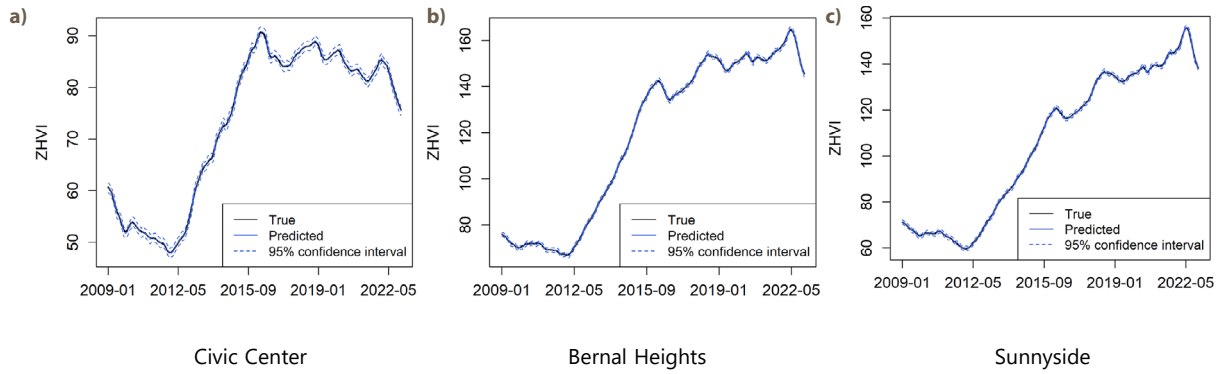


Figure 4. Fitting curve of ZHVI

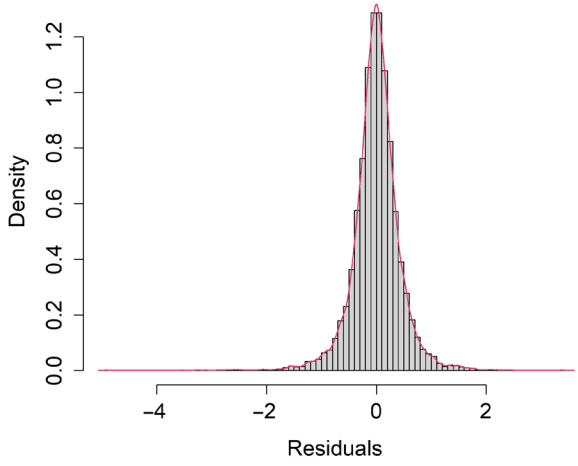


Figure 5. Histogram of residual density

Table 2. Model fit results

Parameter	Mean	5th percentile	50th percentile	95th percentile
β_1	2.3931	2.3912	2.3933	2.3954
β_2	–1.8530	–1.8435	–1.8427	–1.8419
β_3	–0.0374	–0.0876	–0.0366	0.0144
β_4	1.2056	1.2102	1.2134	1.2166
β_5	–1.0263	–1.0288	–1.0267	–1.0246
β_6	0.3182	0.3165	0.3188	0.3211
α_1	0.0302	0.0233	0.0298	0.0385
α_2	0.0258	0.0202	0.0255	0.0327
σ^2	0.2044	0.2008	0.2044	0.2081

4.2. Analysis of random effects

For random effects, we track long-term potential changes in ZHVI through the eigenfunctions of the FPCA. We visualize differences in the trajectories of change across communities. From Figure 6, the mean function $\hat{\mu}(t)$ is a curve without significant fluctuations. It represents the baseline of random variation in ZHVI for all communities. The first eigenfunction $\hat{\psi}_1(t)$ contributes 97.84% of the variation component, reflecting the most dominant pattern of variation in the random effects.

We focus on the two periods mentioned in Section 3.2. From August 2020 to December 2022, $\hat{\psi}_1(t)$ exhibited a downward trend. In May 2022, it crossed the zero baseline. This indicates a downward trend in the rate of increase of the ZHVI during this period, which gradually changes from an increase to a decrease. Corresponding to the intervals in Figure 2, it can be confirmed that $\hat{\psi}_1(t)$ is highly consistent with the actual ZHVI change. This reflects changes in human household behavior that led to increased demand and decreased housing supply shortly after the start of the pandemic. This led to an unusual spike in house prices (Zhao, 2020).

This was combined with a heat map of the principal component scores, as shown in Figure 7a. We observed that the northwestern part of San Francisco had the highest $\hat{\xi}_{i1}$, followed by a higher concentration in the central region. This suggests a more significant trend of $\hat{\psi}_1(t)$ in the northwest and center. These communities could be further explored as hotspots. For example, the Golden Gate Bridge near the Presidio Heights, Seacliff, and Cow Hollow communities provides easy access for commuters. This contributes to the high real estate prices. However, influenced by the pandemic, telecommuting has been gradually replacing traditional work patterns. There is a rapidly

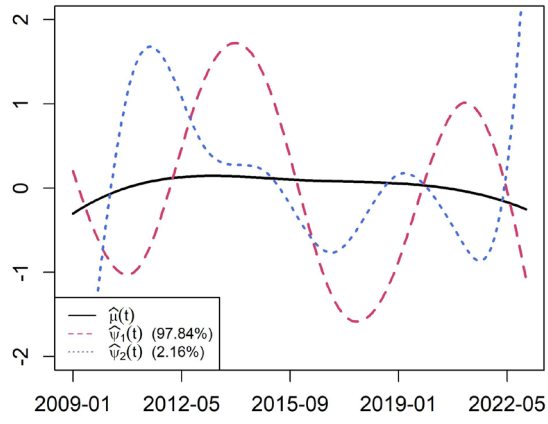
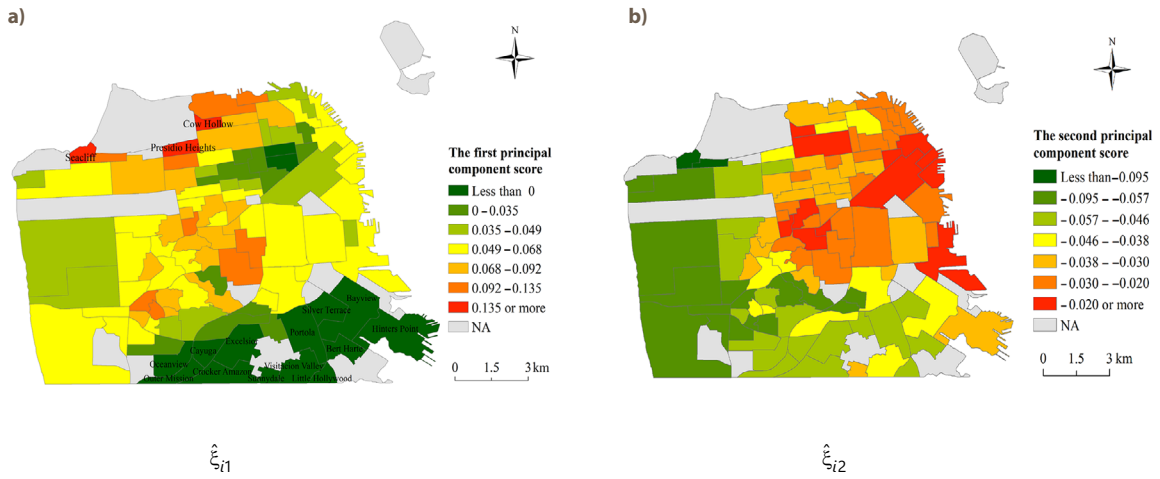


Figure 6. The mean function and eigenfunctions



$\hat{\xi}_{i1}$ $\hat{\xi}_{i2}$

Figure 7. The principal component scores

diminishing demand for people to live in transit-accessible communities, leading to a potential trend for residential values to fall faster (Bartik et al., 2020). Conversely, some communities in South San Francisco have a negative $\hat{\xi}_{i1}$, and the ZHVI shows a growing trend. This proves that the pandemic shifted housing demand from the central and coastal communities to the suburbs, consistent with the findings of Liu and Su (2021).

From November 2011 to September 2015, $\hat{\psi}_1(t)$ was above the zero baseline. This indicates an upward trend in ZHVI. This was because of the impact of the subprime mortgage crisis. With the Federal Reserve reducing interest rates and implementing quantitative easing, housing prices gradually began to recover (Goetzmann et al., 2012). Figure 7a shows that $\hat{\xi}_{i1}$ was relatively high in some coastal communities. House prices in this area have tended to trend upward and more dramatically. This may be due to policies supporting increased commercial investment in densely populated communities (Immergluck, 2011).

Although $\hat{\psi}_2(t)$ explains only 2.16% of the variation, it can be used as a reference. As shown in Figure 6, opposite trends were observed before and after September 2015. It extracts a comparison for San Francisco before and after September 2015. Combined with Figure 7b, the northeast of San Francisco has more significant potential for upward mobility after September 2015.

4.3. Forecast results and analysis

In this study, the comparison criteria of mean absolute error (MAE), mean square error (MSE), root mean square error (RMSE), and mean absolute percentage error (MAPE) were applied to evaluate the prediction accuracy (Chiu, 2024; Guo et al., 2020). However, the input sets varied based on the different requirements of each model structure (see Supplementary Material for details). The competition model is described as follows.

(1) The Single-output LSTM (Chen et al., 2017) model is a variant of a RNN. It is primarily used for single-objective time-dependent learning and time-series prediction.

(2) The Multi-output LSTM (Lee, 2022) model compensates for the inability of the Single-output LSTM to capture correlations from a single-point dataset. The single-point sequence is limited to that location, ignoring surrounding information. The Multi-output LSTM mitigates this weakness by utilizing the correlation between multiple targets.

(3) Multiple experiments have shown that the CNN-LSTM has superior predictive performance compared to LSTM networks (Samal et al., 2022). The CNN-LSTM (Ge, 2019) model aims to predict house prices by capturing and analyzing the impact of nearby locations. The graph CNN captures spatial dependencies and utilizes an LSTM network to learn the historical house prices.

Table 3 presents the results of forecast comparisons. The various error indicators of the FST-SM, CNN-LSTM, Multi-output LSTM, and Single-output LSTM models increase in that order. The FST-SM has significant advantages in terms of its predictive accuracy. This is because the Single-output LSTM model ignores the spatial correlation between locations. The Multi-output LSTM compensates for the shortcomings of the Single-output LSTM model but requires extensive data training to capture spatial information. The CNN-LSTM model can capture spatial dependencies using a graph CNN. Nonetheless, a CNN must compress a spatial matrix, which causes some information to be lost. The FST-SM extracts spatial correlation features through the CAR structure, which overcomes some of the shortcomings of the competitive models mentioned above.

Table 3. Comparison results of error indicators

Error indicator	FST-SM	CNN-LSTM	Multi-output LSTM	Single-output LSTM
MAE	0.9051	1.1137	1.1569	3.4085
MSE	1.2135	2.0985	2.3976	19.0897
RMSE	1.0580	1.3311	1.3937	4.0372
MAPE	0.0066	0.0075	0.0076	0.0231

Figure 8 shows the forecast curves for these three communities. Although all models can predict house price trends to some extent, the degree of agreement is not as good as that of the FST-SM. In particular, the FST-SM model fully demonstrates its extraordinary ability to recognize the timing of turning points. For example, February to April 2024 is the rising ZHVI interval. The FST-SM predicts this trend more accurately, whereas the other models do not capture it. Overall, the FST-SM outperforms the predictions of the other models.

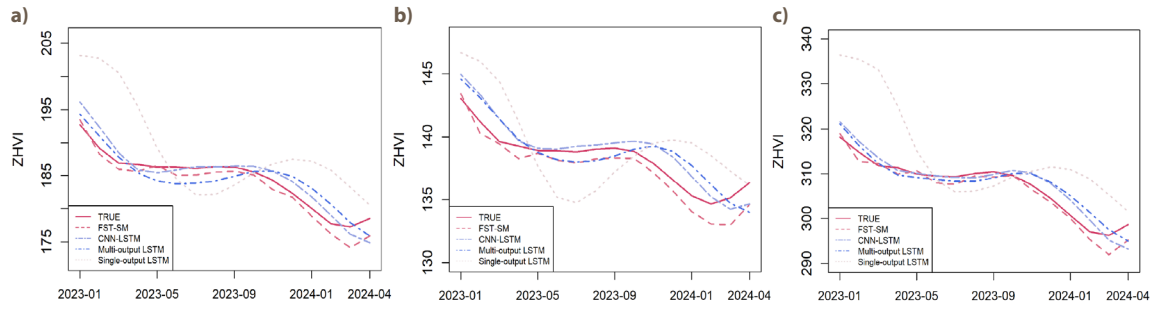


Figure 8. Comparison of ZHVI prediction curves

Table 4. ZHVI residential subset prediction results

Month	Percentile	Overall	Single-family residences	Condos	2 bedrooms	3 bedrooms	4 bedrooms
2023-01	5	88.8161 (0.1904)	130.6183 (0.0371)	69.9278 (0.1003)	92.1121 (0.0014)	103.9428 (0.0199)	114.7249 (0.0850)
	50	142.9784 (0.1438)	230.3824 (0.5316)	122.0919 (0.1916)	135.9638 (0.1417)	167.4671 (0.4854)	208.3976 (0.2827)
	95	246.1873 (0.8068)	454.4879 (0.5635)	157.5119 (0.5001)	166.1920 (0.4028)	275.0367 (1.1311)	441.2342 (1.2064)
2023-04	5	86.4171 (0.3270)	124.6943 (0.6167)	69.1765 (0.2308)	89.9465 (0.5335)	101.5862 (0.5623)	111.5690 (0.6924)
	50	137.3603 (0.7649)	222.5954 (0.6493)	117.5137 (0.2263)	132.3492 (0.4639)	161.4653 (0.9070)	200.8723 (1.2443)
	95	237.8059 (1.4284)	446.1697 (1.1614)	150.0831 (0.6714)	159.2862 (0.8789)	264.4190 (0.7348)	432.1506 (0.1540)
2023-07	5	85.4962 (0.4583)	120.0906 (1.0300)	69.47164 (0.1864)	90.0070 (0.5373)	101.8850 (0.5414)	111.4080 (0.8053)
	50	135.3264 (0.8724)	220.2721 (1.2146)	116.5444 (0.3725)	132.2603 (0.5352)	162.3448 (0.7693)	198.7519 (0.8050)
	95	241.1443 (1.3257)	436.3786 (1.5328)	150.3187 (0.6451)	158.3788 (0.9680)	260.4916 (0.7089)	423.4888 (1.2510)
2023-10	5	83.7889 (0.3519)	117.8929 (0.5134)	69.4270 (0.1299)	90.1206 (0.3729)	102.3589 (0.4141)	111.8329 (0.6676)
	50	135.3976 (0.2438)	216.5315 (0.4857)	117.7401 (0.0834)	132.3739 (0.3778)	162.5441 (0.4972)	197.7688 (0.4513)
	95	246.0838 (0.2441)	429.3126 (0.5742)	153.5688 (0.0646)	158.9336 (0.6217)	258.5721 (0.5765)	415.9487 (1.2039)
2024-01	5	80.8189 (0.7944)	112.0381 (1.4893)	67.6248 (0.3530)	88.3499 (0.5830)	100.3803 (0.6383)	109.5331 (1.0377)
	50	131.9407 (1.2749)	208.2297 (1.0354)	114.8471 (0.5624)	130.3200 (0.6471)	159.4193 (0.0088)	194.1298 (0.3935)
	95	237.9695 (0.5845)	415.8963 (1.0603)	150.6103 (0.6583)	156.0143 (0.5171)	253.5207 (0.2744)	404.1218 (0.2902)
2024-04	5	80.6832 (0.7708)	109.9047 (2.1899)	67.5166 (0.2856)	88.4267 (1.1994)	100.8422 (1.0281)	109.7590 (1.5882)
	50	131.0827 (2.0153)	204.0298 (2.8620)	114.9881 (0.6099)	128.9319 (1.4886)	158.9817 (1.1643)	192.9864 (2.1826)
	95	235.8732 (2.8297)	410.6985 (2.9119)	150.8002 (0.6357)	154.2578 (2.4532)	251.0495 (0.7720)	400.1541 (1.5715)

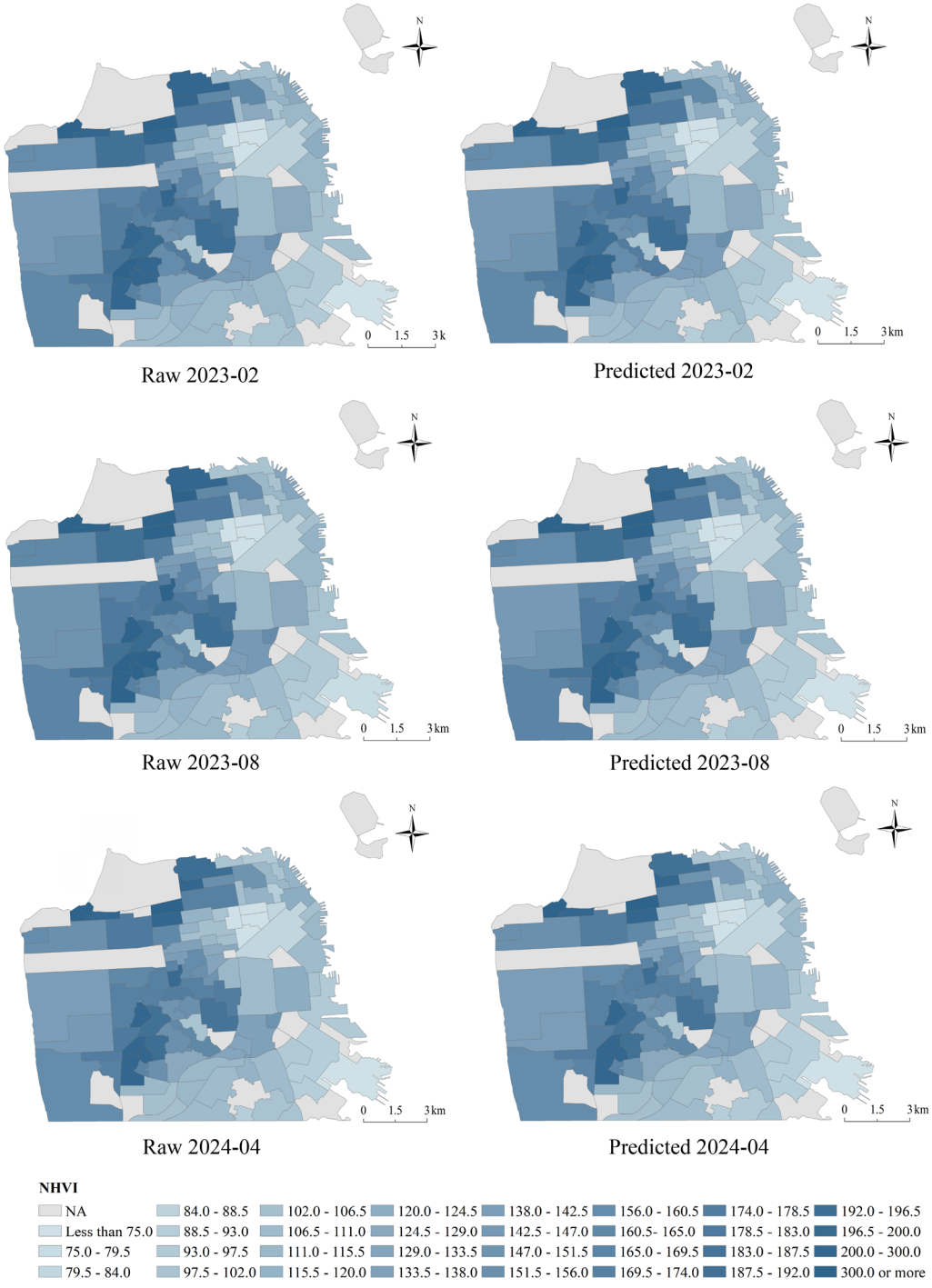


Figure 9. Spatial prediction

Buyers prefer the type of house based on their needs and preferences. For example, some people prefer to live in single-family residences with courtyards, whereas others prefer condos with easy access to public transportation. Overall, the ZHVI may not be an excellent solution for this problem. Therefore, we analyzed the predictive power of a subset of the ZHVI. This may be a more meaningful way to provide information to homebuyers.

Table 4 shows the predictions for the subsets of the ZHVI for six of the months of 2023. We selected five housing types that comprised a significant portion of San Francisco's housing stock. From Figure 8, we predicted a decreasing trend in the ZHVI from January 2023 to April 2024. Homeowners listed many of their postponed houses for sale after the end of the pandemic. Large supply and falling demand put downward pressure on house prices (Gamber et al., 2023). Table 4 shows a decrease in the ZHVI of approximately 11.44%, 5.82%, 5.17%, 5.07%, and 7.45% for each of the five subsets. Single-family residences and houses with four bedrooms declined more than other

houses–their value was more elastic. Policymakers can develop specific responses based on the elasticity of different residential subsets when implementing regulations or interventions.

Previously, we showed that the FST-SM model performs well in time-series prediction. However, it also has strong spatial prediction capability. Figure 9 shows the predicted results for the three months in our selected test set, along with the actual values. By comparing the actual values on the left side with the predicted values on the right side, the ZHVI has a high predictive accuracy in space. The ZHVI exhibits a spatial pattern higher in the northwest and central portions of San Francisco. It gradually decreases from northwest to southeast. This spatially scaled predictive analysis differs from those used in previous studies. This is free from the biases associated with data aggregation. Homebuyers can choose a reasonable area to buy a house based on spatial forecast results and their needs.

5. Conclusions

This study used the FST-SM to model the San Francisco community-level ZHVI as functional spatiotemporal data. A CAR model with spatial dependence was introduced into the principal component scores. It entirely borrowed information on the geographic location and effectively identified the spatiotemporal patterns of ZHVI. Owing to data and methodological limitations, such studies have not been mentioned previously. The first eigenfunction extracted from the random effects explained 97.84% of the variation. This weighting is sufficient to focus the authorities' attention on policy regulation. We monitor housing price hotspots through two significant events: the pandemic and the subprime mortgage crisis. When local governments monitor the housing market, heat maps with principal component scores can be used to identify hotspots.

Moreover, the FST-SM method successfully estimates the parameters and nonparametric functions of the model. This overcomes the limitations of machine learning methods. Our approach focuses on the relationships between variables in the functional domain within a functional spatiotemporal correlation modeling framework. A significant advantage of the ZHVI is its scalability. The ZHVI covers essentially all the metropolitan areas of the United States. This is expected to extend the functional spatiotemporal model to the entire metropolitan area in the United States. Therefore, it is essential to monitor housing information across the United States.

This study demonstrates the practical applicability of FST-SM for spatiotemporal prediction in three ways. First, the FST-SM has apparent advantages in forecasting errors and trends compared with other time-series forecasting models. The improved prediction performance can be attributed to the spatial correlations between communities. Second, we utilize the fine-grained constructive features of ZHVI to classify a subset of ZHVI residences for prediction. This is beneficial for helping homebuyers develop more efficient house-buying programs. Third, we test the model for spatial prediction at a particular transect and obtain more accurate predictions. Community-scale prediction frees us from the bias introduced by data aggregation. If a real estate company monitors relatively high house prices in a particular area, this can be interpreted as high demand in the future. The relevant authorities must promptly adjust their plans.

In summary, monitoring and forecasting trends in the real estate market are major issues faced by policymakers and private investors. This study has significant economic implications and provides policy contributions to real estate activities. On the one hand, accurate projections help homebuyers understand affordability, mortgage requirements, and long-term financial commitments. Real estate investors rely on accurate forecasts to minimize financial risk and ensure a return on investments, especially in volatile markets. Banks and lending institutions use house-price forecasts to assess lending risks and ensure appropriate interest rates and credit terms. Governments use forecasts to inform decisions on infrastructure development and affordable housing policies. On the other hand, spatially fine-grained information monitoring can help governments track price trends in different real estate markets. These trends indicate the health and efficiency of the market and can be monitored for a bubble in house prices or a decline in buyers' potential to purchase houses. Investors can adopt different portfolio-management strategies based on various market conditions. Homebuyers can enhance the purposefulness and transparency of their choice of house area.

Useful extensions of the FST-SM for future research include modeling non-Gaussian transform function responses. Most real estate data do not satisfy a normal distribution, and the direct assumption of a normal distribution introduces errors. There is also a scalable direction for modeling higher levels of fine-grained models. This facilitates the provision of comprehensive housing information to higher-level agencies.

Acknowledgements

This work is supported by the National Natural Science Foundation of China (Grant no. 51778546) and the Fundamental Research Funds for the Central Universities (SWJTU, 2682021ZTPY078). Besides, we would like to thank the R programming language, the open-source software, for its convenience in this research. We also thank the anonymous reviewers of this manuscript for their useful comments.

Funding

This work is supported by the National Natural Science Foundation of China (Grant no. 51778546) and the Fundamental Research Funds for the Central Universities (SWJTU, 2682021ZTPY078).

Author contributions

Yilin Chen: Data curation, Software analysis, Visualization, Writing original draft. Haitao Zheng: Supervision, Management and coordination of research planning and execution.

Disclosure statement

No conflict of interest exits in the submission of this manuscript, and manuscript is approved by all authors for publication.

References

Anselin, L. (1995). Local indicators of spatial association–LISA. *Geographical Analysis*, *27*(2), 93–115. https://doi.org/10.1111/j.1538-4632.1995.tb00338.x

Anselin, L. (2013). *Spatial econometrics: Methods and models* (Vol. 4). Springer Science & Business Media. https://doi.org/10.1007/978-94-015-7799-1

Ansell, B. E. N. (2014). The political economy of ownership: Housing markets and the welfare state. *American Political Science Review*, *108*(2), 383–402. https://doi.org/10.1017/S0003055414000045

Banerjee, S., Carlin, B. P., & Gelfand, A. E. (2014). *Hierarchical modeling and analysis for spatial data* (2nd ed.). Chapman & Hall/CRC. https://doi.org/10.1201/b17115

Bartik, A., Cullen, Z. B., Glaeser, E., Luca, M., & Stanton, C. T. (2020). *What jobs are being done at home during the Covid-19 crisis? Evidence from firm-level surveys* (Working Paper No. 27422). National Bureau of Economic. https://doi.org/10.3386/w27422

Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society: Series B* (*Methodological*), *36*(2), 192–225. https://doi.org/10.1111/j.2517-6161.1974.tb00999.x

Bhakta, K., Duggal, S., Fleisher, C., Karnati, P., & Tusnial, A. (2021). *Forecasting local housing market trends using recurrent neural networks*. https://www.pratyushakarnati.me/assets/documents/DL_Final_Report.pdf?i=1

Can, A. (1992). Specification and estimation of hedonic housing price models. *Regional Science and Urban Economics*, *22*(3), 453–474. https://doi.org/10.1016/0166-0462(92)90039-4

Chen, X., Wei, L., & Xu, J. (2017). *House price prediction using LSTM*. arXiv. https://doi.org/10.48550/arXiv.1709.08432

Chiu, K. C. (2024). A long short-term memory model for forecasting housing prices in Taiwan in the post-epidemic era through big data analytics. *Asia Pacific Management Review*, *29*(3), 273–283. https://doi.org/10.1016/j.apmrv.2023.08.002

Collazos, J. A. A., Dias, R., & Medeiros, M. C. (2023). Modeling the evolution of deaths from infectious diseases with functional data models: The case of COVID-19 in Brazil. *Statistics in Medicine*, *42*(7), 993–1012. https://doi.org/10.1002/sim.9654

Crawford, G. W., & Fratantoni, M. C. (2003). Assessing the forecasting performance of regime-switching, ARIMA and GARCH models of house prices. *Real Estate Economics*, *31*(2), 223–243. https://doi.org/10.1111/1540-6229.00064

Farhi, C., & Young, J. (2010). Forecasting residential rents: The case of Auckland, New Zealand. *Pacific Rim Property Research Journal*, *16*(2), 207–220. https://doi.org/10.1080/14445921.2010.11104302

Gamber, W., Graham, J., & Yadav, A. (2023). Stuck at home: Housing demand during the COVID-19 pandemic. *Journal of Housing Economics*, *59*, Article 101908. https://doi.org/10.1016/j.jhe.2022.101908

Ge, C. (2019). A LSTM and graph CNN combined network for community house price forecasting. In *2019 20th IEEE International Conference on Mobile Data Management* (*MDM*) (pp. 393–394). https://doi.org/10.1109/MDM.2019.00-15

Gelfand, A. E. (2000). Gibbs sampling. *Journal of the American Statistical Association*, *95*(452), 1300–1304. https://doi.org/10.2307/2669775

Glynn, C. (2022). Learning low-dimensional structure in house price indices. *Applied Stochastic Models in Business and Industry*, *38*(1), 151–168. https://doi.org/10.1002/asmb.2653

Goetzmann, W. N., Peng, L., & Yen, J. (2012). The subprime crisis and house price appreciation. *The Journal of Real Estate Finance and Economics*, *44*(1), 36–66. https://doi.org/10.1007/s11146-011-9321-4

Guo, J. Q., Chiang, S. H., Liu, M., Yang, C. C., & Gou, K. Y. (2020). Can machine learning algorithms associated with text mining from internet data improve housing price prediction performance? *International Journal of Strategic Property Management*, *24*(5), 300–312. https://doi.org/10.3846/ijspm.2020.12742

Hael, M. A. (2023). Unveiling air pollution patterns in Yemen: A spatial-temporal functional data analysis. *Environmental Science and Pollution Research*, *30*(17), 50067–50095. https://doi.org/10.1007/s11356-023-25790-3

Holt, J. R., & Borsuk, M. E. (2020). Using Zillow data to value green space amenities at the neighborhood scale. *Urban Forestry & Urban Greening*, *56*, Article 126794. https://doi.org/10.1016/j.ufug.2020.126794

Howard, G., Liebersohn, J., & Ozimek, A. (2023). The short- and long-run effects of remote work on U.S. housing markets. *Journal of Financial Economics*, *150*(1), 166–184. https://doi.org/10.1016/j.jfineco.2023.103705

Iacoviello, M. (2002). *House prices and business cycles in Europe: A VAR analysis* (Working Paper No. 540). Boston College, Department of Economics. https://EconPapers.repec.org/RePEc:boc:bocoec:540

Immergluck, D. (2011). *Foreclosed: High-risk lending, deregulation, and the undermining of America's mortgage market*. Cornell University Press. https://doi.org/10.7591/9780801458828

Kim, H., Kwon, Y., & Choi, Y. (2020). Assessing the impact of public rental housing on the housing prices in proximity: Based on the regional and local level of price prediction models using long short-term memory (LSTM). *Sustainability*, *12*(18), Article 7520. https://doi.org/10.3390/su12187520

Kim, J. (2024). Aging, housing prices, and young adults' homeownership. *Cities*, *149*, Article 104914. https://doi.org/10.1016/j.cities.2024.104914

Kong, J., & Kepili, E. (2023). A survey analysis: The current real estate marketing situation in the China greater bay area in the context of the COVID-19 epidemic. *Real Estate Management and Valuation*, *31*(3), 1–19. https://doi.org/10.2478/remav-2023-0017

Krause, A. L., & Bitter, C. (2012). Spatial econometrics, land values and sustainability: Trends in real estate valuation research. *Cities*, *29*, S19–S25. https://doi.org/10.1016/j.cities.2012.06.006

Lee, C. (2022). Forecasting spatially correlated targets: Simultaneous prediction of housing market activity across multiple areas. *International Journal of Strategic Property Management*, *26*(2), 119–126. https://doi.org/10.3846/ijspm.2022.16786

Lee, C., & Park, K. (2018). Analyzing the rent-to-price ratio for the housing market at the micro-spatial scale. *International Journal*

of Strategic Property Management, *22*(3), 223–233. https://doi.org/10.3846/ijspm.2018.1416

Li, Y. H., & Guan, Y. T. (2014). Functional principal component analysis of spatiotemporal point processes with applications in disease surveillance. *Journal of the American Statistical Association*, *109*(507), 1205–1215. https://doi.org/10.1080/01621459.2014.885434

Li, Y. H., Nguyen, D. V., Kürüm, E., Rhee, C. M., Banerjee, S., & Sentürk, D. (2022). Multilevel varying coefficient spatiotemporal model. *Stat*, *11*(1), Article e438. https://doi.org/10.1002/sta4.438

Liu, S. T., & Su, Y. C. (2021). The impact of the COVID-19 pandemic on the demand for density: Evidence from the US housing market. *Economics Letters*, *207*, Article 110010. https://doi.org/10.1016/j.econlet.2021.110010

Moran, P. A. P. (1950). Notes on continuous stochastic phenomena. *Biometrika*, *37*(1/2), 17–23. https://doi.org/10.2307/2332142

Mullainathan, S., & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives*, *31*(2), 87–106. https://doi.org/10.1257/jep.31.2.87

Nunna, K. C., Zhou, Z., & Shakya, S. R. (2023, December 4–6). Time series forecasting of U.S. housing price index using machine and deep learning techniques. In *2023 IEEE Asia-Pacific Conference on Computer Science and Data Engineering* (*CSDE*) (pp. 1–6). IEEE. https://doi.org/10.1109/CSDE59766.2023.10487733

Pijnenburg, K. (2017). The spatial dimension of US house prices. *Urban Studies*, *54*(2), 466–481. https://doi.org/10.1177/0042098015606595

Ramsay, J. O., & Silverman, B. W. (2002). *Applied functional data analysis: Methods and case studies*. Springer. https://doi.org/10.1007/b98886

Ren, Y., Fox, E. B., & Bruce, A. (2017). Clustering correlated, sparse data streams to estimate a localized housing price index. *The Annals of Applied Statistics*, *11*(2), 808–839. https://doi.org/10.1214/17-AOAS1019

Samal, K. K. R., Babu, K. S., & Das, S. K. (2022). Multi-output spatiotemporal air pollution forecasting using neural network approach. *Applied Soft Computing*, *126*, Article 109316. https://doi.org/10.1016/j.asoc.2022.109316

San Francisco Mayor's Office of Neighborhood Services. (2023). *San Francisco find tool: Geographic locations and boundaries*. San Francisco, USA. https://download.geofabrik.de/north-america/us/california/norcal-latest-free.shp.zip

Scheipl, F., Staicu, A. M., & Greven, S. (2015). Functional additive mixed models. *Journal of Computational and Graphical Statistics*, *24*(2), 477–501. https://doi.org/10.1080/10618600.2014.901914

Selim, S. (2008). Determinants of house prices in Turkey: A hedonic regression model. *Doğuş Üniversitesi Dergisi*, *9*(1), 65–76. https://doi.org/10.31671/dogus.2019.223

Shen, X., Wolfe, D. A., & Zhou, S. (1998). Local asymptotics for regression splines and confidence regions. *The Annals of Statistics*, *26*(5), 1760–1782. https://doi.org/10.1214/aos/1024691356

Shi, S. (2023). Comparison of real estate price prediction based on LSTM and LGBM. *Highlights in Science, Engineering and Technology*, *49*, 294–301. https://doi.org/10.54097/hset.v49i.8521

Straszheim, M. (1974). Hedonic estimation of housing market prices: A further comment. *The Review of Economics and Statistics*, *56*(3), 404–406. https://doi.org/10.2307/1923985

Syriopoulos, T., Makram, B., & Boubaker, A. (2015). Stock market volatility spillovers and portfolio hedging: BRICS and the financial crisis. *International Review of Financial Analysis*, *39*, 7–18. https://doi.org/10.1016/j.irfa.2015.01.015

United States Census Bureau. (2022). Physical housing characteristics for housing units. In *American community survey*. https://data.census.gov/table?t=Types+of+Rooms&g=050XX00US06075&y=2022

Yazdani, M. (2021). *Machine learning, deep learning, and hedonic methods for real estate price prediction*. arXiv. https://EconPapers.repec.org/RePEc:arx:papers:2110.07151

Zhang, H. Z., & Li, Y. H. (2022). Unified principal component analysis for sparse and dense functional data under spatial dependency. *Journal of Business & Economic Statistics*, *40*(4), 1523–1537. https://doi.org/10.1080/07350015.2021.1938085

Zhang, H., Li, Y., & Branco, P. (2024). Describe the house and I will tell you the price: House price prediction with textual description data. *Natural Language Engineering*, *30*(4), 661–695. https://doi.org/10.1017/S1351324923000360

Zhang, L., Baladandayuthapani, V., Zhu, H., Baggerly, K. A., Majewski, T., Czerniak, B. A., & Morris, J. S. (2016). Functional CAR models for large spatially correlated functional datasets. *Journal of the American Statistical Association*, *111*(514), 772–786. https://doi.org/10.1080/01621459.2015.1042581

Zhao, Y. (2020). *US housing market during COVID-19: Aggregate and distributional evidence* (Working Paper No. 212). International Monetary Fund. https://doi.org/10.5089/9781513557816.001