Application of LASSO to the eigenvector selection problem in eigenvector based spatial filtering

Hajime Seya

National Institute for Environmental Studies Center for Global Environmental Research,

16-2, Onogawa, Tsukuba-shi, Ibaraki-ken 305-8506, Japan,

email: seya.hajime@nies.go.jp

Daisuke Murakami

Graduate School of Systems and Information Engineering, University of Tsukuba,

1-1-1, Tennodai, Tsukuba-shi, Ibaraki-ken 305-8573, Japan,

email: muraka51@sk.tsukuba.ac.jp

Morito Tsutsumi

Faculty of Engineering, Information and systems, University of Tsukuba,

1-1-1, Tennodai, Tsukuba-shi, Ibaraki-ken 305-8573, Japan,

email: tsutsumi@sk.tsukuba.ac.jp

Yoshiki Yamagata

National Institute for Environmental Studies Center for Global Environmental Research,

16-2, Onogawa, Tsukuba-shi, Ibaraki-ken 305-8506, Japan,

email: yamagata@nies.go.jp

Abstract

Eigenvector based spatial filtering is one of the well-used approaches to model spatial autocorrelation among the observations or errors in a regression model. In this approach, subset of eigenvectors extracted from a modified spatial weight matrix is added to the model as explanatory variables. The subset is typically specified via the forward stepwise model selection procedure, but it is disappointingly slow when the number of observations n takes a large number. Hence as a complement or alternative, the present paper proposes the use of the least absolute shrinkage and selection operator (LASSO) to select the eigenvectors. The LASSO model selection procedure is applied to the well-known Boston housing dataset and simulation dataset, and its performance is compared with the stepwise procedure. The obtained results suggest that the LASSO procedure is fairly fast compared to the stepwise procedure, and can select eigenvectors effectively even if dataset is relatively large ($n = 10^4$), to which the forward stepwise procedure is not easy to apply.

Keywords: eigenvector based spatial filtering, LASSO, forward stepwise, model selection

Introduction

Spatial autocorrelation is one of the important aspects of spatial data, and analyzing this phenomenon has been remarked by researchers in many fields, including, geography (Haining 1990, 2003), real estate economics (Pace et al. 1998; Dubin et al. 1999), urban analysis (Páez and Scott 2005), ecology (Dormann et al. 2007), population genetics (Diniz-Filho and Bini 2012), among others. From the methodological point of view, both spatial statistics (e.g., Cressie 1993; Banerjee et al. 2004; Rue and Held 2005; Schabenberger and Gotway 2005; Gelfand et al. 2010; Chun and Griffith 2013) and spatial econometrics (Anselin 1988, 2010; LeSage and Pace 2009; Arbia 2011) offer many useful toolboxes. Although Anselin's (1986) indication that "each approach tends to be rather self-contained, with little cross-reference shown in published articles" may be still true, in recent years some intensive works have been done to clarify the similarities and differences between spatial statistics and spatial econometrics (Griffith and Paelinck 2007, 2010).

Besides these approaches, spatial filtering approaches have been developed by quantitative geographers as the "third way" to model spatial autocorrelation (e.g., Getis 1990; Getis and Griffith 2002; Griffith 2000, 2003, 2012; Tiefelsdorf and Griffith 2007). These approaches are in particular used in the field of ecology (Griffith and Peres-Neto 2006), whereas also have many social science applications[1]. Among these approaches, eigenvector based spatial filtering (ESF) developed by Professor Daniel Griffith and co-workers are fairly well-used to model spatial autocorrelation among observations or errors in a regression model. In this approach, subset of eigenvectors extracted from a modified spatial weight matrix is added to the model as explanatory variables. The subset is typically specified by the forward stepwise model selection procedure, but it is disappointingly slow when the number of observations n takes a large number.

Hence as a complement or alternative, the present paper proposes the use of the least absolute shrinkage and selection operator (LASSO) (L1-penalized regression) proposed by Tibshirani (1996) to select the eigenvectors. Different from L2-penalized model such as ridge regression[2] where parameters are shrunken toward zero, LASSO automatically sets many of them "exactly" zeros; hence it can be used for variable selection. Although original LASSO algorithm is computationally very demanding to be used in high-dimensional data (Segal 2006), now several remarkably fast algorithms are proposed such as Path algorithm (Park and Hastie 2007), which is the direct generalization of least-angle regression (LARS) algorithm, (Efron et al. 2004) and gradient ascent algorithm with Newton-Raphson step (Goeman 2010)[3]. The comparison of computation time

[1] Recent social science applications include Fischer et al. (2009), Chun and Griffith (2011), Patuelli et al. (2011), Seya and Tsutsumi (2011), Yamagata et al. (2012), and Cuaresma and Feldkircher (2013).

[2] Therefore, it is typically used as the countermeasure against the multicollinearity problem (Tsutsumi et al. 1997).

[3] Bayesian estimation methods are also proposed (Park and Cassella 2008; Chen et al. 2011), but they requires the MCMC simulation, which is computationally intensive.

of these two algorithms are conducted elsewhere (Keerthi and Shevade 2007; Goeman 2010), and it is clarified that it is unlikely that any algorithm is uniformly faster than others, and the computation speed of those two algorithms are comparable.

In this study, the LASSO model selection procedure is applied to the well-known Boston housing dataset and simulation dataset, and its performance is compared to the conventional forward stepwise procedure proposed by Tiefelsdorf and Griffith (2007).

Introduction of the Eigenvector based spatial filtering approach

This section first introduces a basic multiple regression model and conventional spatial econometric models, followed by the ESF approach.

Basic multiple regression model

Let the following standard multiple linear regression model represent the basic model (BM):

$$\boldsymbol{y} = \beta_0 \boldsymbol{1} + \boldsymbol{x}\boldsymbol{\beta}_1 + \boldsymbol{\varepsilon} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad (1)$$

where $\boldsymbol{y}$ is an $n \times 1$ vector of the observations, $\boldsymbol{1}$ is an $n \times 1$ vector of ones, $\boldsymbol{X} = [\boldsymbol{1}; \boldsymbol{x}]$ is an $n \times k$ matrix of explanatory variables (including intercept at the first column), $\boldsymbol{\beta} = [\beta_0; \boldsymbol{\beta}_1']'$ is a $k \times 1$ regression coefficient vector (where $\boldsymbol{A}'$ denotes transpose of a matrix or a vector $\boldsymbol{A}$), $\boldsymbol{\varepsilon}$ is an $n \times 1$ vector of zero-mean independently and identically distributed (i.i.d.) errors with variance σ_ε^2. All of these parameters can be estimated by the ordinary least squares (OLS) method.

Spatial econometric models

In order to control the spatial autocorrelation among observations or errors, the spatial lag model (SLM) and spatial error model (SEM) are typically used in the field of spatial econometrics[4]. The former is occasionally called the spatial autoregressive (response) model (e.g., LeSage and Pace, 2009) and the latter is known as the simultaneous autoregressive model in the spatial statistics literature (e.g., Cressie 1993).

The SLM is expressed as

$$\boldsymbol{y} = \rho \boldsymbol{W}\boldsymbol{y} + \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}), \quad (2)$$

where ρ is a spatial parameter and $\boldsymbol{W}$ is an $n \times n$ spatial weight matrix. Several methods have been proposed to obtain the elements of $\boldsymbol{W}$ (see Anselin 1988; Getis 2009; Stakhovych and Bijmolt 2009; Bhattacharjee and Jensen-Butler 2013; Seya et al. 2013). $\boldsymbol{W}$ is usually normalized to avoid

[4] From the view point of identification, some econometricians recommend the use of spatial cross regressive model, in which spatial lag of the explanatory variables $\boldsymbol{WX}$ is incorporated in the model (Gibbons and Overman 2012).

singularity of the term ($\boldsymbol{I}$–$\rho\boldsymbol{W}$). Although there are various methods for normalization (e.g., Ord 1975; Kelejian and Prucha 2010), the one most widely used is row-normalization, in which the rows sum to unity. $\boldsymbol{\varepsilon}$ is typically assumed to be normally distributed to estimate the parameters of the model by the maximum likelihood method.

The SEM is expressed as

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}, \quad \boldsymbol{u} = \lambda \boldsymbol{W u} + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}), \tag{3}$$

where λ is a spatial parameter. The consistent estimators of the parameters of SLM and SEM are not obtained by the OLS method (Anselin 1988); therefore the parameters are usually estimated by the maximum likelihood method (Ord 1975; Anselin 1988; Lee 2004), the generalized method of moments (Kelejian and Prucha 1998; Lee 2007), or the Bayesian MCMC method (LeSage 1997; Kakamu and Wago 2008), among others. For more details, see LeSage and Pace (2009).

Eigenvector based spatial filtering (ESF) approach

As its own name suggests, the ESF approach tries to filter out the global and local spatial autocorrelations from the observations or errors, and subsequently conducts parameter estimation with the standard OLS method. Thus, the ESF does not require the special parameter estimation method, which is one of the remarkable merits of this approach.

The ESF utilizes eigenvectors $\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}$ that are extracted from a modified spatial weight matrix. Assume that there are two different types of projection matrices, $\boldsymbol{N}^{[1]} = \boldsymbol{I} - \boldsymbol{1}(\boldsymbol{1}'\boldsymbol{1})^{-1}\boldsymbol{1}'$ and $\boldsymbol{N}^{[X]} = \boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$. Tiefelsdorf and Griffith (2007) indicate that the set of eigenvectors $\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}_{SEM}$ that is extracted from

$$\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}_{SEM} = evec[\boldsymbol{N}^{[X]}\frac{1}{2}(\boldsymbol{W}+\boldsymbol{W}')\boldsymbol{N}^{[X]}], \tag{4}$$

is orthogonal to the explanatory variables $\boldsymbol{X}$, where, *evec*[·] denotes the eigenvectors of [·]. In contrast, the set of eigenvectors $\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}_{SLM}$ that is extracted from

$$\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}_{SLM} = evec[\boldsymbol{N}^{[1]}\frac{1}{2}(\boldsymbol{W}+\boldsymbol{W}')\boldsymbol{N}^{[1]}], \tag{5}$$

is potentially correlated with the explanatory variables $\boldsymbol{X}$**.** Let $\boldsymbol{E}_{SEM}$ and $\boldsymbol{E}_{SLM}$ be matrices whose vectors are subsets of $\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}_{SEM}$ and $\{\boldsymbol{e}_1,...,\boldsymbol{e}_n\}_{SLM}$, respectively.

Then it is important to note that the OLS estimate of $\boldsymbol{\beta}$ in $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{\varepsilon}$ coincident with the OLS estimate of $\boldsymbol{\beta}$ in $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{E}_{SEM}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, because $\boldsymbol{X}$ and $\boldsymbol{E}_{SEM}$ are mutually orthogonal[5].

[5] Recently similar ideas are discussed by statisticians in "spatial confounding" literatures. For more details, see Hodges and Reich (2010), and Hughes and Haran (2013).

However, such property does not hold for $\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{E}_{SLM}\boldsymbol{\gamma} + \boldsymbol{\varepsilon}$, in which $\boldsymbol{E}_{SLM}$ correlate with $\boldsymbol{X}$. Hence the use of $\boldsymbol{E}_{SEM}$ instead of $\boldsymbol{E}_{SLM}$ implies that there is no omitted variable, which correlate with the $\boldsymbol{X}$. This assumption corresponds to the SEM specification (LeSage and Pace 2009). Besides, Brasington and Hite (2005) demonstrate that using the SLM may contribute to the mitigation of omitted variable (or spatial-confounding) bias in the estimate of $\boldsymbol{\beta}$, because omitted variables are typically spatially autocorrelated, and thus it correlate with spatial lag variable $\boldsymbol{Wy}$. However, as discussed in Paciorek (2010) intensively, whether the incorporation of spatial effects will improve the spatial-confounding bias/precision of $\boldsymbol{\beta}$ or not depends on the scale of the spatial autocorrelation in the omitted variables. In this regard, the introduction of $\boldsymbol{E}_{SLM}$ may be useful because it captures various distinct scales of spatial variations. In fact, the eigenvector which correspond to large eigenvalue represents large positive spatial autocorrelation, whereas the one with small positive eigenvalue in absolute value correspond to small scale spatial autocorrelation (Griffith, 2003; 2012). Hence the incorporation of $\boldsymbol{E}_{SLM}\boldsymbol{\gamma}$ may correct the omitted variable bias of OLS estimates of $\boldsymbol{\beta}$.

Now we are required to analyze the manner in which the subset of the eigenvectors must be selected. Tiefelsdorf and Griffith (2007) propose two practical forward step-wise procedures for identifying the subset of eigenvectors; maximizing the explained variation or minimizing the residual spatial autocorrelation in terms of standardized Moran's *I*, which termed Moran's *Z*. In these procedures, the next eigenvector $\boldsymbol{e}_l$ is identified from the search set $\boldsymbol{E}^c$ (initially, this set contains all the eigenvectors) that would continue to increase the proportion of explained variation or reduce the standardized Moran's *I*, until some criterion is met.

Application of the LASSO to the eigenvector selection problem

This section first briefly explains the LASSO, subsequently applies it to the eigenvector selection problem.

LASSO

The OLS estimators of $\boldsymbol{\beta}$ is given by minimizing residual some of squares as:

$$\hat{\boldsymbol{\beta}} = \arg\min\{\| \boldsymbol{y} - \boldsymbol{X\beta} \|\}, \tag{6}$$

where $\|\cdot\|$ is the Euclidean norm, and *argmin*(·) denotes the argument of the minimum, that is to say, the set of points of the given argument for which the given function attains its minimum value. Tibshirani (1996) proposes the algorithm termed the LASSO, which minimizing residual some of squares subject to a constraint of the sum of absolute values of the regression coefficient (usually except intercept). Hence the LASSO estimators of $\boldsymbol{\beta}$ is given by:

$$\hat{\boldsymbol{\beta}} = \arg\min\{\| \boldsymbol{y} - \boldsymbol{X\beta} \| + \theta \| \boldsymbol{\beta}_1 \|^1\}, \tag{7}$$

where $\| \boldsymbol{\beta}_1 \|^1$ is the L1 norm given by:

$$\sum_{q=1}^{k-1} | \beta_q |, \tag{8}$$

where $| \cdot |$ denotes the absolute value of (·). Because the value of each β_q depends on the scale, explanatory variables are usually standardized to have mean 0 and unit length, and the dependent variable is also demeaned to have mean 0. If regularization term is given by L2 norm, then the estimator is termed the Ridge estimator[6]. Although an L2 penalty shrinks parameters toward zero, L1 penalty shrinks many of them "exactly" zeros (Tibshirani 1996). Hence it can be used for variable selection. As discussed below, the proportion of regression coefficients set to zero depends on the LASSO regularization parameter θ.

The LASSO estimator can also be regarded as the penalized likelihood estimator. Let ℓ is the likelihood of the model. Then the LASSO estimator of the regression coefficients vector is defined in terms of the penalized likelihood optimization

$$\hat{\boldsymbol{\beta}} = \arg\max \{ \ell_{pen}(\boldsymbol{\beta}) = \ell (\boldsymbol{\beta}) - \theta \| \boldsymbol{\beta}_1 \|^1 \} . \tag{9}$$

Goeman (2010) proposes efficient Gradient ascent algorithm with Newton-Raphson step to compute the LASSO estimates in eq. (9) with given θ. This study utilizes this algorithm.

Application of the LASSO to the eigenvector selection problem

Let the ESF given by:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \widetilde{\boldsymbol{E}}_p \tilde{\boldsymbol{\gamma}} + \boldsymbol{\varepsilon} , \tag{10}$$

where $\widetilde{\boldsymbol{E}}_p = \{ \boldsymbol{e}_1, ..., \boldsymbol{e}_n \}_p$ (p = SEM or SLM) are $n \times n$ matrix[7] with corresponding $n \times 1$ coefficients vector, $\tilde{\boldsymbol{\gamma}}$. The marked merit of the LASSO is that it is feasible even if the number of coefficients exceed the number of observations. The LASSO estimators of $\boldsymbol{\beta}$ and $\tilde{\boldsymbol{\gamma}}$ is given by:

$$[\hat{\boldsymbol{\beta}}'; \hat{\tilde{\boldsymbol{\gamma}}}']' = \arg\min \{ \| \boldsymbol{y} - \boldsymbol{X\beta} - \widetilde{\boldsymbol{E}}_p \tilde{\boldsymbol{\gamma}} \| + \theta \| \tilde{\boldsymbol{\gamma}} \|^1 \} . \tag{11}$$

In terms of penalized likelihood optimization, the LASSO estimators of coefficients are given by

$$[\hat{\boldsymbol{\beta}}'; \hat{\tilde{\boldsymbol{\gamma}}}']' = \arg\max \{ \ell_{pen}(\boldsymbol{\beta} ; \tilde{\boldsymbol{\gamma}}) = \ell (\boldsymbol{\beta} ; \tilde{\boldsymbol{\gamma}}) - \theta \| \tilde{\boldsymbol{\gamma}} \|^1 \} . \tag{12}$$

---

[6] If we introduce both of the L1 norm and L2 norm, then the model is termed the elastic net (Zou and Hastie 2005). Another famous method is Dantzig selector (Candes and Tao 2007), which is a slightly modified version of the LASSO.

[7] Strictly speaking, some vectors which associated with the eigenvalues around zero (< 1e-10 in our case) are excluded in advance.

Or equally,

$$[\hat{\boldsymbol{\beta}}';\hat{\tilde{\boldsymbol{\gamma}}}']' = \arg\max \{ \ell_{pen}(\boldsymbol{\beta};\tilde{\boldsymbol{\gamma}}) = \ell(\boldsymbol{\beta};\tilde{\boldsymbol{\gamma}}) - \theta \sum_{q=1}^{n} |\tilde{\gamma}_q| \}. \quad (12)'$$

The optimal value of regularization parameter θ, say $\hat{\theta}$ is selected using prediction accuracy as the criterion. More specifically, we use the *K* fold cross-validation combined with Brent algorithm (Brent, 1973) to select $\hat{\theta}$. The algorithm proceeds as follows:

[1] Set the minimum and maximum values for $\hat{\theta}$. The minimum value $\theta_{\min}$ usually takes zero, and the maximum value $\theta_{\max}$ usually takes the smallest value that shrinks all regression coefficients $\tilde{\gamma}$ to zero.

[2] Set the initial value of $\hat{\theta}$ as $\hat{\theta} = \theta_0$, where θ_0 satisfies $\theta_{\min} \le \theta_0 \le \theta_{\max}$.

[3] Randomly divide the observations into *K* roughly equal groups (in this study, we use 20 fold validation), then for each *k* = 1, …, *K* with given $\hat{\theta}$, obtain the parameter estimates for regression coefficients, $\hat{\boldsymbol{\beta}}_{-k}(\hat{\theta})$ and $\hat{\tilde{\boldsymbol{\gamma}}}_{-k}(\hat{\theta})$ by maximizing the penalized likelihood function, where $\hat{\boldsymbol{a}}_{-k}$ is a estimates for $\boldsymbol{a}$ obtained using the observations not belonging to *k*th group.

[4] Evaluate cross-validation score as: $\sum_{k=1}^{K} \sum_{i \in group\ k} \{ y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}_{-k}(\hat{\theta}) - \boldsymbol{e}_i' \hat{\tilde{\boldsymbol{\gamma}}}_{-k}(\hat{\theta}) \}^2$, where $\boldsymbol{x}_i$ and $\boldsymbol{e}_i$ are vectors of explanatory variables and eigenvectors, respectively associated with the observation *i*.

[5] Update $\hat{\theta}$ by using the Brent algorithm (Goeman et al., 2012), and set $\hat{\theta} = \theta_{new}$, where θ_{new} satisfies $\theta_{\min} \le \theta_{new} \le \theta_{\max}$.

[6] Repeat [3] ~ [5] until convergence.

Based on the optimum value of $\hat{\theta}_{opt}$, we have the coefficients estimates on the explanatory variables and eigenvectors as $\hat{\boldsymbol{\beta}}(\hat{\theta}_{opt})$, and $\hat{\tilde{\boldsymbol{\gamma}}}(\hat{\theta}_{opt})$, respectively. Because it is problematic for the frequentist LASSO to produce valid standard errors (Kyung et al. 2010), we construct the $\boldsymbol{E}_{SEM}$ or $\boldsymbol{E}_{SLM}$ from eigenvectors associated with non-zero values of $\hat{\tilde{\boldsymbol{\gamma}}}(\hat{\theta}_{opt})$, and it is introduced to the BM as in usual ESFs. Thus, the LASSO in this study is used just for the eigenvectors selection. With regard to the direct use of the LASSO estimates as the final results, Bayesian framework may be one of the useful options because it is fairly straightforward to derive standard errors of coefficients estimates (Kyung et al. 2010).

Generally, the applicability of one method critically depends on the availability of the

useful software. Geographical analyses itself also have evolved with the development of excellent software (Anselin 2012). Fortunately, there already exist some useful packages to calculate the parameters in eqs. (11) or (12). For example, if we use the penalized package of R, optL1() function can be used to select optimum value of θ via above explained cross validation combined with Brent algorithm, and the maximization of penalized likelihood can be performed using penalized() function. The remarkable merit of penalized() function is that it has the option to allow some unpenalized parameters using the algorithm proposed by Goeman (2010)[8]. Hence we can regulate the coefficients of the eigenvectors only as discussed above.

In the geographical analysis related literatures, few studies have employed the LASSO. Wheeler (2009) applies the LASSO to geographically weighed regression (GWR) model to consider the problem of local multicollinearity. Huang et al. (2010) extends the LASSO to the geostatistical model, and employs the proposed model to the GIS model selection. Demšar et al. (2013) is a review article of principal component analysis (PCA) of spatial data. In this paper, they point out that sparse PCA has not often been used but seems proposing for problems where there are a large number of variables and there is a need to determine the key factors.

Empirical illustration

This section shows two illustration of the LASSO algorithm for the eigenvector selection problem using the Boston housing dataset and simulation dataset.

Example with Boston housing dataset

Here, we use the well-known Boston housing dataset. This dataset was originally provided by Harrison and Rubinfield (1978), and Gilley and Pace (1996) augment the dataset with longitude-latitude of the observations. Pace and Gilley (1997) suggest that these data exhibit various problems common to many hedonic pricing or mass appraisal models. For example, not all variables have the proper sign, that is, the AGE variable (see Table 1) is insignificant and positive, and a high positive spatial autocorrelation exists among the observations. Pace and Gilley (1997) construct two hedonic pricing models based on BM and SEM with this dataset and found that SEM would successfully yield the significantly negative estimate of the AGE variable. This curious result may be caused by the impacts of spatially autocorrelated omitted variables on the included variables are successfully incorporated into the model by using the SEM. Kostov (2010) indicate that this dataset is one of the most popular datasets, and it has stimulated a whole industry that has used this and other datasets to examine and compare alternative statistical methods. We adopt the same variables as Kostov (2010), as indicated in Table 1. Following Kostov (2010), the natural logarithms of MEDV (see Table 1), DIS, RAD, and LSTAT are taken, while the squares of NOX and RM are taken to capture

[8] glmnet package in R also have similar option.

some of the underlying nonlinearities, resulting in ln(MEDV), ln(DIS), ln(RAD), ln(LSTAT), NOX^2, and RM^2. For the descriptive statistics of the data, see Kostov (2010). The numbers of observations and the explanatory variables (including the intercept) are 506 and 14, respectively.

Table 1. Variable description (Kostov, 2010)

Variable	Description
MEDV	Median values of owner-occupier housing in thousands of US dollars
LON	Tract point longitude in decimal degrees
LAT	Tract point latitude in decimal degrees
CRIM	Per capita crime
ZN	Proportion of residential land zoned for lots over 25 000 ft^2 per town
INDUS	Proportion of nonretail business acres per town
CHAS	An indicator: 1 if tract borders Charles River; 0 otherwise
NOX	Nitric oxides concentration (parts per 10 million) per town
RM	Average number of rooms per dwelling
AGE	Proportion of owner-occupied units built prior to 1940
DIS	Weighted distance to five Boston employment centres
RAD	Index of accessibility to radial highways per town
TAX	Property-tax rate per US $10 000 per town
PTRATIO	Pupil – teacher ratio per town
B	Calculated as $1000(N_{Black} - 0.63)^2$ where N_{Black} is the proportion of Blacks
LSTAT	Percentage of lower status population

Using the Boston housing dataset, we compare the empirical performance of the LASSO (ESFLe or ESFLl, hereafter, e: SEM, and l: SLM) procedure with the forward stepwise procedure. Stepwise eigenvector selection based on minimizing standardized Moran's I (ESFMe or ESFMl, hereafter) can be implemented using SpatialFiltering() function of the spdep package in R[9]. Also, the selection based on maximizing explanation power can be implemented using basic step() function in R (ESFBe or ESFBl, hereafter). In the step() function, we set the option of penalty term as "$k = \log n$", which corresponds to the BIC (Bayesian information criterion)[10]. Thus, for

[9] Griffith and Peres-Neto (2006) also provide MATLAB code.

[10] AIC (Akaike information criterion) over-selected eigenvectors in our study. Generally speaking, when number of parameters exceed number of observations, AIC performed badly and BIC or corrected AIC should be used.

implementing stepwise procedure based on maximizing explanation power, we use not the conventional *F*- or *t*- statistics but the information criterion because the former criteria require arbitral parameters such as inclusion probability. Our all calculations are implemented on the Windows 7 64bit system with 24GB memory, and coded using R (version 3.0.1).

The Table 2 represents the parameter estimation results. It is noted that the estimate of the AGE variable of the BM is positive, but those of the ESFMl, ESFBl, and ESFLl are negative (although they are not statistically significant). These results may be caused by the omitted variables biases are improved as in the case of Pace and Gilley (1997).

The residual spatial autocorrelation of the ESFs in terms of Moran's *Z* are improved compared to that of the BM (BM: 14.5; ESFMe: –1.02; ESFMl: –0.800; ESFBe: 4.37; ESFBl: 1.94; ESFLe: 4.12; ESFBl: 2.95), but not random except for the ESFMs. However, some coefficient estimates of the selected eigenvectors of the ESFMs are not statistically significant at 5 % level, although those of the ESFBs and ESFLs are all statistically significant at least 10% level (Table 3). The performance of the ESFL is comparable to that of the ESFB in terms of Moran's *I* and adjusted R^2, and that computation time is much shorter (ESFMe: 34.01; ESFMl: 30.72; ESFBe: 87.94; ESFBl: 172.17; ESFLe: 17.89; ESFBl: 25.33). Hence, the results suggest that LASSO can be a useful compliment or alternative to the stepwise procedure.

Here, we see whether the ESFs select the eigenvectors associated with large eigenvalues or whether these are often associated with small eigenvalues. Figure 1 represents the cumulative number of selected eigenvectors (Left: ESFe, Right: ESFl). Associated eigenvalues are sorted in descending-order, and shown as a thick solid line on right vertical axis. We can find that ESFMs tend to select the eigenvectors associated with large eigenvalues for reducing the residual spatial autocorrelation, but ESFBs or the ESFLs sometimes select the eigenvectors associated with small eigenvalues, which contribute to the improvement of fit to the observations.

Figure 2 represents differences in cumulative number of selected eigenvectors by ten different kinds of random seed in the cross-validation procedure (Left: ESFe, Right: ESFl). Associated eigenvalues are also sorted in descending-order, and shown as a thick solid line on right vertical axis. The figure shows that the differences in the LASSO regulation parameter by the selection of random seed is fairly minor, and corresponding differences in the selected eigenvectors are small compared to the differences due to the selection of methods (ESFMs or ESFBs) (see Fig. 1). Hence, we can say that the LASSO is reasonably insensitive to the small changes in the regulation parameter produced by the cross-validation procedure.

Table 2. Parameter estimation results

Model	BM			ESFMe			ESFMl		
Variable	Coef.	*t*	*p*	Coef.	*t*	*p*	Coef.	*t*	*p*
(Intercept)	4.558	29.5	0.00	4.558	44.5	0.00	4.135	31.2	0.00
CRIM	-0.01186	-9.53	0.00	-0.01186	-14.4	0.00	-0.008429	-9.27	0.00
ZN	0.00008016	0.159	0.87	0.00008016	0.239	0.81	-0.0003881	-0.917	0.36
INDUS	0.0002395	0.101	0.92	0.0002395	0.153	0.88	0.003344	1.73	0.08
CHAS1	0.09140	2.75	0.01	0.09140	4.15	0.00	0.01930	0.759	0.45
NOX	-0.6380	-5.64	0.00	-0.6380	-8.51	0.00	-0.4736	-5.32	0.00
RM	0.006328	4.82	0.00	0.006328	7.28	0.00	0.009408	9.51	0.00
AGE	0.00009074	0.172	0.86	0.00009074	0.260	0.79	-0.0005576	-1.31	0.19
DIS	-0.1913	-5.73	0.00	-0.1913	-8.64	0.00	-0.1405	-4.61	0.00
RAD	0.09571	5.00	0.00	0.09571	7.55	0.00	0.07805	5.28	0.00
TAX	-0.0004203	-3.43	0.00	-0.0004203	-5.17	0.00	-0.0005423	-5.67	0.00
PTRATIO	-0.03112	-6.21	0.00	-0.03112	-9.37	0.00	-0.02183	-5.60	0.00
B	0.0003637	3.53	0.00	0.0003637	5.32	0.00	0.0004333	4.71	0.00
LSTAT	-0.3712	-14.8	0.00	-0.3712	-22.4	0.00	-0.3373	-17.5	0.00
Adjusted R^2	0.801			0.913			0.912		
Residual variance	0.0333			0.0146			0.0146		
Moran *I*	0.4364			-0.1633			-0.1257		
Moran *Z*	14.5			-1.02			-0.800		

Table 2. Parameter estimation results (cont.)

Model	ESFBe			ESFBl			ESFLe			ESFLl		
Variable	Coef.	*t*	*p*	Coef.	*t*	*p*	Coef.	*t*	*p*	Coef.	*t*	*p*
(Intercept)	4.466	39.9	0.00	3.849	34.4	0.00	4.558	43.0	0.00	4.138	34.5	0.00
CRIM	-0.01186	-13.6	0.00	-0.007867	-9.61	0.00	-0.01186	-13.9	0.00	-0.009278	-10.3	0.00
ZN	0.00008016	0.226	0.82	-0.0002603	-0.707	0.48	0.00008016	0.231	0.82	-0.0001728	-0.473	0.64
INDUS	0.0002395	0.144	0.89	0.005486	3.27	0.00	0.0002395	0.148	0.88	0.001252	0.727	0.47
CHAS1	0.09140	3.92	0.00	0.00471	0.217	0.83	0.09140	4.01	0.00	0.05853	2.61	0.01
NOX	-0.6380	-8.03	0.00	-0.4333	-5.75	0.00	-0.6380	-8.22	0.00	-0.5114	-6.21	0.00
RM	0.006328	6.86	0.00	0.01027	11.7	0.00	0.006328	7.03	0.00	0.009680	10.2	0.00
AGE	0.00009074	0.245	0.81	-0.0005719	-1.64	0.10	0.00009074	0.251	0.80	-0.0002386	-0.623	0.53
DIS	-0.1913	-8.15	0.00	-0.02877	-1.18	0.24	-0.1913	-8.35	0.00	-0.1497	-5.67	0.00
RAD	0.09571	7.12	0.00	0.08393	6.71	0.00	0.09571	7.29	0.00	0.06919	5.18	0.00
TAX	-0.0004203	-4.88	0.00	-0.0005175	-6.26	0.00	-0.0004203	-5.00	0.00	-0.0004140	-4.73	0.00
PTRATIO	-0.03112	-8.84	0.00	-0.02062	-6.17	0.00	-0.03112	-9.05	0.00	-0.02336	-6.48	0.00
B	0.0003637	5.02	0.00	0.0004582	6.37	0.00	0.0003637	5.14	0.00	0.0003814	4.42	0.00
LSTAT	-0.3712	-21.1	0.00	-0.3271	-19.3	0.00	-0.3712	-21.6	0.00	-0.3289	-18.1	0.00
Adjusted R^2	0.902			0.934			0.906			0.918		
Residual variance	0.0164			0.0110			0.0157			0.0138		
Lasso parameter							15.66			15.35		
Moran *I*	0.06420			-0.02638			0.05282			0.02035		
Moran *Z*	4.37			1.94			4.12			2.95		

Table 3. Computation time and selected eigenvectors

Model	ESFMe	ESFMl	ESFBe	ESFBl	ESFLe	ESFLl
Computation time (sec.)	34.01	30.72	87.94	172.17	17.89	25.33
Total # of eigenvectors	72	54	41	64	46	45
Significant at 0.1% level	21	21	20	31	21	23
Significant at 1% level	16	9	14	23	14	13
Significant at 5% level	12	14	7	10	11	8
Significant at 10% level	12	6	0	0	0	1
Not significant	11	4	0	0	0	0

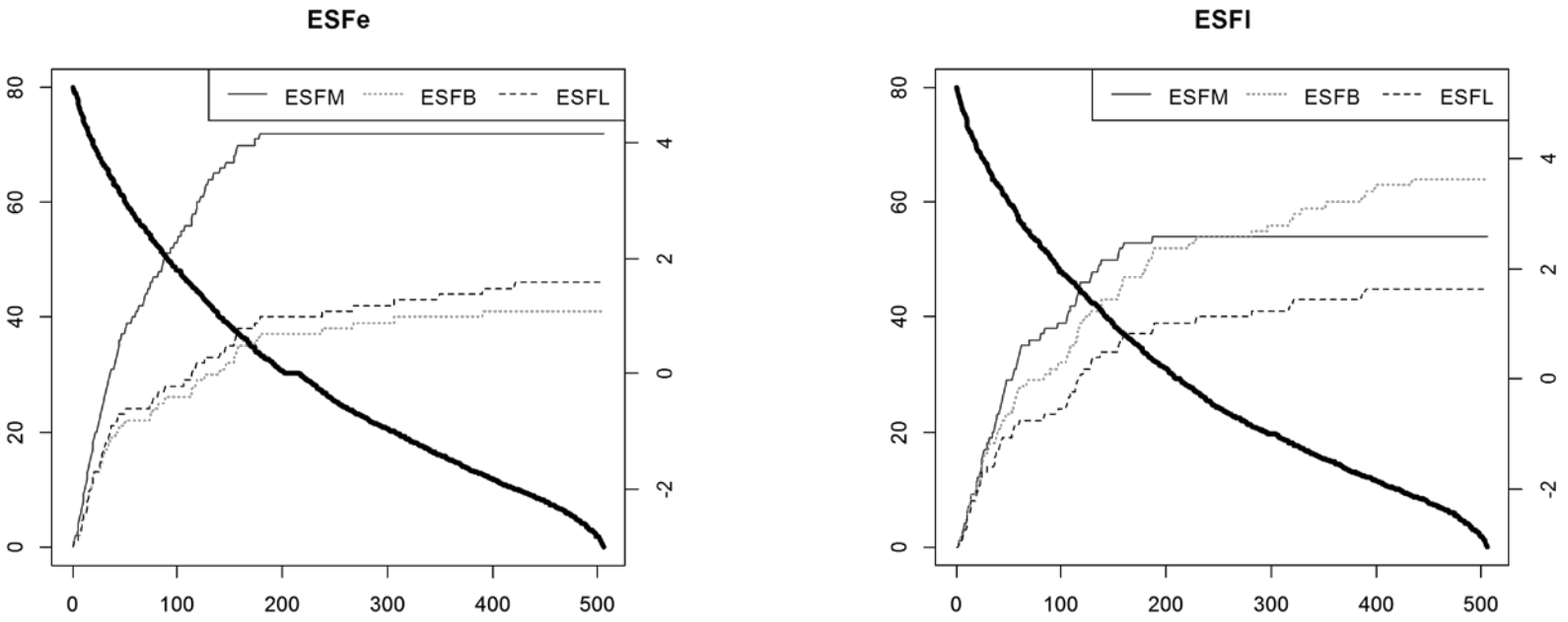

Figure 1. Cumulative number of selected eigenvectors (Left: ESFe, Right: ESFl)

(Associated eigenvalues are sorted in descending-order, and shown as a thick solid line on right vertical axis)

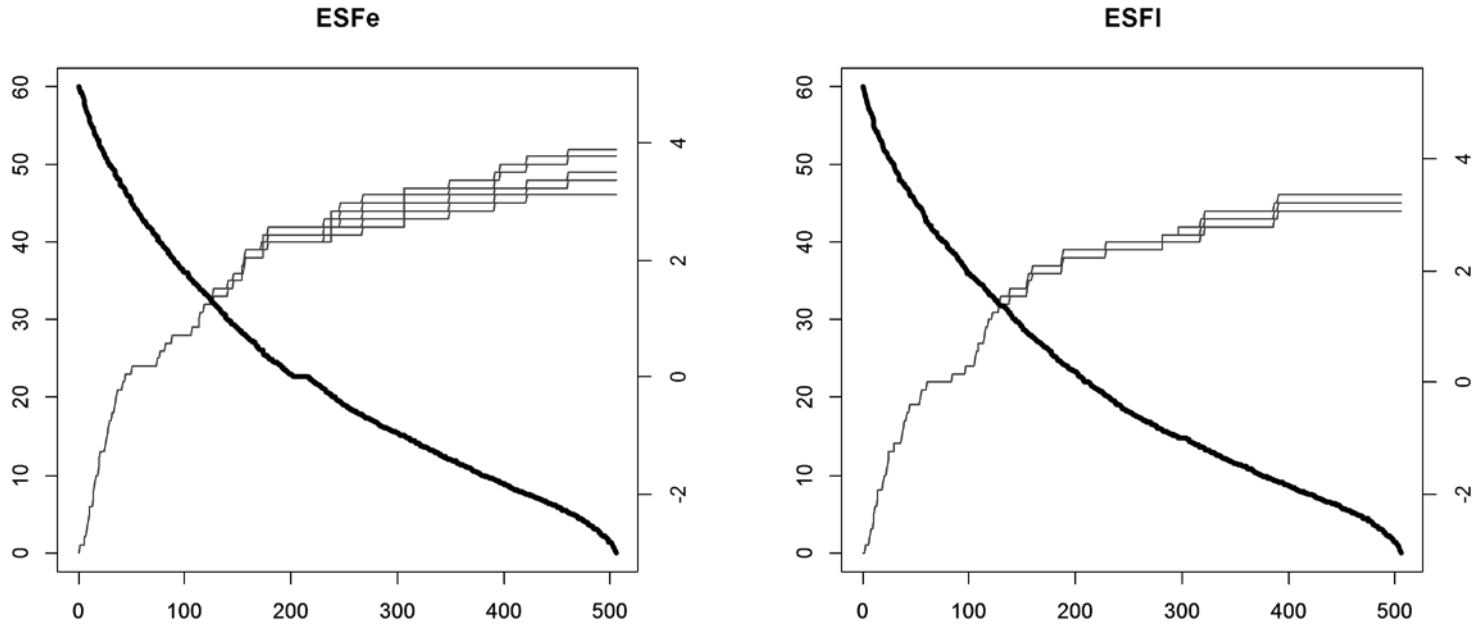

Figure 2. The differences in cumulative number of selected eigenvectors by random seed in the cross-validation procedure (Left: ESFe, Right: ESFl)
(Associated eigenvalues are sorted in descending-order, and shown as a thick solid line on right vertical axis)

Example with simulation dataset

In this section, we try to apply the LASSO procedure for the eigenvector selection problem of much bigger dataset. As the model, here we only consider the ESF with $\boldsymbol{E}_{SEM}$ because computation time of ESF with $\boldsymbol{E}_{SLM}$ and $\boldsymbol{E}_{SEM}$ is expected to be similar. Also, as the stepwise procedure, we only use the ESFM because the ESFB was much slower than ESFM in the previous experiment.

Subsequently, as the data generating process (DGP), we consider the spatial error process (SEP) expressed in the following manner; see Arbia (2006) for more details with regard to the spatially dependent DGPs.

$$\text{SEP:} \quad y_i = \beta_0 + \sum_{p=1}^{10} \beta_p x_{pi} + u_i , \quad u_i = \lambda \sum_{j=1}^{n} w_{ij} u_j + \varepsilon_i , \tag{13}$$

The error term in eq. (13) is assumed to obey a standard normal distribution $N(0,1)$. Sample size is set to n=50, 1000, 2000, and 10^4. Ten explanatory variables are independently generated from a standard normal distribution, and an intercept term is added. A spatial weight matrix is constructed using four-nearest-neighbors criterion where the (x, y) coordinates are generated from a standard normal distribution, and is row-normalized. With regard to the spatial parameter λ, we consider a positive and medium magnitude spatial autocorrelation (λ = 0.6). The true values of the parameters are set to β_p = 1 (p = 0, …, 10). Similar to the illustration with Boston housing dataset, we assume that the explanatory variables are always included in the model.

Tables 4 summarizes the parameter estimation results for n = 500 scenario. It is shown that adjusted R^2 values are improved in the ESFs compared to the BM (from 0.875 to 0.942 or 0.957); and correspondingly, the t-values are much larger in ESFs. Eigenvectors selected by the LASSO are all statistically significant at 5 % level (Table5), resulting in the high R^2 value compared to the ESFM. However, just as the Boston housing data experiment, the residual spatial autocorrelation of the LASSO in terms of Moran's I is not random although rather improved (from 15.5 to 3.28). Hence as expected, if we hope to minimize residual autocorrelation, not ESFB or ESFL but ESFM is the first best choice. Instead, if we hope to maximize fit to observations, maybe ESFB or ESFL are the better choice in medium sample size (e.g., n = 500).

However, the SpatialFiltering() function becomes disappointingly slow when sample size becomes large. For example, for n = 2000 scenario, ESFM requires 7800 second for selecting eigenvector whereas the ESFL requires only 349 second (Table6). In such case, our proposed application of the LASSO can be an important alternative to the stepwise procedures. The LASSO is feasible even the sample size is n = 10^4 (45942.71 sec.), which is a plausible sample size for applied researches.

One of the future challenges is dealing with much bigger dataset order of n = 10^5 or n = 10^6. In such scenarios, ex-ante determination of the number of eigenvectors added to a model may be

needed. Pace et al. (2013), proposed the use of ARPACK proceeds in such direction. Also, because majority of computation time is devoted to the cross-validation combined with the Brent algorithm, the other ways to select eigenvectors not depending on cross-validation, such as Fan and Tang (2013), may needed to be further developed.

Table 4. Parameter estimation results (n = 500)

Model	BM			ESFMe			ESFLe		
Variable	Coef.	*t*	*p*	Coef.	*t*	*p*	Coef.	*t*	*p*
(Intercept)	3.414	62.2	0.00	3.414	91.5	0.00	3.414	106	0.00
x1	0.9908	18.3	0.00	0.9908	26.9	0.00	0.9908	31.0	0.00
x2	0.9716	18.7	0.00	0.9716	27.5	0.00	0.9716	31.7	0.00
x3	0.9504	17.5	0.00	0.9504	25.7	0.00	0.9504	29.7	0.00
x4	0.9422	18.4	0.00	0.9422	27.1	0.00	0.9422	31.3	0.00
x5	1.028	19.5	0.00	1.028	28.7	0.00	1.028	33.1	0.00
x6	1.040	19.3	0.00	1.040	28.4	0.00	1.040	32.7	0.00
x7	0.9454	18.3	0.00	0.9454	27.0	0.00	0.9454	31.1	0.00
x8	0.9868	18.1	0.00	0.9868	26.7	0.00	0.9868	30.8	0.00
x9	0.9557	17.0	0.00	0.9557	25.0	0.00	0.9557	28.8	0.00
x10	1.101	19.9	0.00	1.101	29.2	0.00	1.101	33.7	0.00
Adjusted R^2	0.875			0.942			0.957		
Residual variance	1.48			0.685			0.514		
Lasso parameter	---			---			12.02		
Moran's *I*	0.4574			-0.09128			0.02849		
Moran's *Z*	15.5			0.0909			3.28		

Table 5. Computation time and selected eigenvectors

	ESFMe	ESFLe
Computation time (sec.)	26.07	15.15
Total # of eigenvectors	43	68
Significant at 0.1% level	16	24
Significant at 1% level	11	23
Significant at 5% level	10	21
Significant at 10% level	6	0
Not significant	0	0

Table 6. Computation time and sample sizes

Computation time (sec.)	ESFMe	ESFLe	(P1)	(P2)	(P3)
n = 500	26.07	15.15	0.28	0.47	14.4
n = 1000	424.18	64.52	1.94	2.1	60.48
n = 2000	7800.02	349.09	17.09	14.9	317.1
n = 10000	---	45942.71	2455.97	2676.78	40809.96

(P1): Calculating projection matrix $\boldsymbol{N}^{[x]}$; (P2): Extracting eigenvectors; (P3): Cross validation with Brent algorithm

Concluding remarks

The ESF is one of the well-used used approaches to model spatial autocorrelation among observations or errors in a regression model. In this approach, subset of eigenvectors extracted from a modified spatial weight matrix is added to the model as explanatory variables. The subset is typically specified by the forward stepwise model selection procedure, but it is disappointingly slow when the number of observations *n* takes a large number. Hence the present paper proposed the use of the LASSO to select the eigenvectors. The LASSO model selection procedure was applied to the well-known Boston housing dataset and simulation dataset, and its performance was compared to the stepwise procedure. The obtained results suggest that the LASSO procedure is fairly fast compared the stepwise procedure, and can select eigenvectors effectively even if dataset is relatively large (n = 10000), to which the forward stepwise procedure is not easy to apply. The results suggest that the LASSO can be a useful complement or even alternative to the stepwise procedures.

In the future research, we are going to test the feasibility of the LASSO under the huger dataset with combining the large scale eigenvalue problem solver such as ARPACK (Pace et al. 2013). Also, applying the LASSO based eigenvector selection approach to the other models such as generalized linear model (GLM) is an important remains researches. Furthermore, we are now trying to develop a more efficient approach for the eigenvector selection based on subspace method (Bagan et al. 2009).

Acknowledgement

This study was funded by the “Research Program on Climate Change Adaptation” of the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References

- Arbia, G. (2006). *Spatial Econometrics: Statistical Foundations and Applications to Regional Convergence*. Berlin: Springer.
- Arbia, G. (2011). “A Lustrum of SEA: Recent Research Trends Following the Creation of the Spatial Econometrics Association (2007–2011).” *Spatial Economic Analysis* 6, 377–95.
- Anselin, L. (1986). “Some Further Notes on Spatial Models and Regional Science.” *Journal of Regional Science* 26, 799–802.
- Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer.
- Anselin, L. (2010). “Thirty Years of Spatial Econometrics.” *Papers in Regional Scienc*e 89, 3–25.
- Anselin, L. (2012). “From SpaceStat to CyberGIS Twenty Years of Spatial Data Analysis Software.” *International Regional Science Review* 35, 131–57.
- Bagan, H., W. Takeuchi, Y. Yamagata, X. Wang, and Y. Yasuoka. (2009). “Extended Averaged

Learning Subspace Method for Hyperspectral Data Classification." *Sensors* 9, 4247–70.

- Banerjee, S., B.P. Carlin, and A.E. Gelfand. (2004). *Hierarchical Modeling and Analysis for Spatial Data*. Boca Raton: Chapman & Hall/CRC.
- Bhattacharjee, A. and C. Jensen-Butler. (2013). "Estimation of the Spatial Weights Matrix Under Structural Constraints." *Regional Science and Urban Economics* 43, 617–34.
- Brasington, D.M. and D. Hite. (2005). "Demand for Environmental Quality: A Spatial Hedonic Analysis." *Regional Science and Urban Economics* 35, 57–82.
- Brent, R.P. (1973). *Algorithms for Minimization without Derivatives*. Englewood Cliffs: Prentice-Hall.
- Candes, E. and T. Tao. (2007). "The Dantzig Selector: Statistical Estimation When *p* is Much Larger than *n*." *The Annals of Statistics* 35, 2313–51.
- Chen, X., Z.J. Wang and M.J. McKeown. (2011). "A Bayesian LASSO via Reversible-jump MCMC." *Signal Processing* 91, 1920–32.
- Chun, Y. and D.A. Griffith. (2011). "Modeling Network Autocorrelation in Space-time Migration Flow Data: An Eigenvector Spatial Filtering Approach." *Annals of the Association of American Geographers* 101, 523–36.
- Chun, Y. and D.A. Griffith. (2013). *Spatial Statistics and Geostatistics: Theory and Applications for Geographic Information Science and Technology*, London: SAGE Publications Ltd.
- Cressie, N. (1993). *Statistics for Spatial Data, Revised Edition*. New York: Wiley.
- Cuaresma, J.C. and M. Feldkircher. (2013). "Spatial Filtering, Model Uncertainty and the Speed of Income Convergence in Europe." *Journal of Applied Econometrics* 28,720–41.
- Demšar, U., P. Harris, C. Brunsdon, A.S. Fotheringham, and S. McLoone. (2013). "Principal Component Analysis on Spatial Data: An Overview." *Annals of the Association of American Geographers* 103, 106–28.
- Diniz-Filho, J.A.F. and L.M. Bini. (2012). "Thirty-five Years of Spatial Autocorrelation Analysis in Population Genetics: An Essay in Honour of Robert Sokal (1926–2012)." *Biological Journal of the Linnean Society* 107, 721–36.
- Dormann, C.F., S.M. Bierman, J.M. McPherson, M.B. Araujo, R. Bivand, J. Bolliger, G. Carl, R.G. Davies, A. Hirzel, W. Jetz, W.D. Kissling, I. Kuhm, R. Ohlemuller, P.R. Peres-Neto, B. Reineking, B. Schroder, F.M. Schurr, and R. Wilson. (2007). "Methods to Account for Spatial Autocorrelation in the Analysis of Species Distributional Data: A Review." *Ecography* 30, 609–28.
- Dubin, R., R.K. Pace, and T.G. Thibodeau. (1999). "Spatial Autoregression Techniques for Real Estate Data." *Journal of Real Estate Literature* 7, 79–96.
- Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. (2004). "Least Angle Regression." *Annals*

of Statistics 32, 407–99.

- Fan, Y. and C.Y. Tang. (2013). “Tuning Parameter Selection in High Dimensional Penalized Likelihood.” *Journal of the Royal Statistical Society B* 75, 531–52.
- Fischer, M.M., T. Scherngell, M. Reismann. (2009). “Knowledge Spillovers and Total Factor Productivity: Evidence Using a Spatial Panel Data Model.” *Geographical Analysis* 41, 204–20.
- Gelfand, A.E., P.J. Diggle, M. Fuentes, and P. Guttorp, Editors. (2010). *Handbook of Spatial Statistics*, Boca Raton, Chapman & Hall/CRC.
- Getis, A. (1990). “Screening for Spatial Dependence in Regression Analysis.” *Papers of the Regional Science Association* 69, 69–81.
- Getis, A. (2009). “Spatial Weights Matrices.” *Geographical Analysis* 41, 404–10.
- Getis, A. and D.A. Griffith. (2002). “Comparative Spatial Filtering in Regression Analysis.” *Geographical Analysis* 34, 130–40.
- Gibbons, S. and H.G. Overman. (2012). “Mostly Pointless Spatial Econometrics.” *Journal of Regional Science* 52, 172–91.
- Gilley, O.W. and R.K. Pace. (1996). “On the Harrison and Rubinfeld Data.” *Journal of Environmental Economics and Management* 31, 403–5.
- Goeman, J. (2010). “L1 Penalized Estimation in the Cox Proportional Hazards Model.” *Biometrical Journal* 52, 70–84.
- Goeman, J., R. Meijer and N. Chaturvedi (2012). “L1 and L2 Penalized Regression Models.” Documentation for the penalized Package for the Statistical Environment R.
- Griffith, D.A. (2000). “A Linear Regression Solution to the Spatial Autocorrelation Problem.” *Journal of Geographical Systems* 2, 141–56.
- Griffith, D.A. (2003). *Spatial Autocorrelation and Spatial Filtering: Gaining Understanding through Theory and Scientific Visualization*. Berlin: Springer.
- Griffith, D.A. (2012). “Spatial Statistics: A Quantitative Geographer’s Perspective.” *Spatial Statistics* 1, 3–15.
- Griffith, D.A. and J.H.P. Paelinck. (2007). “An Equation by Any Other Name is Still the Same: On Spatial Econometrics and Spatial Statistics.” *The Annals of Regional Science* 41, 209–27.
- Griffith, D.A. and J.H.P. Paelinck. (2010). *Non-standard Spatial Statistics and Spatial Econometrics*. Berlin: Springer.
- Griffith, D.A. and P.R. Peres-Neto. (2006). “Spatial Modeling in Ecology: The Flexibility of Eigenfunction Spatial Analyses in Exploiting Relative Location Information.” *Ecology* 87, 2603–13.
- Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences.* Cambridge: Cambridge University Press.
- Haining, R. (2003). *Spatial Data Analysis: Theory and Practice*. Cambridge: Cambridge

University Press.

- Harrison, D.Jr. and D.L. Rubinfeld. (1978). “Hedonic Housing Prices and the Demand for Clean Air.” *Journal of Environmental Economics and Management* 5, 81–102.
- Hodges, J.S. and Reich, B.J. (2010). “Adding Spatially-Correlated Errors Can Mess Up the Fixed Effect You Love.” *The American Statistician* 64, 325–34.
- Huang, H.-C., N.-J. Hsu, D. Theobald, and F.J. Breidt. (2010). “Spatial LASSO with Applications to GIS Model Selection.” *Journal of Computational and Graphical Statistics* 19, 963–83.
- Hughes, J. and M. Haran. (2013). “Dimension Reduction and Alleviation of Confounding for Spatial Generalized Linear Mixed Models.” *Journal of the Royal Statistical Society B* 75, 139–59.
- Kakamu, K. and H. Wago. (2008). “Small-sample Properties of Panel Spatial Autoregressive Models: Comparison of the Bayesian and Maximum Likelihood Methods.” *Spatial Economic Analysis* 3, 305–19.
- Keerthi, S.S. and S. Shevade, S. (2007). “A Fast Tracking Algorithm for Generalized LARS/LASSO.” *IEEE Transactions on Neural Networks* 18, 1826–30.
- Kelejian, H.H. and I.R. Prucha. (1998). “A Generalized Spatial Two-stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances.” *Journal of Real Estate Economics* 17, 99–121.
- Kelejian, H.H. and I.R. Prucha. (2010). “Specification and Estimation of Spatial Autoregressive Models with Autoregressive and Heteroskedastic Disturbances.” *Journal of Econometrics* 157, 53–67.
- Kostov, P. (2010). “Model Boosting for Spatial Weighing Matrix Selection in Spatial Lag Models.” *Environment and Planning B* 37, 533–49.
- Kyung, M., J. Gill, M. Ghosh, and G. Casella. (2010). “Penalized Regression, Standard Errors, and Bayesian LASSOs.” *Bayesian Analysis* 5, 369–411.
- Lee, L.-F. (2004). “Asymptotic Distributions of Quasi-maximum Likelihood Estimators for Spatial Autoregressive Models.” *Econometrica* 72, 1899–925.
- Lee, L.-F. (2007). “GMM and 2SLS Estimation of Mixed Regressive, Spatial Autoregressive Models.” *Journal of Econometrics* 137, 489–514.
- LeSage, J.P. (1997). “Bayesian Estimation of Spatial Autoregressive Models.” *International Regional Science Review* 20, 113–31.
- LeSage, J.P., and R.K. Pace. (2009). *Introduction to Spatial Econometrics*, Boca Raton: Taylor & Francis.
- Ord, J.K. (1975). “Estimation Methods for Models of Spatial Interaction.” *Journal of American Statistical Association* 79, 120–26.

- Pace, R.K., R. Barry, and C.F. Sirmans. (1998). “Spatial Statistics and Real Estate.” *The Journal of Real Estate Finance and Economics* 17, 5–13.
- Pace, R.K. and O.W. Gilley. (1997). “Using the Spatial Configuration of the Data to Improve Estimation.” *Journal of Real Estate Finance and Economics* 14, 333–40.
- Paciorek, C.J. (2010). “The Importance of Scale for Spatial-Confounding Bias and Precision of Spatial Regression Estimators.” *Statistical Science* 25, 107–25.
- Pace, R.K., J.P. LeSage, and S. Zhu. (2013). “Interpretation and Computation of Estimates from Regression Models Using Spatial Filtering.” Spatial Economic Analysis 8, 352–69.
- Páez, A. and D.M. Scott. (2005). “Spatial Statistics for Urban Analysis: A Review of Techniques with Examples.” *GeoJournal* 61, 53–67.
- Park, M.Y. and T. Hastie. (2007). “L1-Regularization Path Algorithm for Generalized Linear Models.” *Journal of the Royal Statistical Society B* 69, 659–77.
- Park, M.Y. and G. Cassella. (2008). “The Bayesian LASSO.” *Journal of the American Statistical Association* 103, 681–6.
- Patuelli, R., D.A. Griffith, M. Tiefelsdorf, and P. Nijkamp. (2007). “Spatial Filtering and Eigenvector Stability: Space-time Models for German Unemployment Data.” *International Regional Science Review* 34, 253–80.
- Rue, H. and L. Held. (2005). *Gaussian Markov Random Fields: Theory and Applications*. London: Chapman and Hall/CRC.
- Schabenberger, O. and C.A. Gotway. (2005). *Statistical Methods for Spatial Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- Segal, M.R. (2006). “Microarray Gene Expression Data with Linked Survival Phenotypes: Diffuse Large-B-cell Lymphoma Revisited.” *Biostatistics* 7, 268–85.
- Seya, H. and M. Tsutsumi. (2011). “Application of Model Averaging Techniques to Spatial Hedonic Land Price Models.” *Econometrics: New Developments*, 63–88, edited by S.A. Mendez, and A.M. Vega. New York: NOVA Science Publishers.
- Seya, H., Y. Yamagata, and M. Tsutsumi. (2013). “Automatic Selection of a Spatial Weight Matrix in Spatial Econometrics: Application to a Spatial Hedonic Approach.” *Regional Science and Urban Economics* 43, 429–44.
- Stakhovych, S., and T.H.M. Bijmolt. (2009). “Specification of Spatial Models: A Simulation Study on Weights Matrices.” *Papers in Regional Science* 88, 389–408.
- Tibshirani, R. (1996). “Regression Shrinkage and Selection via the LASSO.” *Journal of the Royal Statistical Society B* 58, 267–88.
- Tiefelsdorf, M. and D.A. Griffith. (2007). “Semiparametric Filtering of Spatial Auto-correlation: The Eigenvector Approach.” *Environment and Planning A* 39, 1193–221.
- Tsutsumi, M., E. Shimizu, and Y. Matsuba. (1997). “A Comparative Study on Counter-measures

for Multicollinearity in Regression Analysis." *Journal of the Eastern Asia Society for Transportation Studies* 2, 1891–904.

- Wheeler, D.C. (2009). "Simultaneous coefficient penalization and model selection in geographically weighted regression: the geographically weighted LASSO." *Environment and Planning A* 41, 722–42.
- Yamagata, Y., H. Seya and D. Murakami. (2012) "New Global Research on the International Network-based Green Growth Analysis with Visualization." presented at the IIASA 40th Anniversary Conference, Vienna, October 24–26, 2012, (http://conference2012.iiasa.ac.at/poster_session.html)
- Zou, H. and T. Hastie. (2005). "Regularization and Variable Selection via the Elastic Net." *Journal of the Royal Statistical Society B* 67, 301–20.

Appendix

R code for the illustration with Boston housing dataset