Modeling and analyzing missing spatial data by spatial autoregressive models and vectorized maximum likelihood method

Samira Zahmatkesh, Mohsen Mohammadzadeh*;
Department of Statistics, Tarbiat Modares University, Tehran, Iran

Received: 1 August 2020 Accepted: 23 September 2020

Extended Abstract

Paper pages (756-772)

Introduction

There is often a considerable amount of missingness in spatial data, due to the conditions of their collection. In spatial data, the intensity of the dependence between observations in the vicinity is stronger than the dependency of further observations. This feature affects the modeling and statistical inferences of the data. Thus, missing values that are close to each other or observations may contain useful information that can be used to lead to more accurate results in data analysis by reconstructing the data and imputation. It is important to consider the appropriate hypothesis about the missing data mechanism. Generally missing data mechanisms can be divided into two categories, ignorable and non-ignorable missingness. Here we continue under the ignorable missingness hypothesis.

Autoregressive models can be used efficiently to model missing spatial data. Considering the spatial autoregressive coefficient in the model, these models correct the prediction of the response variable based on the linear regression model $y = X\beta + \varepsilon$ through the weighted mean of the values in the neighborhood of observations. In this kind of spatial modeling, the spatial dependence of the observations is considered through an appropriate spatial weight matrix, and the intensity of the dependence between observations is described through a spatial autoregressive coefficient. What is essential in using these models is to use a suitable method to obtain an estimate of the model parameters and thus predict in unobserved situations. Studies have shown that in using these models, estimating the maximum likelihood of the model parameters leads to time-consuming calculations and local maximums. An ideal estimator should be able to use high-volume data, also have the ability for fast computations, and should not rely on nonlinear optimization algorithms that provide only a localized rather than a comprehensive value as the

optimal value. A method was proposed for quickly calculating estimates when the dependent variable follows a spatial autoregressive process which is known as the "vectorized maximum likelihood". From a computational point of view, vectoring relative to a parameter of interest avoids the costs of using nonlinear optimizers, which are typically associated with iteration.

Material and methods

In this paper, we have applied the "vectorized maximum likelihood" to analyze missing spatial data. Also, simulation studies and a practical example to evaluate the performance of the method under study are presented.

Results and discussion

According to the results of the simulation study and modeling US election data, modelling data by spatial autoregressive model, estimating and predicting through the introduced method leads better results than the conventional ordinary least square method. As an example, in the analysis of election data via the Spatial Durbin Model (SDM) and vectorized maximum likelihood method, some coefficient relevant to spatially lagged covariates are significant which are ignored through linear regression. This can be the weak points of a simple model.

Conclusion

Also, the adjusted determinative coefficient, the sum of square errors, and root mean square error indicate that the spatial Durbin model, which is fitted by vectorized maximum likelihood method, leads to less prediction error in comparison with simple linear regression and ordinary least square method.

Keywords: Missing spatial data, Spatial autoregressive models, Vectorized maximum likelihood, Weight matrix.

*Corresponding author: mohsen_m@modares.ac.ir