Moran's I Tests for Spatial Dependence in Panel Data Models with Time Varying Spatial Weights Matrices

Bian-Ling OU[1,a,*]

[1]Central University of Finance and Economics, Beijing 100081, China

[a]oubianling@amss.a.cn

*Corresponding author

Keywords: Moran, Time varying spatial weights matrices, Size, Power, Panel data model

Abstract. Moran's I statistic is the most popular test for spatial dependence. When spatial weights matrices are substantially varying over time, Moran's I test based on a time invariant spatial weights matrix may cause substantial bias. This paper first investigates Moran's I tests for spatial dependence in panel data models where spatial weights matrices can be time varying. Based on time varying and time invariant spatial weights matrices, the empirical size and power of Moran's I tests for spatial dependence are evaluated and compared. Monte Carlo results indicate that size of Moran's I tests based on time varying and misspecification of time invariant spatial weights matrices have not significant difference, especially compared with misspecification time invariant spatial weights matrices, power of Moran's I tests for spatial dependence with time varying spatial weights matrices is much higher. TV-Moran tests are superior to NTV-Moran tests with the misspecification of invariant spatial weights matrix, with larger power.

Introduction

Spatial econometrics is a subfield of econometrics that deals with spatial effects (Anselin, 1988). In recent several years, spatial econometrics are developed and applied in economics, regional science and geography (Brasili et al., 2012), agricultural economics (Druska and Horrace, 2004), public economics (Egger et al., 2005), transportation research (Frazier and Kockelman, 2005), good demand (Baltagi and Li, 2006), and so on. Both model estimation and identification are important in spatial econometric analysis. Hypothesis testing for spatial dependence is a common practice for the identification and estimation of a spatial econometric model. Moran's I test statistic is a simple and popular test for spatial dependence (Moran, 1948, Cliff and Ord, 1973, 1981, Anselin, 1988). In panel data models with time invariant spatial weights matrices, Moran's I test statistic is as followed:

$$I = \frac{e'(I_T \otimes W)e}{e'e} \tag{1}$$

Where W is a $N \times N$ zero diagonal and row-normalized spatial weights matrix, e is the OLS residuals of $Y = X\beta + \varepsilon$, $e = Y - \hat{Y} = MY$, $M = I_{NT} - X(X'X)^{-1}X'$. Under the assumption of i.i.d., Moran's I test statistic is asymptotically normally distributed with the mean $E(I)$ and variance $V(I)$ defined by:

$$E(I) = \frac{trace[M(I_T \otimes W)]}{NT - k}$$

$$V(I) = \frac{trace[M(I_T \otimes W)M(I_T \otimes W')] + trace[M(I_T \otimes W)]^2 + [trace[M(I_T \otimes W)]]^2}{(NT - k)(NT - k + 2)} - [E(I)]^2$$

Where $trace(.)$ is the trace of the matrix or the sum of diagonal elements. N is the sample size, T is the time period, and k is the rank of X or the number of independent variables.

Spatial weights matrix is a very important characteristic element of spatial econometric models and is cause of dispute in relation to what is it and how should it be specified. (Mur et al., 2011). It's exogenous and depends on the researchers' subjective judgment. In general, spatial weights matrices are based on contiguity or distances among regions, economic distances, social distances or demographic characteristics. In recent years, spatial weights matrices have received considerable attention afterward (Anselin, 2002, Anselin et al., 2008). Bhattacharjee and Jensen-Bulter (2006) propose a new methodology for estimation of spatial weights matrices which are consistent with an observed pattern of spatial auto-covariance, and for interpreting these estimated spatial weights to identify the real rather than hypothesized determinants of spatial interaction. Mur et al. propose a selection criterion of spatial weights matrices based on objective information existing in the data, which does not depend on the investigator's subjectivity: it is a measure of conditional entropy (Mur et al., 2011). Qu and Lee (2012) present the specification and estimation of the SAR (spatial autoregressive model) model with an endogenous spatial weights matrix, where endogeneity of spatial weights matrix comes from the correlation between error terms in its entries and the disturbances in the SAR outcome equation. Generally, spatial weights matrices are generally not changed over time. However, when elements of a spatial weights matrix are constructed from economic/ socioeconomic / trade characteristics of regions (or districts) in a panel or dynamic setting, these characteristics might be changing over time (Lee and Yu, 2012). Lee and Yu first investigate the quasi-maximum likelihood estimation of spatial dynamic panel data models where spatial weights matrices can be time varying, and find that QML estimate is consistent and asymptotically normal, however, when spatial weights matrices are substantially varying over time, a model misspecification of a time invariant spatial weights matrix may cause substantial bias in estimation (Lee and Yu, 2012). As we know, so far a spatial weights matrix is time invariant in almost kinds of tests for spatial dependence. Could a model misspecification of a time invariant spatial weights matrix also cause substantial bias in tests for spatial dependence?

In this paper, we develop Moran's I tests, first propose Moran's I tests for spatial dependence in panel data models with time varying spatial weights matrices, and investigate the performance of the Moran's I tests from the view of size and power tests. Based on extensive Monte Carlo simulations of Moran's I test statistic, we find that size of Moran's I tests with time varying and misspecification time invariant spatial weights matrices are similar. However, power of Moran's tests with time varying spatial weights matrices is much higher than one with time invariant spatial weights matrices.

In the next section, we investigate Moran's I tests for spatial dependence in panel data models with time varying spatial weights matrices, including its' form and properties. Section 3 presents the design of Monte Carlo experiments for Moran's I tests. Simulation results are summarized in tables and graphs. The conclusion is given in Section 4.

Moran's I Tests for Spatial Dependence

The spatial autoregressive model (SAR) is a standard tool to analyze data with spatial correlation. The spatial error autoregressive model (SEAR) is also one kind of spatial econometric models. When spatial weights matrices are time invariant, SAR follows that:

$$Y = \lambda(I_T \otimes W)Y + X\beta + \varepsilon \tag{2}$$

When spatial weights matrices in panel data models are changed over time, Eq. (2) becomes:

$$Y = \lambda \widetilde{W}Y + X\beta + \varepsilon \tag{3}$$

Or

$$y_{it} = \lambda \sum_{j=1}^{N} \widetilde{w}_{ij,t} \, y_{jt} + x_{it}\beta + \varepsilon_{it} \tag{4}$$

where the *NTXNT* time varying spatial weights matrix *W* is as followed,

$$\tilde{W} = (\tilde{w}_{ij,t})_{i,j=1,\cdots,N,t=1,\cdots,T} = \begin{bmatrix} \tilde{W}_1 & 0 & \cdots & 0 & 0 \\ 0 & \tilde{W}_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \tilde{W}_{T-1} & 0 \\ 0 & 0 & \cdots & 0 & \tilde{W}_T \end{bmatrix}$$

And

$$\tilde{W}_t = \begin{bmatrix} \tilde{w}_{11,t} & \tilde{w}_{12,t} & \cdots & \tilde{w}_{1N,t} \\ \tilde{w}_{21,t} & \tilde{w}_{22,t} & \cdots & \tilde{w}_{2N,t} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{w}_{N1,t} & \tilde{w}_{N2,t} & \cdots & \tilde{w}_{NN,t} \end{bmatrix}$$

Where $\tilde{W}_t (t = 1,2,\cdots,T)$ is a $N \times N$ row-normalized spatial weight matrix with zero diagonal in the t-th period, and when spatial weights matrices is time invariant, the sub-matrix $\tilde{W}_1 = \tilde{W}_2 = \cdots = \tilde{W}_t$ and $\tilde{W} = I_T \otimes W$

Therefore, when spatial weights matrices are changed over time, Moran's I tests for spatial dependence in panel data models becomes:

$$I = \frac{e'\tilde{W}e}{e'e} \tag{5}$$

Where the residual $e = (I_{NT} - X(X'X)^{-1}X')Y = MY$. In this case, the matrix M is still a real symmetric idempotent matrix, and spatial weights matrix W is still row standardized, which is similar to the cross sectional case (Cliff and Ord, 1973, 1981). Tiefelsdorf (2000) derives approximations of Moran's I's distribution by its moments when the disturbance vector ε is independent and identically distributed (i.i.d) random variables. Furthermore, we can develop asymptotic distribution of Moran's I tests from cross sectional data models and panel data models with invariant spatial weights matrices to panel data models with time varying spatial weights matrices. Moran's I statistic in Eq. (5) is asymptotic normal distributed, and then its' expectation and variance follow that:

$$E(I) = \frac{trace(M\tilde{W})}{NT - k} \tag{6}$$

$$V(I) = \frac{trace(M\tilde{W}M\tilde{W}') + trace(M\tilde{W})^2 + [trace(M\tilde{W})]^2}{(NT - k)(NT - k + 2)} - [E(I)]^2 \tag{7}$$

We can obtain the standardized Moran's I statistic and its' distribution by using the z-transformation of Moran's I:

$$Z = \frac{I - E(I)}{\sqrt{V(I)}} \overset{Asy}{\sim} N(0,1) \tag{8}$$

When spatial weights matrices are substantially varying over time, tests for spatial dependence based on Eq. (1) may be bias, and Eq. (8) should be reasonably used. In the next section, we further investigate

the performance of Moran's I tests for spatial dependence with time varying spatial weights matrices and the bias of Moran's I tests for spatial dependence with the misspecification of invariant spatial weights matrices.

Monte Carlo Simulation

In this section, based on Moran's I tests for spatial dependence in panel data models with time varying spatial weights matrices (Eq. (8)), their size and power computed from Monte Carlo experiments are used to evaluate the performance of the tests.

Given a known probability distribution of a test statistic, its P-value is the probability of the test accepting the null hypothesis. Using P-value for hypothesis testing is simple and more informative for empirical size and power evaluation (Davidson and MacKinnon, 1999, 2006, Lin et al., 2011).

Given the null hypothesis of spatial independence, the alternatives under consideration include a spatial autoregressive model (SAR) and a spatial error autoregressive model (SEAR). For different purposes of the study, the data generating process may assume a structure of the basic null model or any of the alternative models. To study the size of Moran's I tests, the data is generated from the basic linear model assuming no spatial correlation. Two alternative models are used to generate data with spatial dependence with time varying spatial weight matrices in order to study the power of Moran's I tests. Namely, the SAR and SEAR model is respectively defined by: $Y = \lambda WY + X\beta + \varepsilon$, and $Y = X\beta + u, u = \rho Wu + \varepsilon$. X is an NTX3 matrix, generated from independent uniformly distribution in [0, 1]. ε is an NTX1 vector, generated from independent normal distribution N(0,1). When the null hypothesis of spatial independence is true, $Y = X\beta + u$, namely, $\lambda = \rho = 0$. The alternative hypothesis model (SAR) is true, $\lambda = $ -0.9,–0.8,$\cdots$,–0.1,0.1,0.2,$\cdots$,0.9, and another alternative hypothesis model (SEAR) is true, $\rho = $ -0.9,–0.8,$\cdots$,–0.1,0.1,0.2,$\cdots$,0.9. The significant level is 0.05. For time varying spatial weights matrices, we choose an alternating pattern. Any element of the t-th spatial weights matrix W_t is randomly drawn from 0 and 1. All these weights matrices are row-normalized and zero diagonal.

We also compare size and power of Moran's I tests for spatial dependence with time varying spatial weights matrices (TV-Moran) and no time varying spatial weights matrices (NTV-Moran) based on 500 times Monte Carlo experiments.

Size of TV-Moran and NTV-Moran tests for spatial dependence is reported in Table 1. We consider two cases: (1) the sample size (N) is different with fixed T, namely N=25, 36, 49, 64, 81, 100, 121, 144, 169, 196, and T=10, (2) the time period (T) is different with fixed N, namely T=5, 10, 15, 20, 25, 30, 35, 40, 45, 50. Table 1 shows that there is no significant difference between size of TV-Moran and NTV-Moran when N or T are different. Furthermore, when there is no spatial dependence, spatial weights matrix doesn't work, and then it's reasonable that there is no difference for TV-Moran tests and NTV-Moran tests.

Table 1 Size of Moran tests for spatial dependence in panel data models with time varying and invariant spatial weights matrices

N	Size		T	Size	
	TV-Moran	NTV-Moran		TV-Moran	NTV-Moran
25	0.043	0.047	5	0.066	0.042
36	0.06	0.047	10	0.055	0.048
49	0.042	0.034	15	0.044	0.043
64	0.049	0.043	20	0.051	0.042
81	0.047	0.042	25	0.041	0.044
100	0.043	0.039	30	0.057	0.054
121	0.05	0.042	35	0.042	0.048
144	0.049	0.036	40	0.051	0.05
169	0.048	0.052	45	0.048	0.06
196	0.055	0.043	50	0.058	0.046

Fig. 1 presents power of Moran tests for spatial dependence in panel data models with time varying and invariant spatial weights matrices at 5% nominal level of significance. In each of the 1 by 2 grid of graphs, "TV-Moran" and "NTV-Moran" respectively denote the power curves of Moran's I tests for spatial dependence with time varying and the misspecification of invariant spatial weights matrices. The blue solid-line curve represents the power of TV-Moran, and the red dash curve represents the power of NTV-Moran. The left graph is the power curves when the alternatives models - SAR are true, and the right graph is the power curves when the alternative models- SEAR are true. We observe that the blue solid-line curve is higher than the red dash curve, especially $\lambda < 0$ in the SAR model and $\rho < 0$ in the SEAR model, namely the power of TV-Moran is much higher than that of NTV-Moran, especially in the negative spatial correlation. When spatial correlation is more than zero ($\lambda > 0$ in the SAR models, $\rho > 0$ in the SEAR models), the power of TV-Moran is similar to one of NTV-Moran. Obviously, when there is negative spatial correlation, a model misspecification of a time invariant spatial weights matrix may cause substantial reduction in power of tests for spatial dependence.
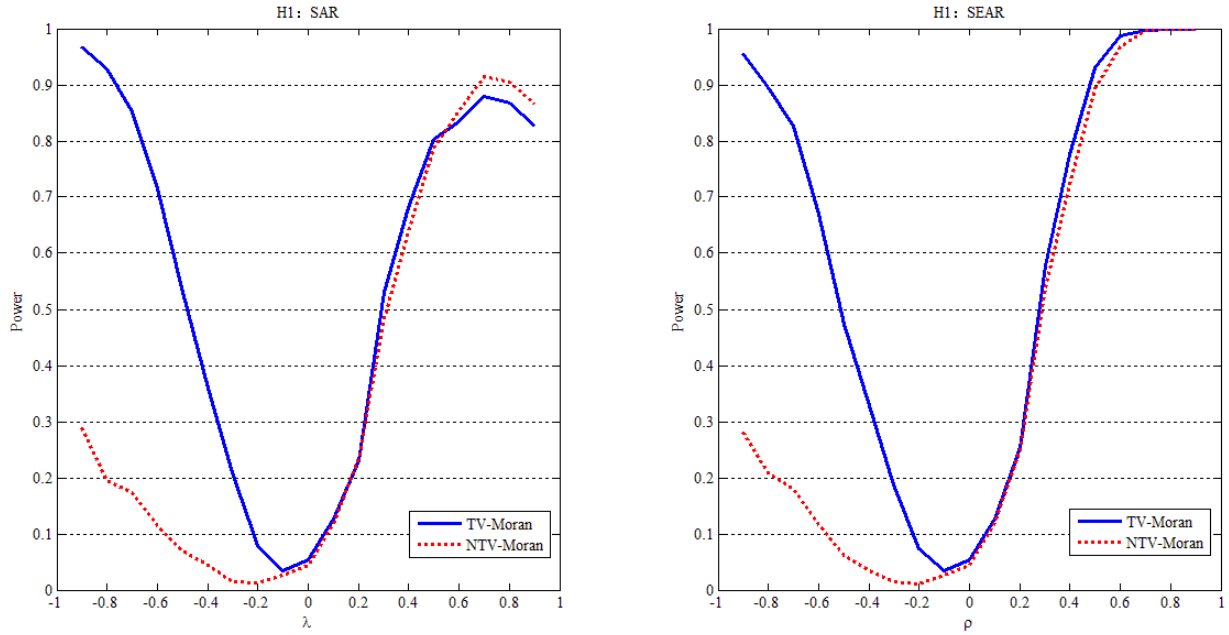
Fig. 1 Power of Moran tests for spatial dependence in panel data models with time varying and invariant spatial weights matrices

In summary, based on our extensive Monte Carlo experiments, TV-Moran tests are proven to have higher power than NTV-Moran tests based on the misspecification of invariant spatial weights matrix, and there is no significant difference between size of TV-Moran and NTV-Moran tests. The good performance of TV-Moran tests is particularly evident for negative spatial correlation. TV-Moran's I tests are the preferred method of testing spatial dependence in panel data models with substantially time varying spatial weights matrices.

Conclusion

In this paper, we first investigate and develop Moran's I tests for spatial dependence in panel data with substantially time varying spatial weights matrices and propose the corresponding TV-Moran tests. It's known that NTV-Moran tests with the misspecification of invariant spatial weights matrix perform poorly with weak power when spatial correlation is negative. Instead, we derive TV-Moran tests based on with time varying spatial weights matrices. Our Monte Carlo experiments confirm that for general models with time varying spatial weights matrices, TV-Moran tests is superior to NTV-Moran test with the misspecification of invariant spatial weights matrix, with larger power.

Acknowledgement

References

[1] A.D. Cliff and J.K. Ord, Spatial Autocorrelation, Pion, London, 1973.

[2] A.D. Cliff and J.K. Ord, Spatial Processes: Models and Applications, Pion, London, 1981.

[3] Baltagi and D. Li, Prediction in the panel data model with spatial correlation: the case of liquor, Spatial Economic Analysis. 1 (2006) 175-185.

[4] Bhattacharjee, A. and C. Jensen-Butler, Estimation of spatial weights matrix in a spatial error model, with an application to diffusion in housing demand, Working paper, 2006.

[5] Brasili, F. Bruno and A. Saguatti, A spatial econometric approach to EU regional disparities between economic and geographical periphery, Statistics. 72 (2012) 299-316.

[6] Frazier and K.M. Kockelman, Spatial econometric models for panel data: incorporating spatial and temporal data, Transportation Research Record: Journal of the Transportation Research Board. 1902 (2005) 80-90.

[7] J. Mur, M. Herrera and M. Ruiz, Selecting the W matrix: parametric vs nonparametric approaches, Working paper, 2011.

[8] K.P. Lin, Z.H. Long and B.L. Ou, The size and power of bootstrap tests for spatial dependence in a linear regression model, Computer Economics. 38(2011) 153-171.

[9] L. Anselin, Spatial externalities, spatial multipliers and spatial econometrics, International Regional Science Review. 26(2003) 153-166.

[10] L. Anselin, Spatial Econometrics: Methods and Models, Dordrecht, Kluwer, 1988.

[11] L. Anselin, Under the hood: issues in the specification and interpretation of spatial regression models, Agricultural Economics. 27 (2002) 247-267.

[12] L. Anselin, J. LeGallo, and H. Jayet, Spatial Panel Econometrics, in: L. Matyas and P. Sevestre(Eds.), The Econometrics of panel data, Springer-Verlag, Berlin, 2008, pp. 625-660.

[13] L.F. Lee and J.H. Yu, QML estimation of spatial dynamic panel data models with time varying spatial weights matrices, Spatial Economic Analysis. 7:1 (2012) 31-74.

[14] M. Tiefelsdorf, Modelling Spatial Processes: The Identification and Analysis of Spatial Relationships in Regression Residuals by Means of Moran's I, Spinger-Verlag, Berlin, 2000.

[15] P. Egger, M. Pfafermayr and H. Winner, An unbalanced spatial panel data approach to US state tax competition, Economics Letters. 88 (2005) 329-335.

[16] P. Moran, The interpretation of statistical maps, Journal of the Royal Statistical Society. B 10 (1948) 243-251.

[17] R. Davidson and J.G. MacKinnon, The size distortion of bootstrap tests, Econometric Theory. 15 (1999) 361-376.

[18] R. Davidson and J.G. MacKinnon, The power of bootstrap and asymptotic tests, Journal of Econometrics. 133(2006) 421-441.

[19] X. Qu and L.F. Lee, Estimating a spatial autoregressive model with an endogenous spatial weight matrix, Working paper, 2012.

[20] V. Druska and W.C. Horrace, Generalized moments estimation for spatial panel data: Indonesian rice farming, American Journal of Agricultural Economics. 86 (2004) 185-198.