Spatial Statistics 21 (2017) 241–261

Contents lists available at ScienceDirect

Spatial Statistics

journal homepage: www.elsevier.com/locate/spasta

Introducing bootstrap methods to investigate coefficient non-stationarity in spatial regression models

Paul Harris [a,*], Chris Brunsdon [b], Binbin Lu [c], Tomoki Nakaya [d], Martin Charlton [b]

[a] *Sustainable Agricultural Sciences, Rothamsted Research, North Wyke, Okehampton, Devon, EX20 2SB, UK*
[b] *National Centre for Geocomputation, Maynooth University, Maynooth, Ireland*
[c] *School of Remote Sensing and Information Engineering, Wuhan University, 129 Luoyu Road, Wuhan 430079, China*
[d] *Department of Geography, Ritsumeiken University, 56-1, Tojin-kita-machi, Kita-ku Kyoto, 603-8577, Japan*

ARTICLE INFO

Article history:
Received 29 January 2017
Accepted 24 July 2017
Available online 29 July 2017

Keywords:
Geographically weighted regression
Spatial regression
Hypothesis testing
Collinearity
GWmodel

ABSTRACT

In this simulation study, parametric bootstrap methods are introduced to test for spatial non-stationarity in the coefficients of regression models. Such a test can be rather simply conducted by comparing a model such as geographically weighted regression (GWR) as an alternative to a standard linear regression, the null hypothesis. In this study however, three spatially autocorrelated regressions are also used as null hypotheses: (i) a simultaneous autoregressive error model; (ii) a moving average error model; and (iii) a simultaneous autoregressive lag model. This expansion of null hypotheses, allows an investigation as to whether the spatial variation in the coefficients obtained using GWR could be attributed to some other spatial process, rather than one depicting non-stationary relationships. The new test is objectively assessed via a simulation experiment that generates data and coefficients with known multivariate spatial properties, all within the spatial setting of the oft-studied Georgia educational attainment data set. By applying the bootstrap test and associated contextual diagnostics to pre-specified, area-based, geographical processes, our study

* Corresponding author.
E-mail address: paul.harris@rothamsted.ac.uk (P. Harris).

http://dx.doi.org/10.1016/j.spasta.2017.07.006
2211-6753/© 2017 Elsevier B.V. All rights reserved.

provides a valuable steer to choosing a suitable regression model for a given spatial process.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

Often when fitting a regression model to spatial data, it is not clear what, if any, spatial effects should be accounted for. Should we focus solely on spatial autocorrelation effects (e.g. Anselin, 1988; Cressie, 1993) or should we focus solely on spatial heterogeneity effects with respect to data relationships (e.g. Fotheringham et al., 2002). Alternatively, should we try to capture both effects (e.g. Haas, 1996; Brunsdon et al., 1998a; Mur et al., 2008; Cho et al., 2010; Harris et al., 2010a; Kim et al., 2010), or investigate ways to link (e.g. Griffith, 2003, 2008; Murakami et al., 2017), or fuse them together (e.g. Gelfand et al., 2003; Finley, 2011), and if so, which are more important? Further, should we ignore both effects altogether, and instead focus on a non-spatial model that is additionally calibrated with key spatial predictor variables, such as the sample coordinates (e.g. Beale et al., 2010)? Further still, should we consider that we are missing vital predictors and that any observed spatial effects are attributable to this omission (e.g. Cressie and Chan, 1989)—and as such, focus our attention on capturing these (likely elusive) missing variables? Unfortunately, such questions are almost always difficult to answer with any objectively, and can involve problematic analytical impasses and confounders (e.g. Anselin, 1990). For example, how to identify first- from second-order effects (e.g. Armstrong, 1984), where relationship heterogeneity is commonly modelled as the former, whist autocorrelation is modelled as the latter effect? These issues are particularly pertinent for spatial data sets, as their collection are rarely part of a statistically-designed experiment—that by definition should negate confounders.

Given such issues, it is commonplace to ignore them, and instead a regression for spatial data is often chosen following a rather subjective exploratory analysis that is itself pre-defined according to the given research hypothesis and/or sometimes biased towards the particular statistical expertise of the analyst. Thus, our study aim is to provide objectivity to a particular aspect of this model selection process, where we introduce parametric bootstrap methods to test for spatial non-stationarity in the coefficients of regression models. The tests are general and can be used to compare any spatially-varying coefficient (SVC) regression as an alternative to any set of constant coefficient regressions (with or without spatial autocorrelation effects). As demonstration, we compare geographically weighted regression (GWR) (Brunsdon et al., 1996, 1998b) as an alternative to the following four null hypotheses: (i) a multiple linear regression model (MLR), (ii) a simultaneous autoregressive error model (ERR); (iii) a moving average error model (SMA); and (iv) a simultaneous autoregressive lag model (LAG). This set of null hypotheses, allows an investigation as to whether the spatial variation in the coefficients obtained using GWR could be attributed to some other spatial process (in this case, some autocorrelation effect), rather than one depicting non-stationary relationships.

To achieve this, a bootstrapping methodology (Efron, 1979, 1981, 1982) is proposed that assesses the variability of the local coefficient estimates found from GWR under the model assumptions for each of the four null hypotheses (i.e. the MLR, ERR, SMA and LAG models). The observed values of coefficient variability are then compared against these as reference distributions. Our bootstrapping methodology complements the bootstrap methods to test for zero coefficients in a mixed GWR model (Mei et al., 2006) and constant coefficients in a basic GWR model in order to specify a mixed GWR model (Mei et al., 2016). Neither studies however, compare GWR with alternative (spatially-autocorrelated) regressions, as we do here. Our paper is structured as follows. Firstly, the study regressions are formally stated; the concept of bootstrapping is reviewed; and our spatial application of bootstrapping is outlined. Secondly, the described methodology is objectively assessed via a simulation experiment based on cokriging (Matheron, 1970) that generates data and regression coefficients, each with known multivariate spatial properties, and all within the spatial setting of the Georgia educational attainment data set (Fotheringham et al., 2002; Griffith, 2008). We complement and contextualise the bootstrap results with associated diagnostics. Thirdly, we discuss and conclude this research.

Download English Version:

https://daneshyari.com/en/article/5119001

Download Persian Version:

https://daneshyari.com/article/5119001

Daneshyari.com