RESEARCH ARTICLE

Kernel-based geographically and temporally weighted autoregressive model for house price estimation

Jooyong Shim¹, Changha Hwang^{2*}

¹ Department of Statistics, Inje University, Gimhae, Korea, ² Department of Applied Statistics, Dankook University, Yongin, Korea

* chwang@dankook.ac.kr

OPEN ACCESS

Citation: Shim J, Hwang C (2018) Kernel-based geographically and temporally weighted autoregressive model for house price estimation. PLoS ONE 13(10): e0205063. <https://doi.org/10.1371/journal.pone.0205063>

Editor: Fenghua Wen, Central South University, CHINA

Received: March 21, 2018

Accepted: August 21, 2018

Published: October 11, 2018

Copyright: © 2018 Shim, Hwang. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are third party and are available upon request from the case study whose authors may be contacted by emailing Bo Huang at bohuang@cuhk.edu.hk. The authors of this study accessed the data in the same manner, and did not have any special privileges to the data.

Funding: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology with grant no. (NRF-2014R1A1A2054917, NRF-2018R1D1A1B07042349). This work was

Abstract

Spatiotemporal nonstationarity and autocorrelation are two crucial points in modeling geographical data. Previous studies have demonstrated that geographically and temporally weighted autoregressive (GTWAR) model accounts for both spatiotemporal nonstationarity and autocorrelation simultaneously to estimate house prices. Therefore, this paper proposes a kernel-based GTWAR (KBGTWAR) model by incorporating the basic principle of support vector machine regression into spatially and temporally varying coefficients model. The efficacy of KBGTWAR model is demonstrated through a case study on housing prices in the city of Shenzhen, China, from year 2004 to 2008. Comparing the existing models, KBGTWAR model obtains the lowest value for the residual sum of squares (RSS) and the highest value for the coefficient of determination R^2 . Moreover, KBGTWAR model improves the goodness of fit of the existing GTWAR model from 12.0 to 4.5 in terms of RSS, from 0.914 to 0.968 in terms of R^2 and from 3.84 to 4.45 in terms of F-statistic. The results show that KBGTWAR model provides a comparatively high goodness of fit and sufficient explanatory power for both spatiotemporal nonstationarity and autocorrelation. The results of this study demonstrate that the proposed KBGTWAR model can be used to effectively formulate policies for real estate management.

Introduction

Analysis of relationships between an output variable and input variables in spatiotemporal fields has recently attracted attention in a data analysis community. Spatiotemporal models occur when data are obtained across time as well as space. Spatial or temporal autocorrelation exists between the observations. In addition, spatial or temporal nonstationarity arises in the relationships. Ordinary regression models have neglected these issues that violate the standard Gauss-Markov assumptions. Spatial econometrics models have been proposed to deal with these issues. Spatiotemporal regression models can contribute to better understanding complex phenomena studied in geographical information science and other fields.

supported by “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20174030201740). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Spatial models not only can account for spatial autocorrelation but also can deal with spatial lag dependence or/and spatial error dependence. There are two approaches for handling spatial autocorrelation. The first is the weight matrix approach, which uses a spatial weight matrix to deal with the spatial relationships between observations. This approach is based on the work of [1]. See for details [2] and [3]. The second is the geostatistical approach, which directly models the covariance matrix of the error terms. This approach is based on the work of [4]. See for details [5] and [6]. Moreover, [3] proposed a spatiotemporal autoregressive (STAR) model to incorporate temporal correlations between observations and found it powerful in the context of residential real estate. [7] developed a two-order STAR model with a Bayesian procedure to effectively detect and correct heteroscedasticity among the residuals.

The spatial models aforementioned have made remarkable contributions to effectively applying spatial and temporal process to regression modeling. However, these models assumed that a global level relationship exists across the study area. By the way, the assumption of stationary or structural stability over space and time is often impractical, because parameters tend to change depending on the study area. To deal with spatial heterogeneity in housing markets, a variety of localized modeling techniques have been proposed. As a result, [8] and [9] demonstrated the usefulness of locally weighted regression in modeling nonmonocentric cities. [10], [11] and [12] proposed geographically weighted regression (GWR), which has grown in popularity amongst real estate modeling methods. GWR model is a local spatial model for exploring spatial nonstationarity. To capture both spatial nonstationarity and spatial autocorrelation in a complex process, [13] proposed a GWR with spatially lagged variables to utilize the spatial variations. But this model did not consider temporal information. To model the effects of temporal heterogeneity, [14] and [15] proposed a geographically and temporally weighted regression (GTWR) to simultaneously capture both spatial and temporal nonstationarity by integrating temporal effects into the traditional GWR model. [16] developed a GTWR model based on travel time distance metrics. But these GTWR models do not consider spatial autocorrelation.

[17] asserted that it is better to deal with both spatial and temporal heterogeneity along with spatial autocorrelation effects in a mixed model. It is because spatial heterogeneity and autocorrelation are generally related in the context of modeling although the two problems are theoretically distinguished. However, it is not easy to build an integrated model with spatiotemporal autocorrelated and heterogeneous effects. [18] proposed a geographically and temporally weighted autoregressive (GTWAR) model to deal with spatiotemporal variations. In this paper we propose an efficient kernel-based GTWAR (KBGTWAR) model by incorporating the basic principle of support vector machine (SVM) regression into spatially and temporally varying coefficients model. SVM, first developed by [19] and his group at AT&T Bell Laboratories, has been successfully applied to a number of real world problems related to classification and regression problems.

The rest of this paper is organized as follows. Section 2 briefly describes the basic principle of GTWR and GTWAR models. Section 3 proposes an efficient KBGTWAR model. Section 4 and section 5 present case study and conclusion, respectively.

GTWAR model

In this section we briefly review the basic framework for GTWR and GTWAR models. We also review how to determine the adjustment parameters related with these two models.

GTWR model

The GWR model is a spatially varying coefficient regression approach for exploring spatial nonstationarity of a regression relationship for spatial data [10, 11, 12]. The GWR model

extends the traditional linear regression model by allowing local rather than global parameters to be estimated. [14] developed a GTWR model to deal with spatial and temporal nonstationarity simultaneously by integrating temporal effects into the GWR model. [15] developed a GTWR model focusing on spatiotemporal kernel function definition and spatiotemporal bandwidth optimization. [16] developed a GTWR model based on non-Euclidean travel distance. The GTWR model can be expressed as follows:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^d \beta_k(u_i, v_i, t_i)x_{ki} + \epsilon_i, i = 1, \dots, n, \quad (1)$$

where (u_i, v_i, t_i) is the coordinate of the observation i in space (u_i, v_i) at time t_i , $\beta_0(u_i, v_i, t_i)$ indicates the intercept value, $\beta_k(u_i, v_i, t_i)$ indicates the slope for each variable k and each space-time point i , and ϵ_i represents the random error with no correlation between different points. Here, local parameters $\beta_0(u_i, v_i, t_i)$ and $\beta_k(u_i, v_i, t_i)$ are continuous functions of the point (u_i, v_i, t_i) . We notice that the GTWR model (1) is a spatially and temporally varying coefficient model.

For a given data set, these local parameters are estimated using the weighted least square procedure. The relevant weights W_{ij} , for $j = 1, \dots, n$, indicate the proximity of each data point to the point (u_i, v_i, t_i) . Let

$$\boldsymbol{\beta}(u_i, v_i, t_i) = (\beta_0(u_i, v_i, t_i), \beta_1(u_i, v_i, t_i), \dots, \beta_d(u_i, v_i, t_i))^t \quad (2)$$

be the vector of the local parameters for the space-time point i . Here, the superscript t represents the transpose of vector or matrix. Then, $\boldsymbol{\beta}(u_i, v_i, t_i)$ is estimated by

$$\hat{\boldsymbol{\beta}}(u_i, v_i, t_i) = (\mathbf{X}^t \mathbf{W}(u_i, v_i, t_i) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_i, v_i, t_i) \mathbf{y}, \quad (3)$$

where $\mathbf{X}$ is the $n \times (d+1)$ matrix of input variables, $\mathbf{y}$ is the n -dimensional vector of output variable, and $\mathbf{W}(u_i, v_i, t_i)$ is an $n \times n$ weighting matrix of the form

$$\mathbf{W}(u_i, v_i, t_i) = \text{diag}\{W_{i1}, W_{i2}, \dots, W_{in}\}. \quad (4)$$

In addition, the fitted value $\hat{\mathbf{y}}$ is obtained as follows:

$$\hat{\mathbf{y}} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{x}_1^a (\mathbf{X}^t \mathbf{W}(u_1, v_1, t_1) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_1, v_1, t_1) \\ \mathbf{x}_2^a (\mathbf{X}^t \mathbf{W}(u_2, v_2, t_2) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_2, v_2, t_2) \\ \vdots \\ \mathbf{x}_n^a (\mathbf{X}^t \mathbf{W}(u_n, v_n, t_n) \mathbf{X})^{-1} \mathbf{X}^t \mathbf{W}(u_n, v_n, t_n) \end{pmatrix} \mathbf{y}, \quad (5)$$

where $\mathbf{x}_i^a$ is the i th row of the matrix $\mathbf{X}$ such that $\mathbf{x}_i^a = (1, x_{1i}, \dots, x_{di})$.

The weights W_{ij} are usually obtained through an adaptive kernel function. The adaptive kernel function attempts to adjust for the density of data points. This adaptive kernel function could use the same number of observed points in each local neighborhood set. The most commonly used adaptive kernel function is the Gaussian function

$$W_{ij} = \exp\left(-\frac{d_{ij}^2}{h^2}\right), \quad (6)$$

where d_{ij} is a spatiotemporal distance between points i and j and h is a nonnegative parameter known as bandwidth, which produces a decay of influence with distance. When the spatial and temporal distances between points i and j are given by

$(d_{ij}^S)^2 = (u_i - u_j)^2 + (v_i - v_j)^2$, $(d_{ij}^T)^2 = (t_i - t_j)^2$, we can construct the spatiotemporal distance as a linear combination between $(d_{ij}^S)^2$ and $(d_{ij}^T)^2$ as follows:

$$(d_{ij}^{ST})^2 = \mu^S[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu^T(t_i - t_j)^2, \quad (7)$$

where μ^S represents the scale factor of spatial distance and μ^T represents the scale factor of temporal distance. Thus, the weights W_{ij} can be expressed as

$$\begin{aligned} W_{ij} &= \exp \left\{ - \left(\frac{\mu^S[(u_i - u_j)^2 + (v_i - v_j)^2] + \mu^T(t_i - t_j)^2}{h_{ST}^2} \right) \right\} \\ &= \exp \left\{ - \left(\frac{(u_i - u_j)^2 + (v_i - v_j)^2}{h_S^2} + \frac{(t_i - t_j)^2}{h_T^2} \right) \right\} \end{aligned} \quad (8)$$

$$\begin{aligned} &= \exp \left\{ - \left(\frac{(d_{ij}^S)^2}{h_S^2} + \frac{(d_{ij}^T)^2}{h_T^2} \right) \right\} \\ &= W_{ij}^S \times W_{ij}^T, \end{aligned} \quad (9)$$

where $h_{ST}^2, h_S^2 = h_{ST}^2 / \mu^S$ and $h_T^2 = h_{ST}^2 / \mu^T$ are the parameters of spatiotemporal, spatial, and temporal bandwidths, respectively. Thus, if the spatial and temporal bandwidths are determined, the weight matrix $W(u_i, v_i, t_i)$ and $\hat{\beta}(u_i, v_i, t_i)$ can be obtained.

[18] developed the improved GTWR (IGTWR) with the following spatiotemporal distance

$$\begin{cases} d_{ij}^{ST} = \mu^S d_{ij}^S + \mu^T d_{ij}^T + 2\sqrt{\mu^S \mu^T d_{ij}^S d_{ij}^T} \cos(v), & t_j < t_i \\ d_{ij}^{ST} = \infty, & t_j > t_i \end{cases}, \quad (10)$$

where $v \in [0, \pi]$. The adjustment parameters μ^S, μ^T and v should be determined.

GTWAR model

To account for both spatiotemporal heterogeneity and spatial autocorrelation effects simultaneously, [18] developed a GTWAR, which combines the IGTWR model with the autocorrelation regression model. Spatial autocorrelation is considered by introducing the spatial lag $\sum_{j=1}^n \bar{W}_{ij} y_j$ in a linear regression relationship [20].

$$y_i = \rho_i \sum_{j=1}^n \bar{W}_{ij} y_j + \epsilon_i, i = 1, 2, \dots, n, \quad (11)$$

where ρ_i represents a spatial autoregressive parameter varying across geographical locations, ϵ_i denotes the random error in the relationship, and $\bar{W}_{ij}$ is the element in the i th row and j th column of the $n \times n$ spatial weight matrix $\bar{W}$ with $\bar{W}_{ii} = 0$. The elements $\bar{W}_{ij}$ are typically row-normalized, such that for each i , $\sum_{j=1}^n \bar{W}_{ij} = 1$. Consequently, the spatial lag may be interpreted as a weighted average of the neighbors. We notice that the model (11) is a locally based autoregressive model. Incorporating (11) into the IGTWR model, [18] proposed the following

GTWAR model:

$$y_i = \beta_0(u_i, v_i, t_i) + \rho(u_i, v_i, t_i) \sum_{j=1}^n \bar{W}_{ij} y_j + \sum_{k=1}^d \beta_k(u_i, v_i, t_i) x_{ki} + \epsilon_i, i = 1, \dots, n, \quad (12)$$

where $\rho(u_i, v_i, t_i)$ is a scalar spatiotemporal autoregressive parameter at point i .

Model selection

Parameter estimation in GTWR and GTWAR is highly dependent on the adjustment parameters μ^S, μ^T and/or v associated with the weighting function used. The selection of the adjustment parameters for GTWR and GTWAR can be determined using a cross validation (CV) approach or the corrected Akaike information criterion (AIC) from [12]. Since the adaptive kernel function generally uses the same number q of the nearest observed points, there is one more adjustment parameter. To obtain an optimal value of q , the CV or corrected AIC approach could be used. [14] argue that only the parameter ratio $\tau = \mu^T / \mu^S$ plays an important role in constructing weights. Hence, they set $\mu^S = 1$ to reduce the number of adjustment parameters, and so only three parameters, q, μ^T and v .

The CV function is defined as

$$CV(\lambda) = \sum_{i=1}^n (y_i - \hat{y}^{(-i)}(\lambda))^2, \quad (13)$$

where λ is the vector of adjustment parameters associated with the GTWR or GTWAR model and $\hat{y}^{(-i)}(\lambda)$ is the fitted value of y_i with the observation i omitted from the calibration process. The corrected AIC function is defined according to [21] as follows:

$$AIC_c(\lambda) = n \log \left(\frac{RSS(\lambda)}{n} \right) + n \log(2\pi) + n \left(\frac{n + \text{tr}(\mathbf{H}(\lambda))}{n - 2 - \text{tr}(\mathbf{H}(\lambda))} \right), \quad (14)$$

where n is the number of data points in data set, RSS is the residual sum of squares defined as $RSS(\lambda) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, and $\text{tr}(\mathbf{H}(\lambda))$ is the trace of the hat matrix $\mathbf{H}(\lambda)$ associated with the GTWR or GTWAR model with the adjustment parameter vector λ , which satisfies the relationship $\hat{\mathbf{y}} = \mathbf{H}(\lambda)\mathbf{y}$. See (5) for the hat matrix related with the GTWR model and [18] for the hat matrix related with the GTWAR model. Here log denotes the natural logarithm. The adjustment parameters are achieved automatically with an optimization technique by minimizing the CV function (13) or the corrected AIC function (14).

KBGTWAR model

In this section we review SVM regression and illustrate how to develop KBGTWAR model. The underlying idea of KBGTWAR model is that the true mean specification is approximated by combining linear SVM regression with nonlinear feature mapping function of the coordinate vector (u_i, v_i, t_i) of the observed point i .

SVM regression

The foundations of SVM have been originally proposed by [19] and are gaining popularity due to many attractive features, and promising empirical performance. We now briefly review SVM regression. See for details [22]. Suppose we are given the data set $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n$ with each covariate vector $\mathbf{x}_i \in \mathbb{R}^d$ and the output $y_i \in \mathbb{R}$. We basically illustrate the case of the

linear SVM regression, taking the form

$$f(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + b, \quad (15)$$

where b is the bias term. The goal of the linear SVM regression is to find a linear regression function $f(\mathbf{x})$ that approximates all pairs $(\mathbf{x}_i, y_i)$ with ϵ precision and is simultaneously as flat as possible. Flatness means that we seek a small $\mathbf{w}$. One way to guarantee this is to minimize the norm $\|\mathbf{w}\|^2$. To make it feasible, we introduce slack variables ξ_i, ξ_i^* representing upper and lower constraints on the outputs. This leads to the convex optimization problem

$$\begin{aligned} \min_{\mathbf{w}, b, \xi, \xi^*} \mathcal{J} &= \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{subject to} \quad &\begin{cases} y_i - \mathbf{w}^t \mathbf{x}_i - b \leq \epsilon + \xi_i \\ \mathbf{w}^t \mathbf{x}_i + b - y_i \leq \epsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \end{cases} \end{aligned} \quad (16)$$

The regularization parameter $C > 0$ determines the trade-off between the flatness of f and the fitting errors.

The key idea is to construct the primal Lagrange function

$$\begin{aligned} \mathcal{L} = \mathcal{J} &- \sum_{i=1}^n \alpha_i (\epsilon + \xi_i - y_i + \mathbf{w}^t \mathbf{x}_i + b) \\ &- \sum_{i=1}^n \alpha_i^* (\epsilon + \xi_i^* + y_i - \mathbf{w}^t \mathbf{x}_i - b) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*), \end{aligned} \quad (17)$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0$ are Lagrange multipliers. The conditions for optimality are given by

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{0} \rightarrow \mathbf{w} = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \mathbf{x}_i \quad (18)$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad (19)$$

$$\frac{\partial \mathcal{L}}{\partial \xi_i^{(*)}} = 0 \rightarrow C - \alpha_i^{(*)} - \eta_i^{(*)} = 0, \quad i = 1, \dots, n \quad (20)$$

Here, we refer to α_i and α_i^* by $\alpha_i^{(*)}$.

Substituting (18), (19) and (20) into (17) yields the dual optimization problem.

$$\begin{aligned} \min_{\alpha_i, \alpha_i^*} \quad &\frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*) (\alpha_j - \alpha_j^*) \mathbf{x}_i^t \mathbf{x}_j + \epsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) \\ \text{subject to} \quad &\sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \quad \text{and} \quad \alpha_i, \alpha_i^* \in [0, C] \end{aligned} \quad (21)$$

Solving the above optimization problem determines the Lagrange multipliers $\hat{\alpha}_i, \hat{\alpha}_i^*$. Thus the optimal regression function can be rewritten as follows:

$$\hat{f}(\mathbf{x}) = \sum_{i=1}^n (\hat{\alpha}_i - \hat{\alpha}_i^*) \mathbf{x}_i^t \mathbf{x} + \hat{b}. \quad (22)$$

Here, the optimal value of b is obtained by employing Karush-Kuhn-Tucker [23] conditions as

follows:

$$\hat{b} = \frac{1}{n_s} \sum_{i \in I_{sv}} \left(y_i - \sum_{j=1}^n (\hat{\alpha}_j - \hat{\alpha}_j^*) \mathbf{x}_j^t \mathbf{x}_i - \epsilon \times \text{sign}(\hat{\alpha}_i - \hat{\alpha}_i^*) \right), \quad (23)$$

where n_s is the size of $I_{sv} = \{i = 1, 2, \dots, n : 0 < |\alpha_i - \alpha_i^*| < C\}$.

We now consider how to make the linear SVM regression algorithm nonlinear. This could be achieved by simply preprocessing the input vectors $\mathbf{x}_i$ by a nonlinear feature mapping function $\phi : \mathbb{R}^d \rightarrow \mathcal{F}$ into some feature space $\mathcal{F}$, and then applying the linear SVM regression algorithm. Thus, we only need to use the kernel trick $K(\mathbf{x}_i, \mathbf{x}_j) = \phi(\mathbf{x}_i)^t \phi(\mathbf{x}_j)$ in the Eqs (21), (22) and (23) associated with the linear SVM regression algorithm [24]. We never need to know explicitly what ϕ is. The most popular kernel is Gaussian kernel defined by

$$K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|^2 / 2\kappa), \quad i, j = 1, \dots, n, \quad (24)$$

where $\kappa > 0$ is the kernel parameter.

KBGTWAR model

Given the data set $\mathcal{D} = \{(\mathbf{u}_i, \mathbf{x}_i, y_i)\}_{i=1}^n$ with each coordinate vector $\mathbf{u}_i = (u_i, v_i, t_i)$, covariate vector $\mathbf{x}_i \in \mathbb{R}^d$ and the output $y_i \in \mathbb{R}$, we consider the following GTWAR model reexpressed from (12):

$$y_i = \rho(\mathbf{u}_i) \sum_{j=1}^n \bar{W}_{ij} y_j + \sum_{k=0}^d \beta_k(\mathbf{u}_i) x_{ki} + \epsilon_i, \quad i = 1, \dots, n, \quad (25)$$

where $x_{0i} = 1$ and x_{ki} is the k th component of $\mathbf{x}_i = (x_{1i}, \dots, x_{di})^t$ for $k = 1, \dots, d$. For simplicity, we use the spatiotemporal weights matrix $\bar{\mathbf{W}}$ constructed from q -nearest neighbors based on the following modified spatiotemporal distance

$$\begin{cases} d_{ij}^{ST} = \sqrt{(d_{ij}^S)^2 + (d_{ij}^T)^2}, & t_j < t_i \\ d_{ij}^{ST} = \infty, & t_j > t_i \end{cases} \quad (26)$$

When the observation j is one of q nearest neighbors of the observation i , $\bar{W}_{ij} = 1/q$, otherwise $\bar{W}_{ij} = 0$. The diagonal elements are $\bar{W}_{ii} = 0$. The spatiotemporal weights matrix $\bar{\mathbf{W}}$ is row-normalized.

To develop KBGTWAR model, we apply the basic principle of the linear SVM regression to the GTWAR model (12) after preprocessing the coordinate vectors $\mathbf{u}_i$ by a nonlinear feature mapping function ϕ into some feature space $\mathcal{F}$. Thus, we first assume that $\rho(\mathbf{u}_i)$ and $\beta_k(\mathbf{u}_i)$ for $k = 0, 1, \dots, d$ are nonlinearly related to the coordinate vector $\mathbf{u}_i$ such that $\rho(\mathbf{u}_i) = \mathbf{w}^t \phi(\mathbf{u}_i) + b \in [0, 1]$ or $[-1, 1]$, $\beta_k(\mathbf{u}_i) = \mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k$, where $\mathbf{w}$ and $\mathbf{w}_k$ are the weight vectors of dimension d_h corresponding to $\phi(\mathbf{u}_i)$. Here, ϕ is defined in an implicit way. An inner product in feature space has an equivalent kernel such that $K(\mathbf{u}_i, \mathbf{u}_j) = \phi(\mathbf{u}_i)^t \phi(\mathbf{u}_j)$, provided certain conditions hold [24]. Several choices of the kernel function are possible. As mentioned before, Gaussian kernel is most widely used. Thus, we focus on the choice of an Gaussian kernel (24) in the sequel.

Using the basic idea of the linear SVM regression, we define the convex optimization problem:

$$\begin{aligned} \min_{\mathbf{w}, \mathbf{w}_k, b, b_k, \xi, \xi^*} \mathcal{J} &= \frac{1}{2} \|\mathbf{w}\|^2 + \frac{1}{2} \sum_{k=0}^d \|\mathbf{w}_k\|^2 + C \sum_{i=1}^n (\xi_i + \xi_i^*) \\ \text{subject to} & \begin{cases} y_i - (\mathbf{w}^t \phi(\mathbf{u}_i) + b) \bar{\mathbf{W}}_i \mathbf{y} - \sum_{k=0}^d x_{ki} (\mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k) \leq \xi_i \\ (\mathbf{w}^t \phi(\mathbf{u}_i) + b) \bar{\mathbf{W}}_i \mathbf{y} + \sum_{k=0}^d x_{ki} (\mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k) - y_i \leq \xi_i^* \\ \xi_i, \xi_i^* \geq 0 \\ 0 \leq \mathbf{w}^t \phi(\mathbf{u}_i) + b \leq 1 \text{ if } \rho \in [0, 1] \\ \text{or } -1 \leq \mathbf{w}^t \phi(\mathbf{u}_i) + b \leq 1 \text{ if } \rho \in [-1, 1] \end{cases}, \end{aligned} \quad (27)$$

where the constant $C > 0$ is the regularization parameter and we use the notation $\bar{\mathbf{W}}_i$ to refer to the i th row of matrix $\bar{\mathbf{W}}$. For simplicity we set up the size ϵ of the insensitive zone to zero.

Now we are going to construct the primal Lagrange function for the case where $\rho \in [0, 1]$ or $\rho \in [-1, 1]$. For the case of $\rho \in [0, 1]$, the Lagrange function is constructed as follows:

$$\begin{aligned} \mathcal{L} = \mathcal{J} &- \sum_{i=1}^n \alpha_i (\xi_i - y_i + (\mathbf{w}^t \phi(\mathbf{u}_i) + b) \bar{\mathbf{W}}_i \mathbf{y} + \sum_{k=0}^d x_{ki} (\mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k)) \\ &- \sum_{i=1}^n \alpha_i^* (\xi_i^* + y_i - (\mathbf{w}^t \phi(\mathbf{u}_i) + b) \bar{\mathbf{W}}_i \mathbf{y} - \sum_{k=0}^d x_{ki} (\mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k)) \\ &- \sum_{i=1}^n \eta_i \xi_i - \sum_{i=1}^n \eta_i^* \xi_i^* - \sum_{i=1}^n v_i (\mathbf{w}^t \phi(\mathbf{u}_i) + b) - \sum_{i=1}^n v_i^* (1 - \mathbf{w}^t \phi(\mathbf{u}_i) - b), \end{aligned} \quad (28)$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*, v_i, v_i^*$ are the Lagrange multipliers.

For the case of $\rho \in [-1, 1]$, the Lagrange function is constructed as follows:

$$\begin{aligned} \mathcal{L} = \mathcal{J} &- \sum_{i=1}^n \alpha_i (\xi_i - y_i + (\mathbf{w}^t \phi(\mathbf{u}_i) + b) \bar{\mathbf{W}}_i \mathbf{y} + \sum_{k=0}^d x_{ki} (\mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k)) \\ &- \sum_{i=1}^n \alpha_i^* (\xi_i^* + y_i - (\mathbf{w}^t \phi(\mathbf{u}_i) + b) \bar{\mathbf{W}}_i \mathbf{y} - \sum_{k=0}^d x_{ki} (\mathbf{w}_k^t \phi(\mathbf{u}_i) + b_k)) \\ &- \sum_{i=1}^n \eta_i \xi_i - \sum_{i=1}^n \eta_i^* \xi_i^* - \sum_{i=1}^n v_i (1 + \mathbf{w}^t \phi(\mathbf{u}_i) + b) - \sum_{i=1}^n v_i^* (1 - \mathbf{w}^t \phi(\mathbf{u}_i) - b), \end{aligned} \quad (29)$$

where $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*, v_i, v_i^*$ are the Lagrange multipliers.

It is understood that the dual variables in (28) and (29) have to satisfy positivity constraints, i.e., $\alpha_i, \alpha_i^*, \eta_i, \eta_i^*, v_i, v_i^* \geq 0$. It follows from the saddle point condition that the partial

derivatives of $\mathcal{L}$ with respect to the primal variables ($\mathbf{w}, \mathbf{w}_k, b, b_k, \zeta_i, \zeta_i^*$) have to vanish for optimality.

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \mathbf{0} \rightarrow \mathbf{w} = \sum_{i=1}^n ((\alpha_i - \alpha_i^*) \bar{\mathbf{W}}_i \mathbf{y} + v_i - v_i^*) \phi(\mathbf{u}_i) \quad (30)$$

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}_k} = \mathbf{0} \rightarrow \mathbf{w}_k = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_{ki} \phi(\mathbf{u}_i), \quad k = 0, 1, \dots, d \quad (31)$$

$$\frac{\partial \mathcal{L}}{\partial b} = 0 \rightarrow \sum_{i=1}^n ((\alpha_i - \alpha_i^*) \bar{\mathbf{W}}_i \mathbf{y} + (v_i - v_i^*)) = 0 \quad (32)$$

$$\frac{\partial \mathcal{L}}{\partial b_k} = 0 \rightarrow \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_{ki} = 0, \quad k = 1, \dots, d \quad (33)$$

$$\frac{\partial \mathcal{L}}{\partial \zeta_i} = 0 \rightarrow C - \alpha_i - \eta_i = 0, \quad i = 1, \dots, n \quad (34)$$

$$\frac{\partial \mathcal{L}}{\partial \zeta_i^*} = 0 \rightarrow C - \alpha_i^* - \eta_i^* = 0, \quad i = 1, \dots, n \quad (35)$$

Classical Lagrangian duality enables the primal problem to be transformed to their dual problem. Substituting (30), (31), (32), (33), (34) and (35) into (28) and (29) yields two dual optimization problems, respectively. For the case of $\rho \in [0, 1]$, the dual optimization problem is obtained as follows:

$$\begin{aligned} \min_{\alpha_i^{(*)}, v_i^{(*)}} & \frac{1}{2} \sum_{i,j} (\bar{\mathbf{W}}_i \mathbf{y} (\alpha_i - \alpha_i^*) + v_i - v_i^*) K_{ij} (\bar{\mathbf{W}}_j \mathbf{y} (\alpha_j - \alpha_j^*) + v_j - v_j^*) \\ & + \frac{1}{2} \sum_{i,j} (\alpha_i - \alpha_i^*) x_{ki} K_{ij} x_{kj} (\alpha_j - \alpha_j^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) + \sum_{i=1}^n v_i^* \\ \text{subject to} & \begin{cases} 0 < \alpha_i < C, \quad 0 < \alpha_i^* < C \\ v_i \geq 0, \quad v_i^* \geq 0 \end{cases} \end{aligned} \quad (36)$$

For the case of $\rho \in [-1, 1]$, the dual optimization problem is obtained as follows:

$$\begin{aligned} \min_{\alpha_i^{(*)}, v_i^{(*)}} & \frac{1}{2} \sum_{i,j} (\bar{\mathbf{W}}_i \mathbf{y} (\alpha_i - \alpha_i^*) + v_i - v_i^*) K_{ij} (\bar{\mathbf{W}}_j \mathbf{y} (\alpha_j - \alpha_j^*) + v_j - v_j^*) \\ & + \frac{1}{2} \sum_{i,j} (\alpha_i - \alpha_i^*) x_{ki} K_{ij} x_{kj} (\alpha_j - \alpha_j^*) - \sum_{i=1}^n y_i (\alpha_i - \alpha_i^*) + \sum_{i=1}^n (v_i + v_i^*) \\ \text{subject to} & \begin{cases} 0 < \alpha_i < C, \quad 0 < \alpha_i^* < C \\ v_i \geq 0, \quad v_i^* \geq 0 \end{cases} \end{aligned} \quad (37)$$

Solving the above quadratic programming (QP) problem (36) or (37) with the constraints determines the optimal Lagrange multipliers $\hat{\alpha}_i, \hat{\alpha}_i^*$ and $\hat{v}_i, \hat{v}_i^*$. Thus, the estimated weight

vectors $\hat{\mathbf{w}}$ and $\hat{\mathbf{w}}_k$ are obtained, respectively as follows:

$$\hat{\mathbf{w}} = \sum_{i=1}^n ((\hat{\alpha}_i - \hat{\alpha}_i^*) \bar{\mathbf{W}}_i \mathbf{y} + \hat{v}_i - \hat{v}_i^*) \phi(\mathbf{u}_i) \quad (38)$$

$$\hat{\mathbf{w}}_k = \sum_{i=1}^n (\hat{\alpha}_i - \hat{\alpha}_i^*) \mathbf{x}_{ki} \phi(\mathbf{u}_i), \quad k = 0, 1, \dots, d \quad (39)$$

Thus, for a point $\mathbf{u}_i$ associated with the training data set $\mathcal{D}$, $\rho(\mathbf{u}_i)$ and $\beta_k(\mathbf{u}_i)$ are obtained, respectively as follows:

$$\hat{\rho}(\mathbf{u}_i) = (\text{Diag}\{\bar{\mathbf{W}}\mathbf{y}\})(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) + \hat{v} - \hat{v}^* \mathbf{K}_i + \hat{b} \quad (40)$$

$$\hat{\beta}_k(\mathbf{u}_i) = (\mathbf{K}_i^t \circ \mathbf{X}^k)^t (\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) + \hat{b}_k, \quad k = 0, 1, \dots, d, \quad (41)$$

where $\mathbf{K}_i$ is the i th row of the kernel matrix $\mathbf{K}$ whose elements are $\phi(\mathbf{u}_i)^t \phi(\mathbf{u}_j) = K(\mathbf{u}_i, \mathbf{u}_j)$, $\mathbf{X}^k$ is the k th column of the $n \times (d+1)$ design matrix $\mathbf{X}$, and $\circ$ is the Hadamard product.

Here, $\hat{b}$ and $\hat{b}_k$ can be determined by the linear regression with the input vector $(\bar{\mathbf{W}}_i \mathbf{y}, \mathbf{X}_i)^t$ and the output variable

$$y_i - (\text{Diag}\{\bar{\mathbf{W}}\mathbf{y}\})(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) + \hat{v} - \hat{v}^* \mathbf{K}_i - ((\mathbf{X}_i \mathbf{X}^t) \circ \mathbf{K}_i)(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*), \quad (42)$$

which is,

$$\begin{aligned} & y_i - (\text{Diag}\{\bar{\mathbf{W}}\mathbf{y}\})(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) + \hat{v} - \hat{v}^* \mathbf{K}_i - ((\mathbf{X}_i \mathbf{X}^t) \circ \mathbf{K}_i)(\hat{\boldsymbol{\alpha}} - \hat{\boldsymbol{\alpha}}^*) \\ &= (b, b_0, b_1, \dots, b_k)(\bar{\mathbf{W}}_i \mathbf{y}, \mathbf{X}_i^t)^t \quad \text{for } i \in I_{sv}, \end{aligned} \quad (43)$$

where $I_{sv} = \{i = 1, 2, \dots, n : 0 < |\alpha_i - \alpha_i^*| < C \text{ or } |v_i - v_i^*| > 0\}$ is obtained by exploiting Karush-Kuhn-Tucker conditions [23]. That is, the estimated values of $\hat{\mathbf{y}}$ is obtained as follows:

$$\hat{\mathbf{y}} = \text{Diag}\{\bar{\mathbf{W}}\mathbf{y}\} \hat{\boldsymbol{\rho}} + \sum_{k=0}^d \mathbf{x}_k \circ \hat{\boldsymbol{\beta}}_k, \quad (44)$$

where $\hat{\boldsymbol{\rho}} = (\hat{\rho}(\mathbf{u}_1), \dots, \hat{\rho}(\mathbf{u}_n))^t$ and $\hat{\boldsymbol{\beta}}_k = (\hat{\beta}_k(\mathbf{u}_1), \dots, \hat{\beta}_k(\mathbf{u}_n))^t$.

Model selection

The functional structure of the KBGTWAR model is characterized by the regularization parameter C , the kernel parameter κ and the number q of the nearest neighbors. These hyperparameters will affect the final model complexity. We now illustrate the model selection method which determines the optimal values of these hyperparameters of the KBGTWAR model. To choose these hyperparameters we utilize the AIC function which is defined according to [25] as follows:

$$AIC(\boldsymbol{\lambda}) = \frac{RSS(\boldsymbol{\lambda})}{n} + \frac{2d}{n} \hat{\sigma}^2(\boldsymbol{\lambda}), \quad (45)$$

where $\boldsymbol{\lambda}$ is the vector of hyperparameters, $RSS(\boldsymbol{\lambda}) = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, $\hat{\sigma}^2(\boldsymbol{\lambda})$ denotes the estimate of noise variance defined as $\hat{\sigma}^2(\boldsymbol{\lambda}) = \frac{1}{n-d} RSS(\boldsymbol{\lambda})$, and d is the number of free parameters, i.e., the size of $I_{sv} = \{i = 1, 2, \dots, n : 0 < |\alpha_i - \alpha_i^*| < C \text{ or } |v_i - v_i^*| > 0\}$.

Case study

This section illustrates the performance of the proposed KBGTWAR model using house price data collected from 2001 to 2008 in Shenzhen, China. This data set was firstly used in [18]. Shenzhen is a special economic zone, which is situated in Guangdong Province immediately north of Hong Kong Special Administrative Region. This city forms part of the Pearl River Delta megalopolis. There are six administrative districts in Shenzhen, which are Luohu, Nanshan, Futian, Yantian, Bao'an and Longgang. By the way, the last two districts are not situated in the Special Zone. See for further details [18]. The house prices of Shenzhen continue to increase at an alarming rate due to rapid industrialization and urbanization. The study data on house prices were provided by the Shenzhen municipal bureau of land resources and housing management. From the study area 406 observations are available.

[18] reported that for empirical study thirteen input variables were used, but only six of them were statistically significant at the 90% confidence level according to their t -probabilities. The six input variables used in this study are land area (LANDA), distance from the nearest major road (DROAD), quality (QUAL), land price (LPRICE), property management spend (MAGT), and proximity to bus (TRAFF). The input variables and (u, v, t) coordinate variables are standardized such as $(z - \min(z))/(\max(z) - \min(z))$. As in [26], a recent selling price is used as output variable, which stands as a proxy for the market value of the house. In fact, the logarithm of recent selling price is considered as output variable.

According to [18], there exist spatiotemporal autocorrelations in the output variable. Therefore, it is appropriate to use GTWAR-based models for this house price data set. For comparison, the global OLS model, the spatial autoregressive (SAR) model and three different GWR-based models including GWR, GTWAR and KBGTWAR are implemented using the same data set. The OLS and SAR models are employed to analyze the house price data by considering space coordinates and time as exogenous variables. GTWAR and KBGTWAR models are used to analyze the house price data with spatial and temporal considerations.

Using the cross validation (CV) technique, [18] determined that the optimal number of the nearest neighbors for GWR and GTWAR models is $q = 51$ and $q = 44$, respectively. In this paper we obtain $q = 53$ using the AIC technique for KBGTWAR model. As mentioned before, the spatiotemporal weights matrix $\bar{W}$ for KBGTWAR model is constructed using the spatiotemporal distance (26) instead of the spatiotemporal distance in [18]. Using this $\bar{W}$, we obtain the value of Moran's I for KBGTWAR model, which is 0.1105. This value indicates that spatiotemporal autocorrelation is positive and thus we need to use KBGTWAR model under the condition $\rho \in [0, 1]$. We examine parameter estimates for models under consideration and goodness of fit in terms of RSS and R^2 . For the case of $\rho \in [0, 1]$, the values of hyperparameters of KBGTWAR model are determined by the AIC method as $(C, \kappa, q) = (50, 0.05, 53)$. The results are reported in Tables 1 and 2. In fact, the estimate of ρ for GTWAR and KBGTWAR is the median of 406 estimated ρ_i 's. Both median values are very similar when $\rho \in [0, 1]$. Fig 1 reports the estimated $\rho \in [0, 1]$ values for 406 observed points. By comparing RSS and R^2 values, KBGTWAR gives significantly better fit of data than OLS, SAR, GWR and GTWAR models. The proposed KBGTWAR model provides the bigger F-statistic value than GWR and GTWAR models. Therefore, this indicates that it is more appropriate to model this particular data set with nonlinear local KBGTWAR model.

We now investigate the spatial and temporal variations of three selected parameters, i.e., autoregressive parameter ρ , DROAD and QUAL coefficients. Figs 2 and 3 show the results. The plots in Fig 2 show the effects of sales time and the geographical location on each individual parameter. As seen from Fig 2, on the whole the autoregressive parameter and QUAL coefficient values do not show the apparent spatial and temporal variations. However, DROAD

Table 1. Parameter estimate results of OLS, SAR and GWR for residential estate price data of Shenzhen.

Parameter	OLS	SAR ($q = 44, \rho \in [-1, 1]$)	GWR ($q = 51$)		
			Min	Med	Max
Intercept	8.0027	8.426	3.519	7.713	9.663
LANDA	0.4178	0.416	-4.750	0.236	5.011
DROAD	0.3627	0.361	-2.400	-0.310	2.928
QUAL	0.2469	0.246	-15.535	-0.304	19.686
LPRICE	1.5161	1.538	-3.734	2.125	12.279
MAGT	1.5734	1.574	-3.021	2.643	6.117
TRAFF	1.0397	1.042	-3.292	0.954	4.842
ρ		-0.0470			
RSS	54.7	54.7		37.2	
R^2	0.612	0.612		0.736	
F-statistic	69.38	69.38		1.63	

<https://doi.org/10.1371/journal.pone.0205063.t001>

Table 2. Parameter estimate results of GTWAR and KBGTWAR for residential estate price data of Shenzhen.

Parameter	GTWAR ($q = 44, \rho \in [0, 1]$)			KBGTWAR ($q = 53, \rho \in [0, 1]$)		
	Min	Med	Max	Min	Med	Max
Intercept	0.805	8.020	16.663	6.656	7.285	7.852
LANDA	-6.699	0.571	6.192	-0.470	1.159	5.773
DROAD	-7.237	-0.016	8.884	-4.052	0.102	3.123
QUAL	-10.280	-0.941	5.490	-4.987	0.126	2.305
LPRICE	-6.377	1.300	10.584	-2.123	1.476	7.198
MAGT	-9.794	1.776	8.753	-2.363	1.702	5.819
TRAFF	-5.589	1.102	9.991	-1.792	0.656	3.109
ρ		0.1101			0.1088	
RSS		12.0			4.5	
R^2		0.914			0.968	
F-statistic		3.84			4.45	

<https://doi.org/10.1371/journal.pone.0205063.t002>

Fig 1. Estimated spatial autoregressive parameter. The estimated spatial autoregressive parameter($\rho \in [0, 1]$) values for 406 observed points.

<https://doi.org/10.1371/journal.pone.0205063.g001>

Fig 2. Effects of sales time and the geographical location on each individual parameter. Autoregressive parameter ρ (top), DROAD (middle) and QUAL (bottom) coefficients on the sales time and the X, Y coordinates of the geographical location.

<https://doi.org/10.1371/journal.pone.0205063.g002>

coefficient show somewhat apparent spatial and temporal variations. In particular, this coefficient changes nonlinearly in the sales time and Y-coordinate. The maps in Fig 3 show the distribution of each individual coefficient. As seen from Fig 3, the autoregressive parameter and QUAL coefficient do not show apparent spatial variations. However, DROAD coefficient changes from low in the south to high in the north. From Figs 2 and 3 we observe that the effect of DROAD coefficient on the house price depends on the location and the sales time.

Conclusions

In this paper, we proposed the KBGTWAR model to simultaneously account for spatiotemporal nonstationarity and autocorrelation that exist in the house prices. To devise KBGTWAR model we applied kernel technique to spatially and temporally varying coefficients and then utilized the basic principle of linear SVM to estimate the relevant parameters. Unlike GTWAR model, KBGTWAR model is basically nonlinear. Therefore, KBGTWAR model can deal with complex nonlinear trends, which are very common in spatiotemporal phenomena. KBGTWAR model can also appropriately explore the dynamic relationship between the output variable and the input variables, since this model is based on varying coefficients model which efficiently describes dynamic patterns of a regression relationship.

KBGTWAR model takes over all advantages of SVM and varying coefficients model that capture nonlinearities in the data, that have good prediction ability, and that are useful tools when the functional form of the relationship between the output variable and the input variables is left unspecified. In particular, KBGTWAR model works well under settings without strong assumptions on the distribution of the data. KBGTWAR model is also interpretable, since varying coefficients model is meaningfully interpretable.

However, as with all SVM-related models, KBGTWAR model requires lots of computing time in determining the optimal hyperparameters, and has serious computational problem for large data because it has to solve a large-scale quadratic programming problem to get the values of relevant parameters. These are disadvantages associated with KBGTWAR model.

This paper analyzed data reflecting the spatiotemporal nonstationarity and autocorrelation of house prices. OLS, SAR, GWR, GTWAR and KBGTWAR models used in the study were built based on house price data collected between 2001 and 2008 in the city of Shenzhen, China. The performances of these models were then compared based on RSS, R^2 and F-statistic. This paper demonstrates that the proposed KBGTWAR model provides good results in

Fig 3. Spatial variation of each individual parameters. Autoregressive parameter (top), DROAD (middle) and QUAL (bottom) coefficients.

<https://doi.org/10.1371/journal.pone.0205063.g003>

goodness of fit for the given example. To conclude, we proposed a more efficient KBGTWAR model to account for spatiotemporal nonstationarity and autocorrelation simultaneously.

Acknowledgments

The authors wish to thank Professor Huang who provided us with house price data. This research was supported by Basic Science Research Program through the National Research

Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology with grant no. (NRF-2014R1A1A2054917, NRF-2018R1D1A1B07042349). This work was supported by “Human Resources Program in Energy Technology” of the Korea Institute of Energy Technology Evaluation and Planning (KETEP), granted financial resource from the Ministry of Trade, Industry & Energy, Republic of Korea (No. 20174030201740).

Author Contributions

Conceptualization: Jooyong Shim, Changha Hwang.

Formal analysis: Jooyong Shim, Changha Hwang.

Investigation: Changha Hwang.

Methodology: Changha Hwang.

Software: Jooyong Shim.

Supervision: Changha Hwang.

Validation: Jooyong Shim, Changha Hwang.

References

1. Cliff AD, Ord JK. Spatial processes: Models and applications. Pion, London; 1981.
2. Can A. Specification and estimation of hedonic housing price models. *Reg Sci Urban Econ*. 1992 Sept; 22(3): 453–474. [https://doi.org/10.1016/0166-0462\(92\)90039-4](https://doi.org/10.1016/0166-0462(92)90039-4)
3. Pace RK, Gilley OW. Generalizing the OLS and grid estimators. *Real Estate Econ*. 1998 June; 26(2): 331–347. <https://doi.org/10.1111/1540-6229.00748>
4. Matheron G. Principles of geostatistics. *Econ Geol*. 1963 Dec; 58(8): 1246–1266. <https://doi.org/10.2113/gsecongeo.58.8.1246>
5. Basu A, Thibodeau T. Analysis of spatial autocorrelation in house price. *J Real Estate Finance Econ*. 1998 Jul; 17(1): 61–85. <https://doi.org/10.1023/A:1007703229507>
6. Gelfand AE, Ghosh SK, Knight JR, Sirmans CF. Spatio-temporal modeling of residential sales data. *J Bus Econ Statist*. 1998 Jul; 16(3): 312–321. <https://doi.org/10.1080/07350015.1998.10524770>
7. Tu Y, Yu SM, Sun H. Transaction-based office price indexes: a spatiotemporal modeling approach. *Real Estate Econ*. 2004 May; 32(2): 297–328. <https://doi.org/10.1111/j.1080-8620.2004.00093.x>
8. McMillen DP. One hundred fifty years of land values in Chicago: a nonparametric approach. *J Urban Econ*. 1996 Jul; 40(1): 100–124. <https://doi.org/10.1006/juec.1996.0025>
9. McMillen DP, McDonald JF. A nonparametric analysis of employment density in a polycentric city. *J Reg Sci*. 1997 Nov; 37(4): 591–612. <https://doi.org/10.1111/0022-4146.00071>
10. Brunson C, Fotheringham AS, Charlton M. Geographically weighted regression: a method for exploring spatial nonstationarity. *Geogr Anal*. 1996 Oct; 28(4): 281–298. <https://doi.org/10.1111/j.1538-4632.1996.tb00936.x>
11. Fotheringham AS, Charlton ME, Brunson C. The geography of parameter space: an investigation of spatial non-stationarity. *Int J Geogr Inf Sci*. 1996 Jul; 10(5): 605–627. <https://doi.org/10.1080/02693799608902100>
12. Fotheringham AS, Brunson C, Charlton M. Geographically weighted regression. John Wiley and Sons, Chichester; 2002.
13. Brunson C, Fotheringham AS, Charlton M. Some notes on parametric significance tests for geographically weighted regression. *J Reg Sci*. 1999 Aug; 39(3): 497–524. <https://doi.org/10.1111/0022-4146.00146>
14. Huang B, Wu B, Barry T. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int J Geogr Inf Sci*. 2010 Mar; 24(3): 383–401. <https://doi.org/10.1080/13658810802672469>
15. Fotheringham AS, Crespo R, Yao J. Geographical and temporal weighted regression (GTWR). *Geogr Anal*. 2015 Oct; 47(4): 431–452. <https://doi.org/10.1111/gean.12071>

16. Liu J, Yang Y, Xu S, Zhao Y, Wang Y, Zhang F. A Geographically Temporal Weighted Regression Approach with Travel Distance for House Price Estimation. *Entropy*. 2016 Aug; 18(8): 303–312. <https://doi.org/10.3390/e18080303>
17. LeSage JP. The theory and practice of spatial econometrics. preprint (1999). Available at <http://www.spatial-econometrics.com>.
18. Wu B, Li R, Huang B. A geographically and temporally weighted autoregressive model with application to housing prices. *Int J Geogr Inf Sci*. 2014 Apr; 28(5): 1186–1204. <https://doi.org/10.1080/13658816.2013.878463>
19. Vapnik VN. The nature of statistical learning theory. Springer, New York; 1995.
20. Can A, Megbolugbe I. Spatial dependence and house price index construction. *J Real Estate Finance Econ*. 1997 Jan; 14(1): 203–222. <https://doi.org/10.1023/A:1007744706720>
21. Hurvich CM, Simonoff JS, Tsai CL. Smoothing parameter selection in nonparametric regression using an improved Akaike Information Criterion. *J R Stat Soc Series B*. 1998 Apr; 60(2): 271–293. <https://doi.org/10.1111/1467-9868.00125>
22. Smola AJ, Schölkopf B. A tutorial on support vector regression. *Stat Comput*. 2004 Aug; 14(3): 199–222. <https://doi.org/10.1023/B:STCO.0000035301.49549.88>
23. Kuhn HW, Tucker AW. Nonlinear programming. *Proceedings of 2nd Berkeley Symposium*, University of California Press, Berkeley; 1951.
24. Mercer J. Function of positive and negative type and their connection with theory of integral equations. *Philosophical Transactions of Royal Society A* 1909; 415–446.
25. Cherkassky V, Ma Y. Comparison of model selection for regression. *Neural Comput*. 2003 Jul; 15(7): 1691–1714. <https://doi.org/10.1162/089976603321891864> PMID: 12816572
26. Sirmans G, Macpherson D, Zietz A. The composition of hedonic pricing models. *J. Real Estate Lit*. 2005 Jan; 13(1): 3–41.