Journal of Econometrics 207 (2018) 92–113

Contents lists available at ScienceDirect

Journal of Econometrics

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

A robust test for network generated dependence[☆]

Xiaodong Liu[a], Ingmar R. Prucha[b,*]

[a] Department of Economics, University of Colorado Boulder, Boulder, CO 80309, USA
[b] Department of Economics, University of Maryland, College Park, MD 20742, USA

ARTICLE INFO

Article history:
Received 30 June 2016
Received in revised form 25 April 2018
Accepted 21 May 2018
Available online 23 July 2018

JEL classification:
C12
C21

Keywords:
Test for network dependence
Generalized Moran $\mathcal{I}$ test
LM test
Laplace approximation
Network endogeneity
Network specification
Robustness

ABSTRACT

The paper introduces a robust testing procedure for network generated cross sectional dependence in the endogenous variables, exogenous variables and/or disturbances. Empirical researchers often face situations where they are unsure about how to model the proximity between cross sectional units in a network. The tests considered provide the empirical researcher an important degree of robustness in such situations. They generalize the Moran (1950) I test for dependence in spatial networks. The asymptotic properties of the tests are established under general conditions. The paper also discusses the use of the test statistics in situations where the network topology is endogenous.
© 2018 Elsevier B.V. All rights reserved.

1. Introduction

The paper introduces a robust test for network generated cross sectional dependence, and derives the statistical properties of the test. Empirical researchers often face situations where they are unsure about how to model the proximity between cross sectional units in a network. The test considered in this paper is aimed at providing to the empirical researcher an important degree of robustness in such situations.

As remarked by Kolaczyk (2009), "... during the decade surrounding the turn of the 21st century network-centric analysis ... has reached new levels of prevalence and sophistication". Applications range widely from physical and mathematical sciences to social sciences and humanities. The importance of network dependencies has, in particular, been recognized early in the regional science, urban economics and geography literature. The focus of this literature is on spatial networks. An important class of spatial network models was introduced by Cliff and Ord (1973, 1981), where, as a formal modeling device, weight matrices are used to capture the existence and directional importance of links in a spatial network. It is important to note that in the Cliff–Ord type models the weights are only viewed as related to a measure of proximity between units,

☆ We thank the editor Oliver Linton, the associate editor, and three anonymous referees for helpful comments and suggestions. We would also like to thank the participants of the 26th EC2 Conference on Theory and Practice of Spatial Econometrics, Edinburgh, the Western Regional Association 56th Meeting Annual Meeting, Santa Fe, the XIth World Congress of the Spatial Econometrics Association, Singapore, and participants at seminars at the ETH Zuerich and the London School of Economics for their helpful comments.

* Corresponding author.
E-mail addresses: xiaodong.liu@colorado.edu (X. Liu), prucha@econ.umd.edu (I.R. Prucha).

https://doi.org/10.1016/j.jeconom.2018.05.005
0304-4076/© 2018 Elsevier B.V. All rights reserved.

but not necessarily to the geographic location of the units.[1] Thus by extending the notion of proximity from geographical proximity to economic proximity, technological proximity, social proximity, etc., these models are useful for a much wider class of applications with cross sectional interactions. This includes social interaction models as discussed in, e.g., Bramoullé et al. (2009), Lee et al. (2010), Liu and Lee (2010), and Kuersteiner and Prucha (2015). For instance, a simple social interaction model can be specified as

$$y_i = \lambda \sum_{j=1}^{n} w_{ij} y_j + \beta x_i + \gamma \sum_{j=1}^{n} w_{ij} x_j + u_i \qquad \text{and} \qquad u_i = \varepsilon_i + \rho \sum_{j=1}^{n} w_{ij} \varepsilon_j, \tag{1}$$

where y_i, x_i and ε_i represent, respectively, the outcome, observed exogenous characteristic, and unobserved individual heterogeneity of cross sectional unit i. The weight w_{ij} captures social proximity of i and j in the network. Suppose $w_{ij} = n_i^{-1}$ if i and j are friends and $w_{ij} = 0$ otherwise, where n_i denotes the number of friends of i. Then, using the terminology in Manski (1993), $\sum_{j=1}^{n} w_{ij} y_j$ is the average outcome of i's friends with the coefficient λ representing *the endogenous peer effect*, $\sum_{j=1}^{n} w_{ij} x_j$ is the average of the observable characteristics of i's friends with the coefficient γ representing *the contextual effect*, and $\sum_{j=1}^{n} w_{ij} \varepsilon_j$ is the average of the unobservable characteristics of i's friends with the coefficient ρ representing *the correlated effect*. Of course, the above model also covers the simple group-average model as a special case with $w_{ij} = (n_g - 1)^{-1}$ if i and j belong to the same group of size n_g and $w_{ij} = 0$ otherwise; compare, e.g., Lee (2007), Davezies et al. (2009), and Carrell et al. (2013).

In the spatial network literature one of the most widely used tests for cross sectional dependence is the Moran (1950) $\mathcal{I}$ test. This test statistic is formulated in terms of a normalized quadratic form of the variables to be tested for spatial dependence. Moran's original formulation assumed that the variables are observed and based the quadratic form on a weight matrix with zero or one elements, depending on whether or not two units were considered neighbors. Cliff and Ord (1973, 1981) considered testing for spatial dependence in the disturbance process of a classical linear regression model, and generalized the test statistic to a quadratic form of ordinary least squares residuals, allowing for general weight matrices.[2] They derived the finite sample moments of the test statistic under the assumption of normality. Burridge (1980) showed that the Moran $\mathcal{I}$ test can be interpreted as a Lagrange Multiplier (LM) test if the disturbance process under the alternative hypothesis is either a spatial autoregressive or spatial moving average process of order one. He also discusses its close conceptual connection to the Durbin–Watson test statistic in the time series literature. King (1980, 1981) demonstrated that the Moran I test is a Locally Best Invariant test, when the alternative is one-sided, and the errors come from an elliptical distribution. A more detailed discussion of optimality properties of the Moran I test is given in Hillier and Martellosio (2018), including a discussion of conditions under which the Moran I test is a Uniformly Most Powerful Invariant test. Kelejian and Prucha (2001) introduced a central limit theorem (CLT) for linear–quadratic forms, and used that result to establish the limiting distribution of the Moran $\mathcal{I}$ test statistic as $N(0, 1)$ under a fairly general set of assumptions. They allowed for heteroskedasticity, which facilitates, among other things, applications to models with limited dependent variables, and they introduced necessary modifications for the Moran $\mathcal{I}$ test statistic to accommodate endogenous regressors. Pinkse (1998, 2004) also considered Moran $\mathcal{I}$ flavored tests, including applications to discrete choice models.

Anselin (1988), Anselin and Rey (1991) and Anselin et al. (1996) considered LM and modified LM tests for spatial autoregressive model with spatially autoregressive disturbances, and provide extensive Monte Carlo results on their small sample properties. Baltagi and Li (2000), Baltagi et al. (2003, 2007) derived LM test for first order spatial panel data models, and also analyzed their small sample behavior based on an extensive Monte Carlo study.[3] For cross sectional data Born and Breitung (2011) considered LM tests for first order spatial dependence in the dependent variable and the disturbances, allowing for unknown heteroskedasticity. Baltagi and Yang (2013) considered small sample improved LM tests, and Yang (2015) provided a bootstrap refinement. Robinson and Rossi (2014) introduced improved LM tests based on an Edgeworth expansion.[4]

One problem in using the Moran $\mathcal{I}$ and available LM tests for spatial models is that researchers are often unsure about how to specify the weight matrix employed by the test. Take a spatial network as an example, the weight w_{ij} could be a binary indicator variable depending on whether or not i and j are neighbors, or w_{ij} could be specified as the inverse of the geographical distance between i and j, etc. Researchers may consequently adopt a sequential testing procedure based on different specifications of the weight matrix. The sequential testing procedure, however, raises issues regarding the overall significance level of the test. Motivated by this problem we define in this paper a single test statistic, which in its simplest form is defined as a weighted inner product of a vector of quadratic forms, with each quadratic form corresponding to a different weight matrix. In this sense the test statistic combines a set of Moran $\mathcal{I}$ tests into a single test.

---

1 This is in contrast to the literature on spatial random fields where units are indexed by location; see, e.g., Conley (1999) and Jenish and Prucha (2009, 2012) for contributions to the spatial econometrics literature.

2 Of course, in the absence of regressors this setup included the original Moran $\mathcal{I}$ test statistic as a special case.

3 Pesaran (2004), Pesaran et al. (2008) and Baltagi et al. (2012) also considered LM flavored tests for cross sectional dependence for panel data. Those tests are based on sample correlations and not specifically geared towards network generated dependence.

4 Of course, there is also a large literature on ML and GMM estimation of Cliff–Ord type spatial models, which allows for likelihood ratio and Wald-type testing of the significance of spatial autoregressive parameters. Robinson and Rossi (2015) introduced an alternative test based on the biased OLS estimator for a first order spatial autoregressive model, using Edgeworth expansions and the bootstrap to appropriately adjust the size of the test.

Download English Version:

https://daneshyari.com/en/article/11020489

Download Persian Version:

https://daneshyari.com/article/11020489

Daneshyari.com