JOURNAL OF TIME SERIES ANALYSIS
*J. Time Ser. Anal.* **40**: 327–342 (2019)
Published online 18 December 2018 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1111/jtsa.12442

SPECIAL ISSUE

# ON A SEMIPARAMETRIC DATA-DRIVEN NONLINEAR MODEL WITH PENALIZED SPATIO-TEMPORAL LAG INTERACTIONS

DAWLAH AL-SULAMI,[a] ZHENYU JIANG,[b] ZUDI LU[c] AND JUN ZHU[d*]

[a] *Department of Statistics, King Abdulaziz University Jeddah, Saudi Arabia*
[b] *Statistical Sciences Research Institute, University of Southampton, Southampton, UK*
[c] *Statistical Sciences Research Institute and School of Mathematical Sciences, University of Southampton, Southampton, UK*
[d] *Department of Statistics and Department of Entomology, University of Wisconsin-Madison, Madison, WI, USA*

To study possibly nonlinear relationship between housing price index (HPI) and consumer price index (CPI) for individual states in the USA, accounting for the temporal lag interactions of the housing price in a given state and spatio-temporal lag interactions between states could improve the accuracy of estimation and forecasting. There lacks, however, methodology to objectively identify and estimate such spatio-temporal lag interactions. In this article, we propose a semiparametric data-driven nonlinear time series regression method that accounts for lag interactions across space and over time. A penalized procedure utilizing adaptive Lasso is developed for the identification and estimation of important spatio-temporal lag interactions. Theoretical properties for our proposed methodology are established under a general near epoch dependence structure and thus the results can be applied to a variety of linear and nonlinear time series processes. For illustration, we analyze the US housing price data and demonstrate substantial improvement in forecasting via the identification of nonlinear relationship between HPI and CPI as well as spatio-temporal lag interactions.

*Received 14 October 2017; Accepted 7 November 2018*

Keywords: adaptive Lasso; asymptotic property; nonlinear regression; regularization; spatial statistics; time series

MOS subject classification: 62M10; 62M30.

## 1. INTRODUCTION

Of broad practical interest are nonlinear relationships between covariate variables and a response variable with spatio-temporal effects for not necessarily Gaussian data. For example, for a given state in the United States , ignoring the temporal lag interactions of the housing price return in this state and between states may result in biased estimates of possibly nonlinear effect of the consumer price index (CPI) change on the housing price index (HPI) return (see Section 5 for more detail). Such spatio-temporal lag interactions in a nonlinear regression setting with irregularly spaced time series data are not well understood in the literature (Al-Sulami *et al.*, 2017). The purpose of this article is to develop a semiparametric data-driven nonlinear time series regression method that objectively selects spatio-temporal lag interactions based on data.

It is well known that too many unnecessary predictors may give inefficient estimation and prediction and therefore, selecting the more important predictors among a large number of predictors is of keen interest. Variable selection procedures such as stepwise selection, Akaike's information criterion (AIC) (Akaike, 1973) and Bayesian information criterion (BIC) (Schwarz, 1978) are extensively employed to choose the appropriate covariates in linear regression as well as the appropriate lag order in time series analysis (see, e.g., Ramanathan, 1992; McQuarrie and Tsai, 1998; Shumway and Stoffer, 2000). These traditional methods have a number of drawbacks including instability (Breiman, 1996), as the estimation and variable selection are executed in separate steps, and the stochastic error is not taken into account in the variable selection step (Fan and Li, 2001).

* Correspondence to: Jun Zhu, Department of Statistics and Department of Entomology, University of Wisconsin-Madison, 1300 University Avenue, Madison, WI 53706, USA. E-mail: jzhu@stat.wisc.edu

To overcome the limitations of the traditional variable selection methods and enhance the prediction accuracy, a variety of penalized methods have been developed for linear regression and gained popularity. See, for example, bridge regression (Frank and Friedman, 1993), least absolute shrinkage and selection operator (Lasso) (Tibshirani, 1996), and elastic net (Zou and Hastie, 2005). By applying an $\mathcal{L}_1$-penalty function with a single regularization parameter, Lasso produces a sparse model and yields a consistent estimator under appropriate conditions. The necessary conditions to achieve the consistency were explored by Zou (2006) and Zhao and Yu (2007). Fan and Li (2001) proposed a penalized likelihood approach with a set of penalty functions including Lasso as a special case, and showed that penalized likelihood with a smoothly clipped absolute deviation (SCAD) penalty function enhances the performance in model selection. Zou (2006) developed adaptive Lasso by using an $\mathcal{L}_1$-penalty function which assigns different weights to different interactions, and demonstrated that adaptive Lasso enjoys *oracle* properties and ease of applications. Variable selection techniques have also been developed for generalized linear models, survival data, and time series data (see, e.g., Tibshirani, 1997; Wang *et al.*, 2007; Van De Geer, 2008; Hsu *et al.*, 2008; Haufe *et al.*, 2009; Ren and Zhang, 2010).

For spatial lattice data, Zhu *et al.* (2010) applied adaptive Lasso to select covariates and spatial neighborhoods in a spatial linear regression model (see also, Huang *et al.*, 2010). Reyes *et al.* (2012) further developed adaptive Lasso for the selection of covariates and spatio-termporal coefficients. However, the spatial and spatio-termporal neighborhood structures considered in Zhu *et al.* (2010) and Reyes *et al.* (2012) are relatively simple (see also Hallin *et al.*, 2004; Gao *et al.*, 2006; Lu *et al.*, 2007, 2009). Bayesian model selection has also been considered for spatial data such as Song and De Oliveira (2012) and Hefley *et al.* (2017).

In spatial econometrics, a spatial weight matrix measures possibly complex spatial interactions between spatial locations or spatial units on a lattice and is crucial to construct when applying a spatial or spatio-temporal model (Anselin, 1988). The influence of the construction of a spatial weight matrix on both model testing and parameter estimation has been documented by many (see, e.g., Stetzer, 1982; Griffith and Lagona, 1998; Smith, 2009; Stakhovych and Bijmolt, 2009). In many applications, however, the spatial weight matrix is assumed to be known *a priori*, which may or may not reflect the true underlying spatial interaction. Thus considerable attention has been devoted to the estimation of the spatial weight matrix in various settings with low- or high-dimensional spatial panels in recent years. For example, estimation methods were proposed for the spatial weight matrix in low-dimensional spatial context under different structural constraints (Bhattacharjee and Holly, 2011, 2013; Bhattacharjee and Jensen-Butler 2013). de Souza (2012) proposed a spatial autoregressive model with exogenous covariates and estimated the spatial weight matrix by Lasso, but required *a priori* knowledge about the structure of the spatial interactions. Lam and Souza (2013) developed a spatial lag model with exogenous covariates and estimated the spatial weight matrix together with the model parameters via adaptive Lasso, whereas Manresa (2013) considered a non-autoregressive spatial model with exogenous covariates and estimated the spatial weight matrix by a pooled Lasso technique. Ahrens and Bhattacharjee (2015) proposed a spatial autoregressive model with exogenous covariates and developed a two-step Lasso method to estimate the spatial weight matrix, which was demonstrated to be effective to uncover the underlying spatial dependence structure via a simulation study. All of these methods above in spatial econometrics, however, focused on the estimation of the spatial weight matrix for linear models under spatial stationarity.

In contrast to the research development above, here we construct and estimate the spatio-temporal weights for nonlinear models with non-Gaussian errors. In our previous work (Al-Sulami *et al.*, 2017), we considered a partially nonlinear regression model for spatial time series. However, the neighborhood structure was pre-specified and thus can be quite subjective. Here, we develop a penalized method to simultaneously identify and estimate the spatio-temporal lag interactions in the setting of a semiparametric data-driven nonlinear model (see, e.g., Zhang *et al.*, 2003; Lu *et al.*, 2008, 2009). Although the theory for non-parametric or nonlinear estimation is well established, here we consider the adaptive Lasso for selecting lagged variables in time series and space-time model under a general near epoch dependence structure, which can be applied to a wide range of linear and nonlinear time series processes (c.f., Lu and Linton, 2007; Li *et al.*, 2012).

The rest of this article is organized as follows. In Sections 2 and 3, we describe our proposed model and estimation procedure. In Section 4, the asymptotic properties for the estimation procedure are established. For illustration,

  
DOI: 10.1111/jtsa.12442

we apply our method to explore possible nonlinear relationship between housing price and consumer price in selected states of the USA in Section 5. Concluding remarks are given in Section 6. Technical details including proofs of theorems are relegated to Appendix S1, Supporting Information.

## 2. MODEL

Let $Y_t(\mathbf{s}_0)$ and $X_t(\mathbf{s}_0)$ denote two time series which are observed at discrete time points $t = 1, \ldots, T$, at a given spatial location $\mathbf{s}_0 = (u_0, v_0) \in \mathbb{R}^2$, where $u_0$ and $v_0$ are the $x$ and $y$ coordinates respectively, representing a spatial unit on a lattice. In the US housing price data example, $u_0$ and $v_0$ are the latitude and longitude of the centroid of a given state. Furthermore, $Y_t(\mathbf{s}_0)$ is a univariate response variable and $X_t(\mathbf{s}_0)$ is the covariate vector of dimension $d$, where $X_t$ can be the same for different $\mathbf{s}_0$ while $d$ should be small to avoid the curse of dimensionality. A possibly nonlinear relationship between the response $Y_t(\mathbf{s}_0)$ and the covariate $X_t(\mathbf{s}_0)$ is of interest.

For a given spatial unit represented by the spatial location $\mathbf{s}_0$, we may have $N$ other spatial units represented by spatial locations on a possibly irregular lattice, say $\mathbf{s}_k := (u_k, v_k) \in \mathbb{R}^2$ for $k = 1, \ldots, N$, over which $Y_t(\mathbf{s}_k)$ are observed at the $T$ time points $t = 1, \ldots, T$ affecting $Y_t(\mathbf{s}_0)$. At a given spatial location $\mathbf{s}_0$, we consider a semiparametric nonlinear regression time series model written as,

$$Y_t(\mathbf{s}_0) = g_0(X_t(\mathbf{s}_0)) + \sum_{i=1}^{p} \sum_{k=1}^{N} \lambda_{0k,i} Y_{t-i}(\mathbf{s}_k) + \sum_{l=1}^{q} \alpha_{0,l} Y_{t-l}(\mathbf{s}_0) + \varepsilon_t(\mathbf{s}_0) \tag{1}$$

where $t = r+1, \ldots, T$ with $r = \max(p, q)$ and $g_0(\cdot)$ is an unknown function characterizing the relationship between $X_t(\mathbf{s}_0)$ and $Y_t(\mathbf{s}_0)$. Furthermore, $\lambda_{0k,i}$ are unknown coefficients representing the spatio-temporal lag interactions of orders $p$ between spatial units, while $\alpha_{0,l}$ are unknown coefficients representing the temporal lag interactions of order $q$ for the given spatial unit $\mathbf{s}_0$. Finally, $\varepsilon_t(\mathbf{s}_0)$ are i.i.d. errors, not necessarily Gaussian, with mean $E\varepsilon_t(\mathbf{s}_0) = 0$ and variance $E\varepsilon_t^2(\mathbf{s}_0) = \sigma_0^2$.

Model (1) can be viewed as an extension of the model in Al-Sulami *et al.* (2017) at a given spatial location $\mathbf{s}_0$. Unlike Al-Sulami *et al.* (2017), however, here the lag effects between two spatial units $\lambda_{0k,i}$ are assumed to be unknown and depend on the time lag $i$ in the characterization of spatio-temporal lag interactions. In fact, when $\lambda_{0k,i} = \lambda_i(\mathbf{s}_0) w_{0k}$ with unknown $\lambda_i(\mathbf{s}_0)$ but known $w_{0k}$ such that $\sum_{k=1}^{N} w_{0k} = 1$, our model (1) reduces to the partially nonlinear model of Al-Sulami *et al.* (2017) at the spatial location $\mathbf{s}_0$.

## 3. ESTIMATION

How well the unknown function $g_0(\cdot)$ is estimated relies on the selection and estimation of the unknown spatio-temporal lag interactions $\lambda_{0k,i}$ and $\alpha_{0,l}$. A challenging issue here is that the unknown vector $\lambda(\mathbf{s}_0)$, consisting of all the spatio-temporal lag interactions $\lambda_{0k,i}$, has $Np$ unknown components, which can be quite large. For example, in the US housing price data example, we have $N = 50$. With say $p = 6$ time lags, the dimension of $\lambda(\mathbf{s}_0)$ is 300. In addition, for different time lags $i$, the number of non-zero interactions (i.e., number of $\mathbf{s}_k$'s with $\lambda_{0k,i} \neq 0$) for $\mathbf{s}_0$ can be different. Thus, the estimation of model (1) is more challenging than that of Al-Sulami *et al.* (2017) with pre-specified spatial weights at a fixed location.

Before giving the estimation detail, we introduce some notation. At a given spatial location $\mathbf{s}_0$, let $\boldsymbol{\eta}(\mathbf{s}_0) = \left(\lambda(\mathbf{s}_0)', \boldsymbol{\alpha}(\mathbf{s}_0)'\right)'$, with $\lambda(\mathbf{s}_0) = \left(\lambda_1(\mathbf{s}_0)', \lambda_2(\mathbf{s}_0)', \ldots, \lambda_N(\mathbf{s}_0)'\right)'$, $\lambda_k(\mathbf{s}_0) = \left(\lambda_{0k,1}, \lambda_{0k,2}, \ldots, \lambda_{0k,p}\right)'$, and $\boldsymbol{\alpha}(\mathbf{s}_0) = \left(\alpha_{0,1}, \alpha_{0,2}, \ldots, \alpha_{0,q}\right)'$. Moreover, let $\mathbf{Z}_t(\mathbf{s}_0) = \left(\mathbf{Z}_{1,t}(\mathbf{s}_0)', \mathbf{Z}_{2,t}(\mathbf{s}_0)'\right)'$, where $\mathbf{Z}_{1,t}(\mathbf{s}_0) = \left(\mathbf{Y}_{t-1}(\mathbf{s}_1)', \mathbf{Y}_{t-1}(\mathbf{s}_2)', \ldots, \mathbf{Y}_{t-1}(\mathbf{s}_N)'\right)'$, with $\mathbf{Y}_{t-1}(\mathbf{s}_k) = \left(Y_{t-1}(\mathbf{s}_k), Y_{t-2}(\mathbf{s}_k), \ldots, Y_{t-p}(\mathbf{s}_k)\right)'$, and $\mathbf{Z}_{2,t}(\mathbf{s}_0) = \left(Y_{t-1}(\mathbf{s}_0), Y_{t-2}(\mathbf{s}_0), \ldots, Y_{t-q}(\mathbf{s}_0)\right)'$. Then model (1) can be written more succinctly as

$$Y_t(\mathbf{s}_0) = g_0\left(X_t(\mathbf{s}_0)\right) + \mathbf{Z}_{1,t}(\mathbf{s}_0)'\lambda(\mathbf{s}_0) + \mathbf{Z}_{2,t}(\mathbf{s}_0)'\boldsymbol{\alpha}(\mathbf{s}_0) + \varepsilon_t(\mathbf{s}_0). \tag{2}$$

  
DOI: 10.1111/jtsa.12442

By taking conditional expectation of the terms in (2), we have

$$\begin{aligned} g_0\left(X_t(\mathbf{s}_0)\right) &= E\left[Y_t(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right] - E\left[\mathbf{Z}_{1,t}(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right]'\lambda(\mathbf{s}_0) - E\left[\mathbf{Z}_{2,t}(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right]'\boldsymbol{\alpha}(\mathbf{s}_0) \\ &\equiv g_1\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) - \mathbf{g}_{21}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right)'\lambda(\mathbf{s}_0) - \mathbf{g}_{22}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right)'\boldsymbol{\alpha}(\mathbf{s}_0), \end{aligned} \tag{3}$$

where $g_1\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) \in \mathbb{R}^1$, $\mathbf{g}_{21}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) \in \mathbb{R}^{Np}$ and $\mathbf{g}_{22}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) \in \mathbb{R}^q$ denote the conditional means of $E\left[Y_t(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right]$, $E\left[\mathbf{Z}_{1,t}(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right]$ and $E\left[\mathbf{Z}_{2,t}(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right]$ respectively, at the spatial location $\mathbf{s}_0$.

### 3.1. Estimation of $g_0\left(X_t(\mathbf{s}_0)\right)$

To estimate $g_0\left(X_t(\mathbf{s}_0)\right)$, we first consider estimating the three conditional means in (3), $g_1\left(x,\mathbf{s}_0\right)$, $\mathbf{g}_{21}\left(x,\mathbf{s}_0\right)$, and $\mathbf{g}_{22}\left(x,\mathbf{s}_0\right)$ at $X_t(\mathbf{s}_0) = x$. Define $a_0 = a_0(x,\mathbf{s}_0) = g_1(x,\mathbf{s}_0)$ and let $a_1 = a_1(x,\mathbf{s}_0) = \dot{g}_1(x,\mathbf{s}_0)$ denote the first-order derivative of $g_1$ with respect to $x$. By local linear fitting (Fan and Gijbels, 1996; Hallin *et al.*, 2004), we obtain the estimators $\hat{a}_0 = \hat{g}_1(x,\mathbf{s}_0)$ and $\hat{a}_1$ by

$$(\hat{a}_0,\hat{a}_1)' = \underset{(a_0,a_1)'\in\mathbb{R}^2}{\operatorname{argmin}} \sum_{t=r+1}^{T} \{Y_t(\mathbf{s}_0) - a_0 - a_1(X_t(\mathbf{s}_0)-x)\}^2 K_b(X_t(\mathbf{s}_0)-x), \tag{4}$$

where $T_0 = T - r$ is the effective sample size with $r = \max\{p,q\}$, $b = b_{T_0}$ is a bandwidth tending to zero as $T \to \infty$, $K(\cdot)$ is a bounded kernel function, and $K_b(\cdot) = b^{-1}K(\cdot/b)$.

Let $\mathbf{A}(x)$ be a $T_0\times 2$ matrix with the $(t-r)$th-row $\left(1, b^{-1}\left(X_t(\mathbf{s}_0)-x\right)\right)$ for $t = r+1,\dots,T$, and let $\mathbf{B}(\mathrm{x})$ be a $T_0\times T_0$ diagonal matrix with the $t$th diagonal element $K_b\left(X_t(\mathbf{s}_0)-x\right)$ for $t = r+1,\dots,T$. Let $\mathbf{Y} = \left(Y_{r+1}(\mathbf{s}_0),\dots,Y_T(\mathbf{s}_0)\right)'$ denote a $T_0$-dimensional vector of responses. Then the local linear estimator is given by

$$\left(\hat{a}_0, b\hat{a}_1\right)' = \mathbf{U}_{T_0}^{-1}\mathbf{V}_{T_0},$$

where $\mathbf{U}_{T_0} = \mathbf{A}(x)'\mathbf{B}(x)\mathbf{A}(x) = \begin{pmatrix} u_{T_0,00} & u_{T_0,01} \\ u_{T_0,10} & u_{T_0,11} \end{pmatrix}$ is a $2\times 2$ matrix and $\mathbf{V}_{T_0} = \mathbf{A}(x)'\mathbf{B}(x)\mathbf{Y} = \left(v_{T_0,0}, v_{T_0,1}\right)'$ is a $2\times 1$ vector. By denoting $\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right)^0 = 1$, we have

$$\begin{aligned} u_{T_0,jk} &= (T_0 b)^{-1} \sum_{t=r+1}^{T} \left(\frac{X_t(\mathbf{s}_0)-x}{b}\right)^j \left(\frac{X_t(\mathbf{s}_0)-x}{b}\right)^k K\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right),\ j,k = 0,1 \\ v_{T_0,j} &= (T_0 b)^{-1} \sum_{t=r+1}^{T} Y_t(\mathbf{s}_0) \left(\frac{X_t(\mathbf{s}_0)-x}{b}\right)^j K\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right),\ j = 0,1. \end{aligned}$$

Thus, with $\mathbf{e}_1 = (1,0)' \in \mathbb{R}^2$, the local linear estimator of $g_1(x,\mathbf{s}_0)$ is

$$\hat{g}_1(x,\mathbf{s}_0) = \hat{a}_0 = \mathbf{e}_1'\mathbf{U}_{T_0}^{-1}\mathbf{V}_{T_0}. \tag{5}$$

Similarly, we estimate $\mathbf{g}_{21}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) \in \mathbb{R}^{Np}$ with

$$\mathbf{g}_{21}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) = \left((g_{21}^{1,k}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right),\dots,g_{21}^{p,k}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right))' : k = 1,\dots,N\right),$$

  
DOI: 10.1111/jtsa.12442

where $g_{21}^{i,k}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) = E\left[Y_{t-i}(\mathbf{s}_k)|X_t(\mathbf{s}_0)\right]$. Let $\mathbf{Z}_1^{i,k} = \left(Y_{(r+1)-i}(\mathbf{s}_k),\dots,Y_{T-i}(\mathbf{s}_k)\right)'$ be a $T_0 \times 1$ vector and $\mathbf{R}_{1T_0}^{i,k} = \mathbf{A}(x)'\mathbf{B}(x)\mathbf{Z}_1^{i,k} = \left(\mathbf{r}_{1T_0,0}^{i,k},\mathbf{r}_{1T_0,1}^{i,k}\right)'$ a $2\times 1$ vector, where

$$\mathbf{r}_{1T_0,j}^{i,k} = (T_0 b)^{-1}\sum_{t=r+1}^{T} Y_{t-i}(\mathbf{s}_k)\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right)^j K\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right);\ j=0,1.$$

Then the local linear estimator of $g_{21}^{i,k}\left(x,\mathbf{s}_0\right)$ is

$$\hat{g}_{21}^{i,k}(x,\mathbf{s}_0) = \mathbf{e}_1'\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{i,k}. \tag{6}$$

Therefore, the local linear estimator of the unknown vector $\mathbf{g}_{21}\left(x,\mathbf{s}_0\right)$ is given by

$$\hat{\mathbf{g}}_{21}\left(x,\mathbf{s}_0\right) = \mathbf{e}_1'\left(\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{1,1},\dots,\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{p,1},\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{1,2},\dots,\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{p,2},\dots,\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{1,N},\dots,\mathbf{U}_{T_0}^{-1}\mathbf{R}_{1T_0}^{p,N}\right)'.$$

Furthermore, we estimate $\mathbf{g}_{22}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) = \left(g_{22}^1\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right),\dots,g_{22}^q\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right)\right)'$ with

$$g_{22}^l\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) = E\left[Y_{t-l}(\mathbf{s}_0)|X_t(\mathbf{s}_0)\right],$$

for $l = 1,\dots,q$. Let $\mathbf{Z}_2^l = \left(Y_{(r+1)-l}(\mathbf{s}_0),\dots,Y_{T-l}(\mathbf{s}_0)\right)'$ be a $T_0\times 1$ vector and let $\mathbf{R}_{2T_0}^l = \mathbf{A}(x)'\mathbf{B}(x)\mathbf{Z}_2^l = \left(\mathbf{r}_{2T_0,0}^l,\mathbf{r}_{2T_0,1}^l\right)'$ be a $2\times 1$ vector, where

$$\mathbf{r}_{2T_0,j}^l = (T_0 b)^{-1}\sum_{t=r+1}^{T} Y_{t-l}(\mathbf{s}_0)\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right)^j K\left(\frac{X_t(\mathbf{s}_0)-x}{b}\right), j=0,1.$$

Then the local linear estimator of $g_{22}^l\left(x,\mathbf{s}_0\right)$ is

$$\hat{g}_{22}^l(x,\mathbf{s}_0) = \mathbf{e}_1'\mathbf{U}_{T_0}^{-1}\mathbf{R}_{2T_0}^l. \tag{7}$$

Therefore, the local linear estimator of the unknown vector $\mathbf{g}_{22}\left(x,\mathbf{s}_0\right)$ is given by

$$\hat{\mathbf{g}}_{22}\left(x,\mathbf{s}_0\right) = \left(\hat{g}_{22}^1(x,\mathbf{s}_0),\dots,\hat{g}_{22}^q(x,\mathbf{s}_0)\right)' = \left(\mathbf{e}_1'\mathbf{U}_{T_0}^{-1}\mathbf{R}_{2T_0}^1,\dots,\mathbf{e}_1'\mathbf{U}_{T_0}^{-1}\mathbf{R}_{2T_0}^q\right)'.$$

Finally, by (3), (5)–(7), and given both $\boldsymbol{\lambda}(\mathbf{s}_0)$ and $\boldsymbol{\alpha}(\mathbf{s}_0)$, the unknown function $g_0\left(X_t(\mathbf{s}_0)\right)$ can be estimated by

$$\hat{g}_0\left(X_t(\mathbf{s}_0)\right) = \hat{g}_1\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right) - \hat{\mathbf{g}}_{21}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right)'\lambda(\mathbf{s}_0) - \hat{\mathbf{g}}_{22}\left(X_t(\mathbf{s}_0),\mathbf{s}_0\right)'\boldsymbol{\alpha}(\mathbf{s}_0). \tag{8}$$

### 3.2. Estimation of Parameters

Replacing $g_0\left(X_t(\mathbf{s}_0)\right)$ in (2) by its estimator (8), we re-write model (2) as

$$\hat{Y}_t(\mathbf{s}_0) = \hat{\mathbf{Z}}_t(\mathbf{s}_0)'\boldsymbol{\eta}(\mathbf{s}_0) + \varepsilon_t(\mathbf{s}_0),$$

  
DOI: 10.1111/jtsa.12442

where $\hat{Y}_t(\mathbf{s}_0) = Y_t(\mathbf{s}_0) - \hat{g}_1\left(X_t(\mathbf{s}_0), \mathbf{s}_0\right)$, $\hat{\mathbf{Z}}_t(\mathbf{s}_0) = \mathbf{Z}_t(\mathbf{s}_0) - \hat{\mathbf{g}}_2\left(X_t(\mathbf{s}_0), \mathbf{s}_0\right)$ and $\boldsymbol{\eta}(\mathbf{s}_0) = \left(\boldsymbol{\lambda}(\mathbf{s}_0)', \boldsymbol{\alpha}(\mathbf{s}_0)'\right)'$, $\mathbf{Z}_t(\mathbf{s}_0) = \left(\mathbf{Z}_{1,t}(\mathbf{s}_0)', \mathbf{Z}_{2,t}(\mathbf{s}_0)'\right)'$ and $\hat{\mathbf{g}}_2\left(X_t(\mathbf{s}_0), \mathbf{s}_0\right) = \left(\hat{\mathbf{g}}_{21}\left(X_t(\mathbf{s}_0), \mathbf{s}_0\right)', \hat{\mathbf{g}}_{22}\left(X_t(\mathbf{s}_0), \mathbf{s}_0\right)'\right)'$.

Following the idea of adaptive Lasso, we estimate the interactions $\boldsymbol{\lambda}(\mathbf{s}_0)$ and $\boldsymbol{\alpha}(\mathbf{s}_0)$ at the spatial location $\mathbf{s}_0$ by minimizing the following penalized sum of squared errors

$$Q(\boldsymbol{\eta}(\mathbf{s}_0)) = L(\boldsymbol{\eta}(\mathbf{s}_0)) + T_0 \sum_{k=1}^{N} \sum_{i=1}^{p} \gamma_i^k(\mathbf{s}_0)|\lambda_{0k,i}| + T_0 \sum_{l=1}^{q} \beta_l(\mathbf{s}_0)|\alpha_{0l}|, \tag{9}$$

where $L(\boldsymbol{\eta}(\mathbf{s}_0)) = \sum_{t=r+1}^{T} \hat{\varepsilon}_t(\mathbf{s}_0)^2$ is the sum of squared errors with $\hat{\varepsilon}_t(\mathbf{s}_0) = \hat{Y}_t(\mathbf{s}_0) - \hat{\mathbf{Z}}_t(\mathbf{s}_0)'\boldsymbol{\eta}(\mathbf{s}_0)$. The last two terms in (9) are adaptive Lasso penalties with regularization parameters $\{\gamma_i^k(\mathbf{s}_0)\}_{i=1}^{p}{}_{k=1}^{N}$ and $\{\beta_l(\mathbf{s}_0)\}_{l=1}^{q}$. Thus, the penalized parameter estimator of $\boldsymbol{\eta}(\mathbf{s}_0)$ is

$$\hat{\boldsymbol{\eta}}(\mathbf{s}_0) = \underset{\boldsymbol{\eta} \in \mathbb{R}^{Np+q}}{\operatorname{argmin}} \left\{ Q(\boldsymbol{\eta}(\mathbf{s}_0)) \right\}, \tag{10}$$

where the objective function $Q\left(\boldsymbol{\eta}(\mathbf{s}_0)\right)$ is given by (9).

To estimate the regularization parameters $\{\gamma_i^k(\mathbf{s}_0)\}_{i=1}^{p}{}_{k=1}^{N}$ for $k = 1, \ldots, N$ and $\{\beta_l(\mathbf{s}_0)\}_{l=1}^{q}$ in (9) with $(Np + q)$ parameters, we let

$$\gamma_i^k(\mathbf{s}_0) = \gamma(\mathbf{s}_0)\frac{\log(T_0)}{T_0 \mid \tilde{\lambda}_i^k(\mathbf{s}_0) \mid} \text{ and } \beta_l(\mathbf{s}_0) = \beta(\mathbf{s}_0)\frac{\log(T_0)}{T_0 \mid \tilde{\alpha}_l(\mathbf{s}_0) \mid}, \tag{11}$$

for $k = 1, \ldots, N, i = 1, \ldots, p, j = 1, \ldots, q$, where $\tilde{\lambda}_i^k(\mathbf{s}_0)$ and $\tilde{\alpha}_l(\mathbf{s}_0)$ are the initial least squares estimators obtained by minimizing the objective function $Q\left(\boldsymbol{\eta}(\mathbf{s}_0)\right)$ without penalty (Wang *et al.*, 2007; Zhu *et al.*, 2010). That is, $\tilde{\lambda}_i^k(\mathbf{s}_0)$ and $\tilde{\alpha}_l(\mathbf{s}_0)$ are the minimizers of $\sum_{t=r+1}^{T}\left\{\hat{Y}_t(\mathbf{s}_0) - \hat{\mathbf{Z}}_t(\mathbf{s}_0)'\boldsymbol{\eta}(\mathbf{s}_0)\right\}^2$. In the case of multicollinearity, $\tilde{\lambda}_i^k(\mathbf{s}_0)$ and $\tilde{\alpha}_l(\mathbf{s}_0)$ could be the ridge regression estimators (Zou, 2006). By (11), we reduce from $(Np + q)$ to just two regularization parameters, $\gamma(\mathbf{s}_0)$ and $\beta(\mathbf{s}_0)$. We select $\gamma(\mathbf{s}_0)$ and $\beta(\mathbf{s}_0)$ by a BIC,

$$\text{BIC}\left(\gamma(\mathbf{s}_0), \beta(\mathbf{s}_0)\right) = \log\left(\hat{\sigma}^2(\mathbf{s}_0)\right) + \kappa \log(T_0)/T_0, \tag{12}$$

where $\hat{\sigma}^2(\mathbf{s}_0) = T_0^{-1}\sum_{t=r+1}^{T}\left\{\hat{Y}_t(\mathbf{s}_0) - \hat{\mathbf{Z}}_t(\mathbf{s}_0)'\hat{\boldsymbol{\eta}}(\mathbf{s}_0)\right\}^2$, $\hat{\boldsymbol{\eta}}(\mathbf{s}_0)$ is the penalized parameter estimator in (10) corresponding to the regularization parameters $\gamma(\mathbf{s}_0)$ and $\beta(\mathbf{s}_0)$, and $\kappa$ is the effective number of parameters. For all possible combinations of $\gamma(\mathbf{s}_0)$ and $\beta(\mathbf{s}_0)$, we choose the combination of regularization parameters that gives the minimum value of $\text{BIC}(\gamma(\mathbf{s}_0), \beta(\mathbf{s}_0))$.

After estimating $\boldsymbol{\lambda}(\mathbf{s}_0)$ and $\boldsymbol{\alpha}(\mathbf{s}_0)$, we obtain the final estimator of the function $g_0(x)$ by substituting $\hat{\boldsymbol{\lambda}}(\mathbf{s}_0)$ and $\hat{\boldsymbol{\alpha}}(\mathbf{s}_0)$ into (8),

$$\hat{g}_0(x) = \hat{g}_1\left(x, \mathbf{s}_0\right) - \hat{\mathbf{g}}_{21}\left(x, \mathbf{s}_0\right)' \hat{\boldsymbol{\lambda}}(\mathbf{s}_0) - \hat{\mathbf{g}}_{22}\left(x, \mathbf{s}_0\right)' \hat{\boldsymbol{\alpha}}(\mathbf{s}_0). \tag{13}$$

In the methodology developed here, the order of spatio-temporal lag interactions $p$ and the order of temporal lag interactions $q$ need to be pre-specified in model (1). In the housing price data example (Section 5), we recommend to consider $p = q = 6$, which is slightly larger than what was chosen by AIC with correction (AICc) in Appendix of Al-Sulami *et al.* (2017). This recommendation is owing to the advantage of the data-driven method which enables the selection of the more important lag interactions by penalization. Alternatives to local linear fitting considered above include local constant fitting, Nadaraya–Watson kernel estimate, and spline based methods. However, local linear fitting is known to outperform local constant fitting and Nadaraya–Watson kernel estimate (see, e.g., Fan and Gijbels, 1996).

  *J. Time Ser. Anal.* **40**: 327–342 (2019)
DOI: 10.1111/jtsa.12442

## 4. ASYMPTOTIC PROPERTIES

In semiparametric nonlinear regression time series model (1), there are two sets of parameters $\lambda_{0k,i}$ and $\alpha_{0,l}$, with $\lambda_{0k,i}$ reflecting the interaction between spatial locations $\mathbf{s}_k$ and $\mathbf{s}_0$ at time lag $i$, while $\alpha_{0,l}$ reflecting the temporal lag interaction for a given spatial location $\mathbf{s}_0$ at time lag $l$. Denote by $S_N(\mathbf{s}_0) = \{\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_N\}$ the set of $N$ spatial locations that are potentially interacting with spatial location $\mathbf{s}_0$. Let $A_i(\mathbf{s}_0) = \{\mathbf{s}_k \in S_N(\mathbf{s}_0) : \lambda_{0k,i} \neq 0\}$ denote the set of spatial locations that do interact with spatial location $\mathbf{s}_0$ at time lag $i$, whereas $A_i^c(\mathbf{s}_0) = \{\mathbf{s}_k \in S_N(\mathbf{s}_0) : \lambda_{0k,i} = 0\}$ the set of spatial locations that do not interact with spatial location $\mathbf{s}_0$ at time lag $i$. We let $n_i(\mathbf{s}_0) = \#A_i(\mathbf{s}_0)$ denote the cardinality of $A_i(\mathbf{s}_0)$ and $n_i^*(\mathbf{s}_0) = N - n_i(\mathbf{s}_0)$ denote the cardinality of $A_i^c(\mathbf{s}_0)$.

We partition $\lambda(\mathbf{s}_0)$, the $Np$-dimensional vector of spatio-temporal lag interactions $\lambda_{0k,i}$, into $\lambda(\mathbf{s}_0) = \left(\lambda_1(\mathbf{s}_0)', \lambda_2(\mathbf{s}_0)'\right)'$, where $\lambda_1(\mathbf{s}_0)$ is a vector of $\lambda_{0k,i}$ with $i, k$ such that $\mathbf{s}_k \in A_i(\mathbf{s}_0)$, and $\lambda_2(\mathbf{s}_0)$ is a vector of $\lambda_{0k,i}$ with $i, k$ such that $\mathbf{s}_k \in A_i^c(\mathbf{s}_0)$. Therefore, with $(Np)_0 = \sum_{i=1}^p n_i(\mathbf{s}_0)$, $\lambda_1(\mathbf{s}_0)$ is an $(Np)_0$-dimensional vector of non-zero spatio-temporal lag interactions and $\lambda_2(\mathbf{s}_0)$ is an $(Np - (Np)_0)$-dimensional vector of zero spatio-temporal lag interactions.

Similarly, we partition $\boldsymbol{\alpha}(\mathbf{s}_0)$, the $q$-dimensional vector of temporal lag interactions $\alpha_{0,l}$, into $\boldsymbol{\alpha}(\mathbf{s}_0) = \left(\boldsymbol{\alpha}_1(\mathbf{s}_0)', \boldsymbol{\alpha}_2(\mathbf{s}_0)'\right)'$. Let $B(\mathbf{s}_0) = \{1 \le l \le q : \alpha_{0,l} \neq 0\}$ denote a set of $q_0 = \#B(\mathbf{s}_0)$ non-zero temporal lag interactions and $B^c(\mathbf{s}_0) = \{1 \le l \le q : \alpha_{0,l} = 0\}$, a set of $(q - q_0)$ zero temporal lag interactions. Then, $\boldsymbol{\alpha}_1(\mathbf{s}_0)$ is a $q_0$-dimensional vector of all $\alpha_{0,l}$ such that $l \in B(\mathbf{s}_0)$ and $\boldsymbol{\alpha}_2(\mathbf{s}_0)$ is a $(q - q_0)$-dimensional vector of all $\alpha_{0,l}$ such that $l \in B^c(\mathbf{s}_0)$.

Furthermore, let $\boldsymbol{\eta}^0(\mathbf{s}_0) = \left(\boldsymbol{\eta}_1^0(\mathbf{s}_0)', \boldsymbol{\eta}_2^0(\mathbf{s}_0)'\right)'$ denote an $(Np + q)$-dimensional vector of true interactions, where $\boldsymbol{\eta}_1^0(\mathbf{s}_0) = (\lambda_1(\mathbf{s}_0)', \boldsymbol{\alpha}_1(\mathbf{s}_0)')'$ is an $((Np)_0 + q_0)$-dimensional vector of non-zero interactions and $\boldsymbol{\eta}_2^0(\mathbf{s}_0) = (\lambda_2(\mathbf{s}_0), \boldsymbol{\alpha}_2(\mathbf{s}_0))'$ is an $(Np - (Np)_0 + q - q_0)$-dimensional vector of zero interactions. For the regularization parameters, we define $a_{T_0}^*(\mathbf{s}_0) = \max\left\{\gamma_i^k(\mathbf{s}_0), \beta_l(\mathbf{s}_0) : \mathbf{s}_k \in A_i(\mathbf{s}_0), l \in B(\mathbf{s}_0)\right\}$, and $d_{T_0}^*(\mathbf{s}_0) = \min\left\{\gamma_i^k(\mathbf{s}_0), \beta_l(\mathbf{s}_0) : \mathbf{s}_k \in A_i^c(\mathbf{s}_0), l \in B^c(\mathbf{s}_0)\right\}$, where the maximum and minimum are taken over $i = 1, \dots, p$, $k = 1, \dots, N$, and $l = 1, \dots, q$. We let $|\cdot|$ and $\|\cdot\|$ denote the $L_1$ and $L_2$ norm respectively. We let $\xrightarrow{P}$ and $\xrightarrow{D}$ denote convergence in probability and convergence in distribution respectively.

For establishing the asymptotic properties, we impose regularity conditions given in Appendix S1. In the theorems below, we state the assumptions made about the regularization parameters and the asymptotic results. At a given spatial location $\mathbf{s}_0$, we establish the consistency, sparsity, and asymptotic normality of the penalized parameter estimator $\hat{\boldsymbol{\eta}}(\mathbf{s}_0)$ obtained in (10) as follows.

**Theorem 1.** Suppose that the regularity conditions in Appendix S1 hold and that the regularization parameter satisfies $a_{T_0}^*(\mathbf{s}_0) = o_p(T_0^{-1/2})$ as $T_0 \to \infty$. Then, there exists a global minimizer $\hat{\boldsymbol{\eta}}(\mathbf{s}_0)$ of the objective function $Q(\boldsymbol{\eta}(\mathbf{s}_0))$ such that $\| \hat{\boldsymbol{\eta}}(\mathbf{s}_0) - \boldsymbol{\eta}^0(\mathbf{s}_0) \| = O_p\left(T_0^{-1/2} + a_{T_0}^*(\mathbf{s}_0)\right)$.

By Theorem 1, when the regularization parameters associated with the non-zero interactions converge to zero at a rate of $\sqrt{T_0}$, the penalized parameter estimator $\hat{\boldsymbol{\eta}}(\mathbf{s}_0)$ is a global minimizer and $\sqrt{T_0}$-consistent.

**Theorem 2.** Suppose that the regularity conditions in Appendix S1 hold and that the regularization parameter satisfies $\sqrt{T_0} d_{T_0}^*(\mathbf{s}_0) \to \infty$ as $T_0 \to \infty$ and $\| \hat{\boldsymbol{\eta}}(\mathbf{s}_0) - \boldsymbol{\eta}^0(\mathbf{s}_0) \| = O_p\left(T_0^{-1/2}\right)$. Then, $\hat{\boldsymbol{\eta}}_2(\mathbf{s}_0) = \mathbf{0}$ with probability tending to one. That is, as $T_0 \to \infty$, $P\left(\hat{\lambda}_2(\mathbf{s}_0) = \mathbf{0}\right) \to 1$ and $P\left(\hat{\boldsymbol{\alpha}}_2(\mathbf{s}_0) = \mathbf{0}\right) \to 1$.

Theorem 2 shows the sparsity of the penalized parameter estimator $\hat{\boldsymbol{\eta}}(\mathbf{s}_0)$, which is $\sqrt{T_0}$-consistent. That is, with the regularization parameters in (11), with probability tending to one, the zero interactions are estimated to be 0.

  
DOI: 10.1111/jtsa.12442

**Theorem 3.** Suppose that the regularity conditions in Appendix S1 hold, $\sqrt{T_0}a^*_{T_0}(\mathbf{s}_0) \to 0$ and $\sqrt{T_0}d^*_{T_0}(\mathbf{s}_0) \to \infty$. Then,

$$\sqrt{T_0}\left(\hat{\boldsymbol{\eta}}_1(\mathbf{s}_0) - \boldsymbol{\eta}_1^0(\mathbf{s}_0)\right) \xrightarrow{D} N\left(\mathbf{0}, \boldsymbol{\Psi}_1(\mathbf{s}_0)\right),$$

where $\boldsymbol{\Psi}_1(\mathbf{s}_0) = \sigma_0^2\boldsymbol{\Sigma}_1^{-1}(\mathbf{s}_0)$ and $\boldsymbol{\Sigma}_1(\mathbf{s}_0) = \begin{pmatrix} \boldsymbol{\Sigma}_{\lambda_1}(\mathbf{s}_0) & \boldsymbol{\Sigma}_{\lambda_1\boldsymbol{\alpha}_1}(\mathbf{s}_0) \\ \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_1\lambda_1}(\mathbf{s}_0) & \boldsymbol{\Sigma}_{\boldsymbol{\alpha}_1}(\mathbf{s}_0) \end{pmatrix}$, with $\boldsymbol{\Sigma}_{\lambda_1}(\mathbf{s}_0)$, $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}_1}(\mathbf{s}_0)$ and $\boldsymbol{\Sigma}_{\lambda_1\boldsymbol{\alpha}_1}(\mathbf{s}_0)$ defined in Appendix S1.

Theorem 3 establishes a central limit theorem for the penalized parameter estimator of the non-zero interactions.

**Theorem 4.** Suppose that the regularity conditions in Theorem 3 hold and the bandwidth $b$ satisfies conditions (C7) (ii,iii) for $x$ in the support of $X(\mathbf{s}_0)$ at the spatial location $\mathbf{s}_0$. Then, as $T \to \infty$, we have

$$\sqrt{T_0 b}\left[\left\{\hat{g}_0(x) - g_0(x)\right\} - (1/2)b^2 B_0(x,\mathbf{s}_0)\right] \xrightarrow{D} N\left(\mathbf{0}, \Gamma(x,\mathbf{s}_0)\right),$$

where $B_0(x,\mathbf{s}_0) = \frac{\partial^2 g_0(x)}{\partial x^2}\int u^2K(u)\,\mathrm{d}u$, $\Gamma(x,\mathbf{s}_0) = \frac{\sigma^2(x,\mathbf{s}_0)}{p(x,\mathbf{s}_0)}\int K^2(u)\,\mathrm{d}u$, $p(x,\mathbf{s}_0)$ is the marginal density function of $X_t(\mathbf{s}_0)$, and

$$\sigma^2(x,\mathbf{s}_0) = Var\left[\left(Y_t(\mathbf{s}_0) - \mathbf{Z}_{1,t}(\mathbf{s}_0)'\lambda(\mathbf{s}_0) - \mathbf{Z}_{2,t}(\mathbf{s}_0)'\boldsymbol{\alpha}(\mathbf{s}_0)\right)|X_t(\mathbf{s}_0) = x\right].$$

Theorem 4 establishes the asymptotic properties of $\hat{g}_0(x)$, which is the non-parametric estimator of the unknown possibly nonlinear function $g(x)$.

## 5. DATA EXAMPLE: US HPI

We now return to the study of US HPI, which is known to fluctuate between boom and recession periods. The CPI is an important economic factor that may impact housing prices negatively or positively. The methodology developed in Sections 2–4 can be applied to identify important spatio-temporal lag interactions as well as possibly nonlinear relationship between the HPI and the CPI.

### 5.1. Data and Exploratory Analysis

The HPI data are obtained from the Federal Home Loan Mortgage Corporation (www.freddiemac.com). This time series comprises 453 monthly observations of HPI from January 1975 to September 2012 for each of the 50 states and the District of Columbia (DC). The HPI time series for DC and three other states (Hawaii, Texas, and Washington) are plotted in Figure 1(a) and are clearly non-stationary in time. These three states are chosen for illustration, because their geographical locations are outlying and thus it is not straightforward to decide which states are considered to their neighbors by the standard spatial econometric methods (see, e.g., Anselin, 1988). Furthermore, time series of the geometric return of the HPI are plotted in Figure 1(b), defined as $Y_t(\mathbf{s}_0) = \log P_t(\mathbf{s}_0)/P_{t-1}(\mathbf{s}_0)$ where $P_t(\mathbf{s}_0)$ is the HPI in month $t$ at a given state $\mathbf{s}_0$. Unlike the HPI, the geometric returns of the HPI appear to be stationary in time. The box plots of the geometric returns of the HPI for the 50 states and DC are shown in Figure 2 and the temporal averages are mapped across states in Figure 3.

The CPI, on the other hand, is an average change of the prices of products over a time period and is an indicator of inflation in a country. Here we consider the CPI for all urban consumers (CPI-U), which represents about 80% of the American population, published by the Bureau of Labor Statistics (`www.bls.gov`). In particular, we

  

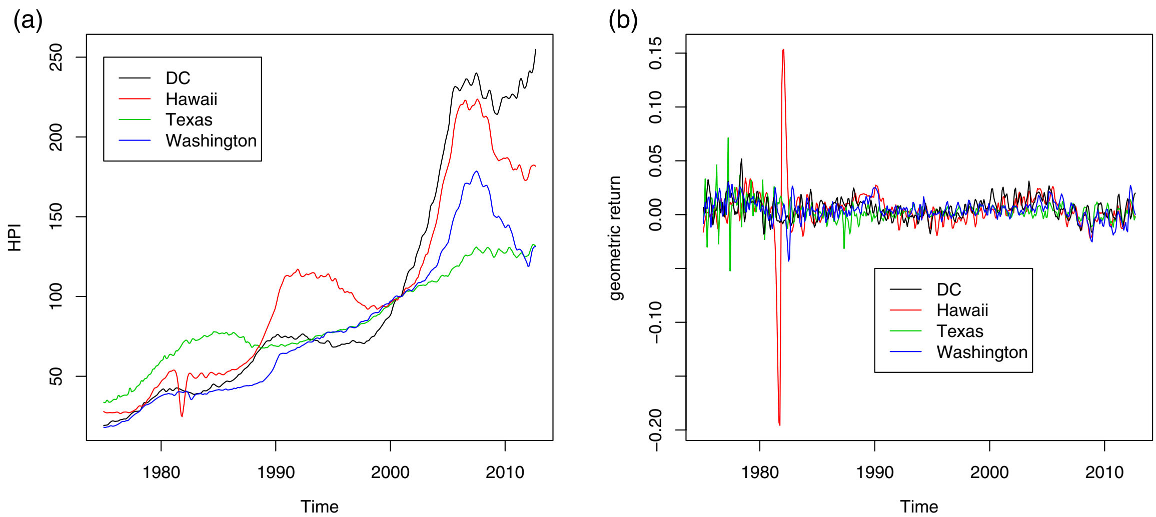


Figure 1. Time series plots of (a) the monthly HPI and (b) its geometric return for DC, Hawaii, Texas and Washington from January 1975 to September 2012 [Color figure can be viewed at wileyonlinelibrary.com]

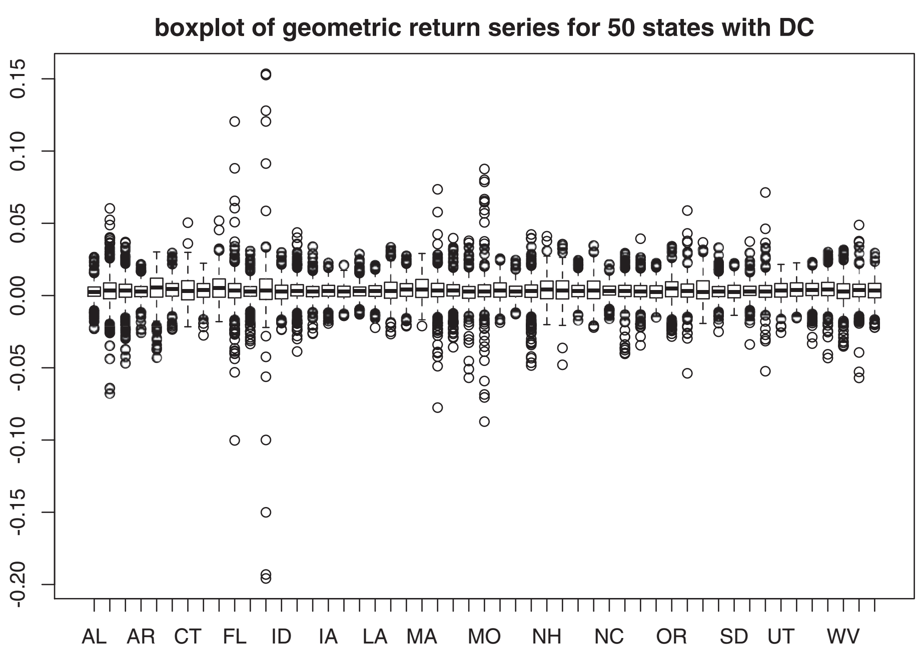


Figure 2. Boxplots of the geometric returns of HPI from January 1975 to September 2012 for each of the 50 US states and 1 district. The states are placed along the *x*-axis in the order of AL: Alabama; AK: Alaska; AZ: Arizona; AR: Arkansas; CA: California; CO: Colorado; CT: Connecticut; DE: Delaware; DC: District of Columbia; FL: Florida; GA: Georgia; HI: Hawaii; ID: Idaho; IL: Illinois; IN: Indiana; IA: Iowa; KS: Kansas; KY: Kentucky; LA: Louisiana; ME: Maine; MD: Maryland; MA: Massachusetts; MI: Michigan; MN: Minnesota; MS: Mississippi; MO: Missouri; MT: Montana; NE: Nebraska; NV: Nevada; NH: New Hampshire; NJ: New Jersey; NM: New Mexico; NY: New York; NC: North Carolina; ND: North Dakota; OH: Ohio; OK: Oklahoma; OR: Oregon; PA: Pennsylvania; RI: Rhode Island; SC: South Carolina; SD: South Dakota; TN: Tennessee; TX: Texas; UT: Utah; VT: Vermont; VA: Virginia; WA: Washington; WV: West Virginia; WI: Wisconsin; WY: Wyoming


DOI: 10.1111/jtsa.12442

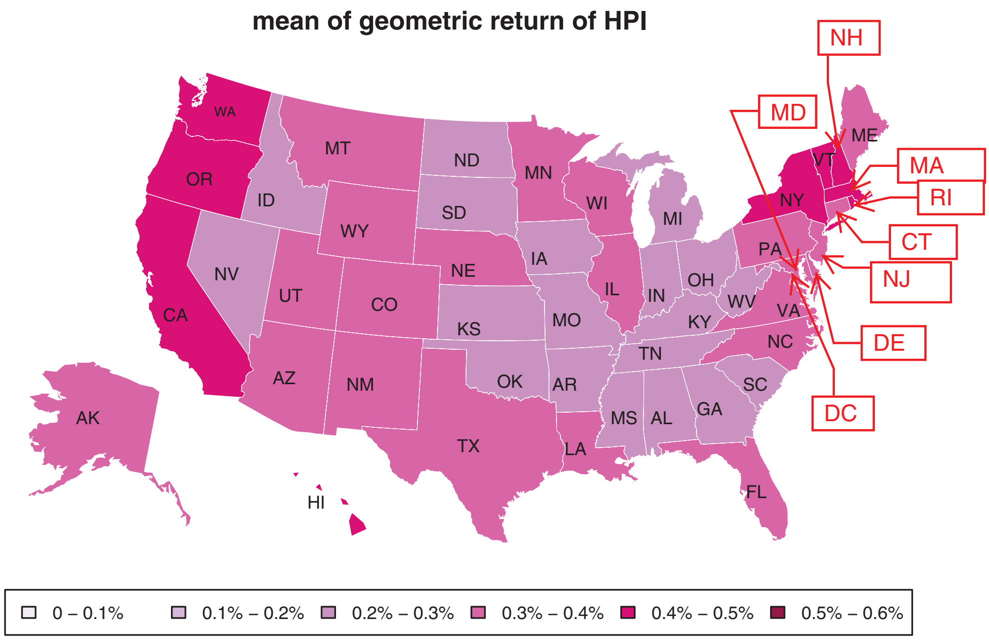


Figure 3. Map of the mean geometric return of the HPI averaged from January 1975 to September 2012 across the 50 states and Washington, DC [Color figure can be viewed at wileyonlinelibrary.com]

focus on the CPI-U of primary residence (CPI-UR) and the time series of monthly CPI-UR in Figure 4(a) shows non-stationarity. Thus we consider a monthly increment of the CPI-UR, defined as $x_t = \mathrm{CPI}_t - \mathrm{CPI}_{t-1}$, where $\mathrm{CPI}_t$ is the CPI-UR in month $t$. Figure 4(b) plots the monthly increment CPI-UR indicating approximate stationarity, while Figure 4(c) plots its kernel density estimate in comparison to a Gaussian density estimate indicating that the distribution may not be Gaussian.

### 5.2. Model Fitting

For illustration, we evaluate possibly nonlinear relationships between the geometric return of the HPI and the increment of the CPI-UR in DC, Hawaii, Texas, and Washington (Figure 1), each with possible interactions with the other $N = 50$ states (or district). We fit the semiparametric nonlinear regression time series model (1) to the data such that the monthly geometric return of the HPI (multiplied by 100) $Y_t(\mathbf{s}_0)$ is the response variable for a given state $\mathbf{s}_0$ and the monthly increment of CPI-UR $X_t(\mathbf{s}_0) = x_{t-1}$ is the covariate that is the same for all the states for a given month. We use the temporal lag 1 of the increment of CPI (i.e., $x_{t-1}$, not $x_t$) here to avoid potential endogeneity of CPI in modeling HPI (c.f., Kuang and Liu, 2015; Panagiotidis and Printzis, 2016) and also for the purpose of forecasting. As mis-specification of the spatio-temporal lag interactions in model (1) may bias the estimate of the relationship between the geometric return of the HPI and the monthly increment of the CPI-UR, we consider model (1) with temporal lag orders $p = q = 6$, which are chosen slightly larger than the orders $p = q = 5$ selected by AICc in Al-Sulami *et al.* (2017). The results are similar for orders greater than $p = q = 6$, which is not unexpected as interactions for larger temporal lags are much smaller (see Figures 5 and A1–A6).

Thus, for a given state $\mathbf{s}_0$ and $N = 50$, the following forecasting model can be easily obtained based on model (1):

$$Y_t(\mathbf{s}_0) = g_0(x_{t-1}) + \sum_{i=1}^{6} \sum_{k=1}^{N} \lambda_{0k,i} Y_{t-i}(\mathbf{s}_k) + \sum_{l=1}^{6} \alpha_{0,l} Y_{t-l}(\mathbf{s}_0) + \varepsilon_t(\mathbf{s}_0), \tag{14}$$




DOI: 10.1111/jtsa.12442

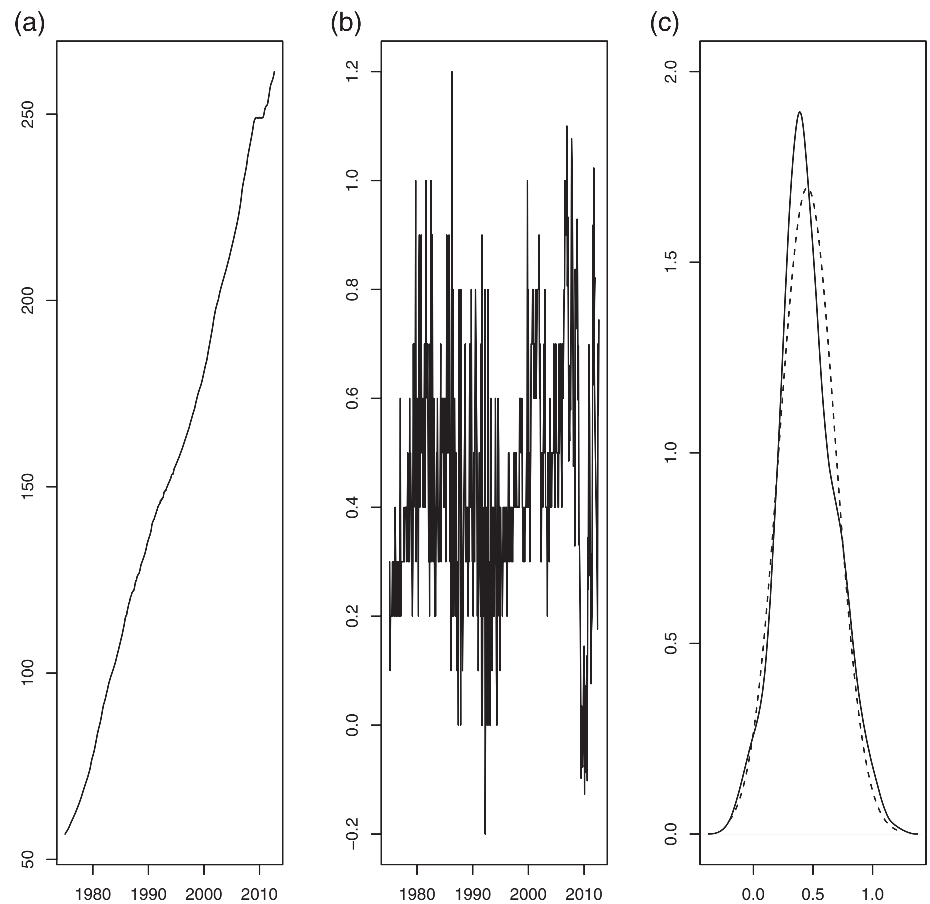


Figure 4. Plots of CPI: (a) original time series; (b) monthly increments; and (c) kernel density estimate (solid curve) compared with Gaussian density estimate with the same mean and variance (dashed curve)

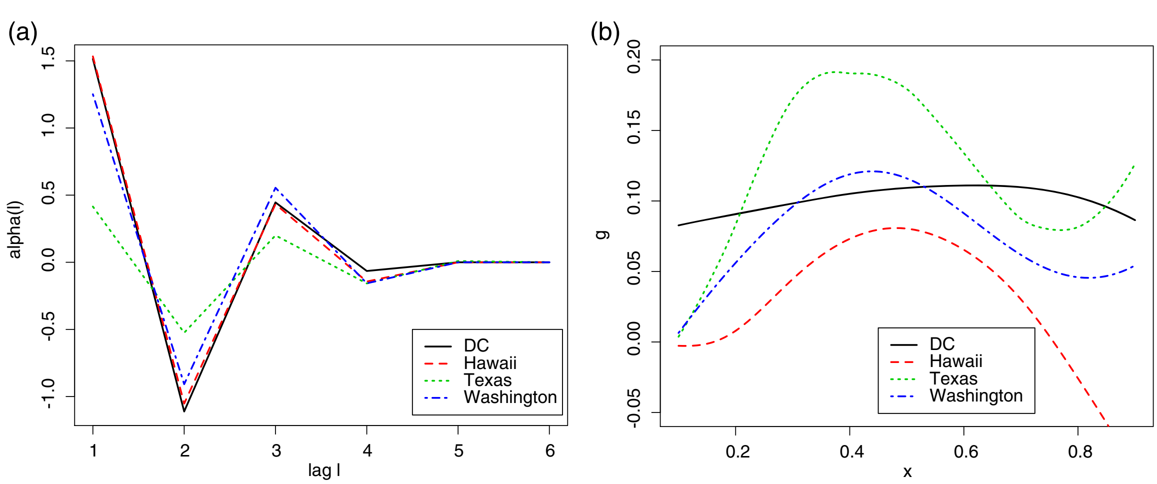


Figure 5. Plots of (a) the estimated temporal lag interactions $\hat{\alpha}_{\ell}$ for $l = 1, \ldots, 6$; (b) the estimated function $\hat{g}_0$ for DC, Hawaii, Texas and Washington [Color figure can be viewed at wileyonlinelibrary.com]

for $t = 1, \ldots, T(= 452)$ and $k = 1, \ldots, N(= 50)$. By (14), we have the fitted response variable

$$\hat{Y}_t(\mathbf{s}_0) = \sum_{i=1}^{6} \sum_{k=1}^{N} \lambda_{0k,i} \hat{Y}_{t-i}(\mathbf{s}_k) + \sum_{l=1}^{6} \alpha_{0,l} \hat{Y}_{t-l}(\mathbf{s}_0), \tag{15}$$

where the conditional means $\hat{Y}_t(\mathbf{s}_0) = Y_t(\mathbf{s}_0) - \hat{E}\left[Y_t(\mathbf{s}_0)|x_{t-1}\right]$, $\hat{Y}_{t-i}(\mathbf{s}_k) = Y_{t-i}(\mathbf{s}_k) - \hat{E}\left[Y_{t-i}(\mathbf{s}_k)|x_{t-1}\right]$, and $\hat{Y}_{t-l}(\mathbf{s}_0) = Y_{t-l}(\mathbf{s}_0) - \hat{E}\left[Y_{t-l}(\mathbf{s}_0)|x_{t-1}\right]$ are estimated by local linear regression with bandwidth selected by cross-validation (CV). Then, we obtain the penalized parameter estimate of $\lambda_{0k,i}$ and $\alpha_{0,l}$ by minimizing the following penalized sum of squared errors with a single regularization parameter,

$$Q\left(\boldsymbol{\eta}(\mathbf{s}_0)\right) = \sum_{t=7}^{T} \left[\hat{Y}_t(\mathbf{s}_0) - \hat{\mathbf{Z}}_t(\mathbf{s}_0)'\boldsymbol{\eta}(\mathbf{s}_0)\right]^2 + T_0 \sum_{i=1}^{6} \sum_{k=0}^{N} \gamma_i^k(\mathbf{s}_0)|\lambda_{0k,i}|,$$

where $\boldsymbol{\eta}(\mathbf{s}_0) = \left(\lambda_{01,1}, \ldots, \lambda_{0N,1}, \ldots, \lambda_{01,6}, \ldots, \lambda_{0N,6}, \lambda_{00,1}, \ldots, \lambda_{00,6}\right)'$ with $\lambda_{00,i}$ denoting $\alpha_{0,i}$. Furthermore, $\gamma_i^k(\mathbf{s}_0) = \gamma(\mathbf{s}_0)\frac{\log(T_0)}{T_0|\tilde{\lambda}_{0k,i}|}$, where $\tilde{\lambda}_{0k,i}$ are the initial estimates of $\lambda_{0k,i}$ by the least squares for model (15) and $\gamma(\mathbf{s}_0)$ is selected by the BIC. For simplicity, here we consider only one tuning parameter $\gamma(\mathbf{s}_0)$ for both spatio-temporal and temporal interactions.

For DC, the estimates of the temporal lag interactions $\alpha_{0,l}$ and the unknown function $g(\cdot)$ are plotted in Figure 5(a) and (b) respectively. Furthermore, Figure 5(a) shows that the estimates of the temporal lag interactions for DC $\alpha_{0,\ell}$ decrease in magnitude over the time lags and appear relatively strong for time lags 1–4 but weak for time lags 5–6. It is also interesting that the temporal lag interactions oscillate around 0, which may indicate a self-adjustment or reversion of the housing price change along time. As commented by a referee, this is also probably due to the weakly correlated errors that are compensating each other at adjacent lags. In Figure 5(b), the estimated function $g(\cdot)$ is nearly flat, implying a weak or no relationship between the geometric return of HPI return and monthly increment of CPI-UR for DC, after taking into account the spatio-temporal lag interactions. We have also considered a residual analysis for model diagnostics and in particular, we plot the autocorrelation functions to evaluate the assumption of independence for the innovations. For example, we find that compared with the strong autocorrelation of the original return series $Y_t$, the time series of residuals of (14), denoted by $e_t$, is largely uncorrelated for DC (Figure 6).

For the other three states of (Hawaii, Texas, and Washington), we have presented the estimates for $\alpha_{0,\ell}$ and $g(\cdot)$ in Figure 5. Figures A1–A6 in Appendix S1 plot the estimated spatio-temporal lag interactions for all 50 states and DC. Figure 5 shows that the patterns of the temporal lag interactions for Hawaii, Texas, and Washington are similar to that for the DC. However, the relationships between HPI and CPI-UR for the three states differ from that for DC, with obvious nonlinearity. For Texas, the relationship is positive when the monthly increment of CPI-UR is, approximately, below 0.3 and above 0.75, relatively flat when the monthly increment of CPI-UR is between about 0.3 and 0.5, and negative when the monthly increment of CPI-UR is between about 0.5 and 0.75. For Washington, the relationship between HPI and CPI-UR is positive when the CPI-UR is, approximately, below 0.4 and above 0.8, and turns negative when the CPI-UR is between 0.4 and 0.8. For Hawaii, the relationship between HPI and CPI-UR starts off negative when the CPI-UR is below 0.2, turns positive when the CPI-UR is between 0.2 and 0.75, and turns negative again when the CPI-UR is above 0.75. These are potentially interesting patterns about housing prices and investments in different parts of the country. In particular, important spatio-temporal lag interactions could occur between states that are not necessarily close in geographical distance.

### 5.3. Prediction Performance

To further evaluate the model fitting in Section 5.2, we consider prediction and comparison with alternative approaches. We partition the data into two parts. The first part has the first $T = 402$ observations and is

  
DOI: 10.1111/jtsa.12442

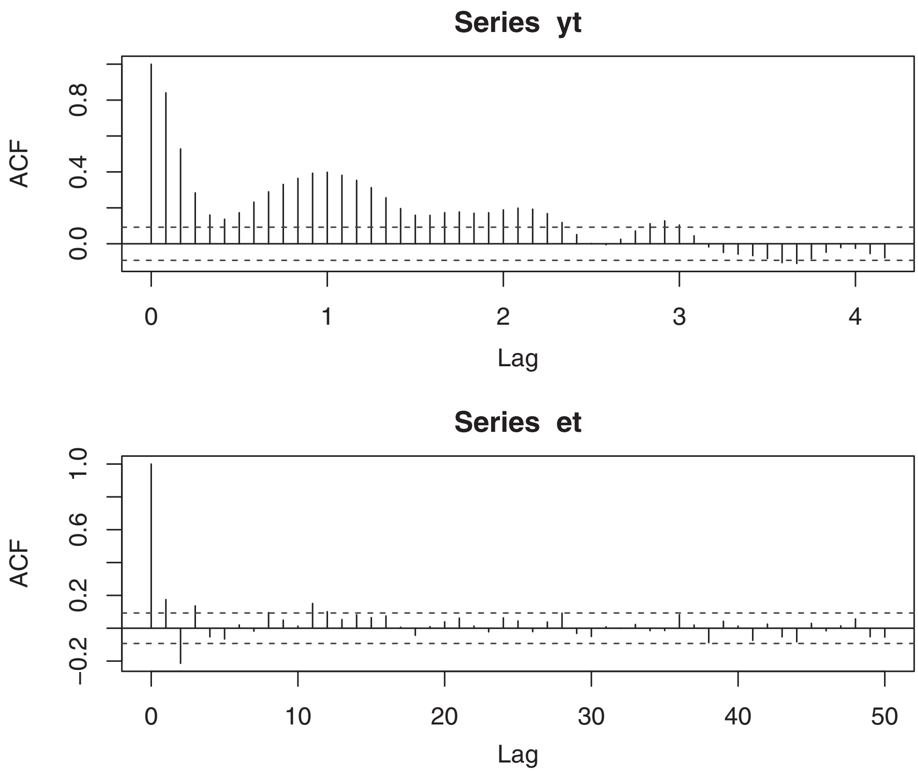


Figure 6. The ACF for the original return of housing price index (HPI) $Y_t$ and the residual $e_t$ for Washington, DC

Table I. MSPE based on the last 50 months of HPI geometric returns assuming linearity $g_L$, nonlinearity $g_{NP}$, and using ordinary least squares $g_{LS}$ without penalization

| | $g_L$ | $g_{NP}$ | $g_{LS}$ |
|---|---|---|---|
| DC | 0.090 | 0.069 | 1.574 |
| Hawaii | 0.154 | 0.138 | 5.540 |
| Texas | 0.288 | 0.282 | 1.237 |
| Washington | 0.261 | 0.227 | 0.972 |

used for model fitting. The second part has the last 50 observations and we consider one-step ahead prediction for testing. The performance of the prediction is compared between two different forms of the function $g$ in terms of mean squared prediction error (MSPE). In the first form, $g$ is assumed to be a linear function, $g_L(x_{t-1}) = a_0(\mathbf{s}_0) + a_1(\mathbf{s}_0)x_{t-1}$, where the linear coefficients $a_0(\mathbf{s}_0)$ and $a_1(\mathbf{s}_0)$ are for a given location $\mathbf{s}_0$. In the second form, we apply our non-parametric specification of $g$ in model (14) and denote it as $g_{NP}$. The MSPE values computed based on the two forms of $g$ and the testing data are provided in the first two columns of Table I. The results suggest that a flexible, possibly nonlinear, form for $g$ that relates the geometric return of the HPI to the monthly increment of the CPI-UR is helpful and improves the accuracy of prediction for the three states and DC.

Furthermore, to evaluate the penalized approach that enables the simultaneous selection and estimation of the spatio-temporal lag interactions, we compare our methodology with the least squares approach without regularization, which we denote as $g_{LS}$ in the third column of Table I. The MSPE values under least squares are much larger, demonstrating that the penalization has helped to improve prediction accuracy.

  
DOI: 10.1111/jtsa.12442

In summary, our data analysis based on the proposed model (1) is fully data-driven in the specification of spatio-temporal interactions between spatial locations, which allows spatial weights to vary over different temporal lags. Like purely non-parametric approach, our method can help to extract useful information and knowledge for improvement of model forecasting, but itself may not always perform the best in forecasting when compared with other more prior information imposed spatial time series models. However, we do find that for the DC, the forecasting based on the proposed model with refined optimal tuning parameter achieves an MSPE as small as 0.06905, which outperforms that of Al-Sulami *et al.* (2017) with pre-specified inverse distance based spatial weights, which has an MSPE of 0.09416. The data-driven based spatial weight matrix is more adaptive to the data and hence likely more robust in forecasting.

## 6. CONCLUSIONS AND DISCUSSION

In this article, we have proposed a semiparametric data-driven nonlinear method that allows potentially nonlinear relationship between a response variable and a covariate, as well as spatial interactions that could vary by temporal lag. In economic applications, spatial weights are usually specified *a priori* between two spatial locations or units that are neighbors of each other according to a neighborhood structure, which can be somewhat subjective. In contrast, our penalized estimation method identifies the important lag interactions across space and over time, providing a data-driven way to determine the spatio-temporal weights more objectively. If it is desirable to consider only interactions between neighboring spatial units guided by theory or for ease of interpretation, our method could be modified by setting the interactions to zero between spatial units that are not neighbors of each other.

We have applied our methodology to analyze a US housing price data set focusing on DC, Hawaii, Texas, and Washington. The data analysis has revealed nonlinear relationships between housing price and CPI for the three states but not DC. The prediction based on our method can be more accurate than the more standard approaches that assume linearity or without penalization for the US housing price data. Our approach here is different from the time simultaneous perspective of linear spatial autoregressive models (see, e.g., Ahrens and Bhattacharjee, 2015; Qu and Lee, 2015). Extending our methodology to spatial autoregressive models with nonlinear covariate structure would be interesting for understanding simultaneous spatial interactions, although it would not be suitable for forecasting.

Furthermore, our proposed methodology works well for a given spatial location in relation to other spatial locations on a lattice, but may not be optimal in identifying the whole network. It would be interesting to extend the methodology to simultaneously examine all the locations on the lattice, although we may face the challenge of estimating a much larger number of parameters. For example, with $N = 51$ and $p = q = 6$, there will be 15606 parameters. Application of dimension reduction techniques of semi-data-driven estimation with prior information may be desirable. We leave these for future research.

Finally, we have established the asymptotic properties of the estimation procedure. Theoretical justifications for the non-parametric and nonlinear estimation, as well as the adaptive LASSO for selecting lagged variables in time series and space-time model are well studied under general near epoch dependence structure (c.f., Appendix S1). Thus, the results obtained here can be applied to a wide range of linear and nonlinear time series processes in statistics and econometrics (c.f., Li *et al.*, 2012; Lu and Linton, 2007).

## ACKNOWLEDGEMENTS

We thank the co-editors and two reviewers for their constructive comments that have improved the content and presentation of this article. Lu's research was partially supported by an EU's Marie Curie Career Integration Grant, which is acknowledged.

## SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

  
DOI: 10.1111/jtsa.12442

REFERENCES

Ahrens A, Bhattacharjee A. 2015. Two-step Lasso estimation of the spatial weights matrix. *Econometrics* **3**(1): 128–155.
Akaike H. 1973. Information theory and an extension of the maximum likelihood principle. In *Second International Symposium on Information Theory*, Vol. 1. Berlin: Springer-Verlag; 267–281.
Al-Sulami D, Jiang Z, Lu Z, Zhu J. 2017. Estimation for semiparametric nonlinear regression of irregularly located spatial time-series data. *Econometrics and Statistics* **2**: 22–35.
Anselin L. 1988. *Spatial Econometrics: Methods and Models*. Berlin: Springer Science & Business Media.
Bhattacharjee A, Holly S. 2011. Structural interactions in spatial panels. *Empirical Economics* **40**(1): 69–94.
Bhattacharjee A, Holly S. 2013. Understanding interactions in social networks and committees. *Spatial Economic Analysis* **8**(1): 23–53.
Bhattacharjee A, Jensen-Butler C. 2013. Estimation of the spatial weights matrix under structural constraints. *Regional Science and Urban Economics* **43**(4): 617–634.
Breiman L. 1996. Heuristics of instability and stabilization in model selection. *The Annals of Statistics* **24**(6): 2350–2383.
de Souza P. C. L. 2012. Estimating networks: Lasso for spatial weights. In *The 34th Meeting of the Brazilian Econometric Society* Rio de Janeiro: Brazilian Society of Econometrics.
Fan J, Gijbels I. 1996. *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.
Fan J, Li R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association* **96**(456): 1348–1360.
Frank I, Friedman J. 1993. A statistical view of some chemometrics regression tools. *Technometrics*: 109–135.
Gao J, Lu Z, Tjøstheim D. 2006. Estimation in semiparametric spatial regression. *The Annals of Statistics* **34**(3): 1395–1435.
Griffith DA, Lagona F. 1998. On the quality of likelihood-based estimators in spatial autoregressive models when the data dependence structure is misspecified. *Journal of Statistical Planning and Inference* **69**(1): 153–174.
Hallin M, Lu Z, Tran L. 2004. Local linear spatial regression. *The Annals of Statistics* **32**(6): 2469–2500.
Haufe S, Nolte G, Mueller K, Kraemer N. 2009. Sparse causal discovery in multivariate time series. Arxiv preprint arXiv 0901.2234.
Hefley T, Hooten M, Hanks E, Russell R, Walsh D. 2017. The Bayesian group lasso for confounded spatial data. *Journal of Agricultural, Biological and Environmental Statistics*: 42–59.
Hsu N, Hung H, Chang Y. 2008. Subset selection for vector autoregressive processes using Lasso. *Computational Statistics & Data Analysis* **52**(7): 3645–3657.
Huang H, Hsu N, Theobald D, Breidt F. 2010. Spatial LASSO with applications to GIS model selection. *Journal of Computational and Graphical Statistics* **19**(4): 963–983.
Kuang W, Liu P. 2015. Inflation and house prices: theory and evidence from 35 major cities in China. *International Real Estate Review* **18**(2): 217–240.
Lam C, Souza PC. 2013. Regularization for spatial panel time series using the adaptive lasso. Technical report, Mimeo.
Li D, Lu Z, Linton O. 2012. Local linear fitting under near epoch dependence: uniform consistency with convergence rates. *Econometric Theory* **28**(5): 935–958.
Lu Z, Linton O. 2007. Local linear fitting under near epoch dependence. *Econometric Theory* **23**(1): 37.
Lu Z, Lundervold A, Tjøstheim D, Yao Q. 2007. Exploring spatial nonlinearity using additive approximation. *Bernoulli* **13**(2): 447–472.
Lu Z, Steinskog D, Tjøstheim D, Yao Q. 2009. Adaptively varying-coefficient spatiotemporal models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **71**(4): 859–880.
Lu Z, Tjøstheim D, Yao Q. 2008. Spatial smoothing, nugget effect and infill asymptotics. *Statistics & Probability Letters* **78**(18): 3145–3151.
Manresa E. 2013. Estimating the structure of social interactions using panel data (Unpublished manuscript). CEMFI, Madrid.
McQuarrie A, Tsai C. 1998. *Regression and Time Series Model Selection*. Singapore: World Scientific.
Panagiotidis T, Printzis P. 2016. On the macroeconomic determinants of the housing market in Greece: A VECM approach. *International Economics and Economic Policy* **13**(3): 387–409.
Qu X, Lee L-f. 2015. Estimating a spatial autoregressive model with an endogenous spatial weight matrix. *Journal of Econometrics* **184**(2): 209–232.
Ramanathan R. 1992. *Introductory Econometrics with Applications*. San Diego, CA: Harcourt Brace Colleges.
Ren Y, Zhang X. 2010. Subset selection for vector autoregressive processes via adaptive Lasso. *Statistics & Probability Letters* **80**(23-24): 1705–1712.
Reyes P, Zhu J, Aukema B. 2012. Selection of spatial-temporal lattice models: assessing the impact of climate conditions on a mountain pine beetle outbreak. *Journal of Agricultural, Biological, and Environmental Statistics* **17**: 508–525.
Schwarz G. 1978. Estimating the dimension of a model. *The Annals of Statistics* **6**(2): 461–464.
Shumway R, Stoffer D. 2000. *Time Series Analysis and its Applications*. Berlin: Springer-Verlag.
Smith TE. 2009. Estimation bias in spatial models with strongly connected weight matrices. *Geographical Analysis* **41**(3): 307–332.




DOI: 10.1111/jtsa.12442

Song JJ, De Oliveira V. 2012. Bayesian model selection in spatial lattice models. *Statistical Methodology* **9**(1–2): 228–238.
Stakhovych S, Bijmolt TH. 2009. Specification of spatial models: a simulation study on weights matrices. *Papers in Regional Science* **88**(2): 389–408.
Stetzer F. 1982. Specifying weights in spatial forecasting models: the results of some experiments. *Environment and Planning A* **14**(5): 571–584.
Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*: 267–288.
Tibshirani R. 1997. The Lasso method for variable selection in the Cox model. *Statistics in Medicine* **16**(4): 385–395.
Van De Geer S. 2008. High-dimensional generalized linear models and the lasso. *The Annals of Statistics* **36**(2): 614–645.
Wang H, Li G, Tsai C. 2007. Regression coefficient and autoregressive order shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **69**(1): 63–78.
Zhang W, Yao Q, Tong H, Stenseth N. 2003. Smoothing for spatiotemporal models and its application to modeling Muskrat–Mink interaction. *Biometrics* **59**(4): 813–821.
Zhao P, Yu B. 2007. On model selection consistency of Lasso. *Journal of Machine Learning Research* **7**(2): 2541.
Zhu J, Huang H, Reyes P. 2010. On selection of spatial linear models for lattice data. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **72**(3): 389–402.
Zou H. 2006. The adaptive lasso and its oracle properties. *Journal of the American Statistical Association* **101**(476): 1418–1429.
Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **67**(2): 301–320.

  
DOI: 10.1111/jtsa.12442