Bootstrap LM-lag test for spatial dependence in panel data models

Zhihe Long[1, a] and Bianling Ou[2, b]

[1] School of Business Administration, South China University of Technology, Guangzhou 510641, China

[2]School of Management Science and Engineering, Central University of Finance and Economics, Beijing 102206, China

[a]l-zh720@163.com, [b]oubianling@amss.ac.cn

Keywords: Bootstrap, LM-lag, Panel data models, Edgeworth expansion.

Abstract. This paper applies bootstrap methods to LM-lag test for spatial dependence in panel data models, and LM-lag test is asymptotic pivotal. The consistencies of LM tests and their bootstrap versions are proved, and then the asymptotic refinements of bootstrap LM tests are obtained. It shows that the first order asymptotic distribution of LM-lag test converges as $O((NT)^{-1})$, and the convergence rate of bootstrap LM-lag test is $O((NT)^{-2})$. The error made by the bootstrap LM-lag test approximation to the asymptotic distribution is $O((NT)^{-3/2})$. Compared to asymptotic distribution of LM-lag test, bootstrap LM-lag test gets asymptotic refinements.

1. Introduction

The bootstrap is a method for estimating the distribution of an estimator or test statistic by resampling one's data or a model estimated from the data (Horowitz, 2001). So far, bootstrap is widely applied in cross sectional data, time series and panel data. Many scholars investigate the performance of bootstrap methods in terms of simulation analysis, and few provide the corresponding mathematical proof.

In the classical econometrics, the performance of bootstrap methods is derived based on Edgeworth expansion by Beran (1988), Hall (1992), Horowitz (2001), Hall and Horowitz (1996), Cameron and Trivedi (2005), Godfrey (2009). Recently, bootstrap is extensively focused on its applications in spatial econometrics. Lin et al. (2011) investigate the performance of bootstrap Moran's I test for spatial dependence in cross sectional data, and find that bootstrap Moran's test for spatial dependence is superior to asymptotic Moran's I test in terms of mathematical derivation and Monte Carlo experiments.

Tests for spatial dependence are one of the important essentials in spatial econometric analysis. Both Moran's I and LM tests are usually applied for spatial dependence in cross sectional data and panel data. Many Monte Carlo experiments show that LM tests for spatial dependence have good finite sample properties when the disturbances are normal, however when the disturbances are not normal or with heteroskedasticity, LM tests can have large size distortion and lower power (Anselin et al. (1996), Yang (2015)). So far there are three important and closely associated studies in mathematically derivation. Long et al. (2012) use Edgeworth expansion to prove the validity of bootstrap Moran's I tests for spatial dependence in cross sectional data. Jin and Lee (2015) prove the existence of the Edgeworth expansions for LQ forms with the normal disturbances, and establish an asymptotic expansion of LQ forms based on martingales with no normal disturbances. Yang (2015) find that in finite samples, the LM test referring to asymptotic critical values may suffer from the problems of size distortion and low power, which become worse with a denser spatial weight matrix, and introduces residual-based bootstrap methods to refine asymptotically approximations to the finite sample critical values of the LM statistics. In a word, Edgeworth expansion proposed by Hall (1992) is still the most feasible method that mathematically derives the validity of bootstrap methods. However, all above studies are focused on tests for spatial dependence in cross sectional data. This paper will firstly extend bootstrap LM tests for spatial dependence from cross sectional

data to panel data, and discuss the performance of bootstrap LM tests for spatial dependence in panel data models with fixed effects. LM tests in this paper include LM-lag test for spatially lagged dependent variable, and LM-error test for spatial residual autocorrelation.

The rest of the paper is organized as follows: Section 2 introduces LM-lag test. Section 3 first provides assumptions, then shows that the bootstrap is consistent for LM tests, and finally proposes the asymptotic refinement of bootstrap LM tests. Section 4 concludes.

2. LM-lag test for spatial dependence

LM tests for spatial dependence in panel data models mainly include LM-lag test, LM-error test. LM-lag and LM-error tests can respectively identify spatial correlation in the error and spatial dependence in the dependent variable. LM-lag test in panel data models is as follows (Elhorst, 2010; Anselin, et al., 2008):

$$LM - lag = \frac{[e'(I_T \otimes W)Y / \sigma^2]^2}{J} \tag{1}$$

Where $\otimes$ denotes Kronecker product, $T_W = trace[(W + W\,')W]$ is the trace of the matrix (W+W')W, $J = ((I_T \otimes W)X\hat{\beta})'(I_{NT} - X(X\,'X)^{-1}X\,')((I_T \otimes W)X\hat{\beta}) / \sigma^2 + TT_W$, $\sigma^2 = e'e / NT$ denotes the variance of the residuals, W is a row-standardized exogenous spatial weights matrix, T is the number of time periods. LM-lag test asymptotically follows the $\chi^2(1)$ distribution. To conveniently study the performance of bootstrap LM-lag test, in view of related jobs (Burridge, 1980; Yang, 2015), the transformation of equation (1) is the following:

$$LM - lag = \frac{NT}{\sqrt{J}} \frac{e'(I_T \otimes W)Y}{e'e} \tag{2}$$

LM-lag test in equation (2) is asymptotically distributed as $N(0,1)$.

3. Asymptotic refinements of Bootstrap LM-lag test

In the literature, Anselin (1988), Anselin et al. (1996, 2008) and Elhorst et al. (2010) investigate LM tests for spatial dependence. Beran (1988), Hall (1992) and Horowitz (1994, 2001) discuss and study the performance of bootstrap methods. Inside, Long et al. (2012), Jin and Yu (2015) and Yang (2015) consider bootstrap methods in cross sectional data. In this section, we make use of bootstrap methods to establish LM-lag test for spatial dependence based on OLS residuals in panel data models, and derive the validity of bootstrap LM-lag test with the standard normal disturbance in finite sample.

3.1 Assumptions

In order to prove the validity of bootstrap LM-lag test in panel data models, we assume the following assumptions:

Assumption 1 W is a non-negative, non-stochastic and row-normalized spatial weights matrix and fixed through time. Moreover, w_{ii}=0 for $i = 1, \cdots, N$. Finally, the matrix is assumed to be uniformly bounded in row and column sums in absolute value.

Assumption 2 The $u_{it}(i = 1, \cdots, N; t = 1, \cdots, T)$ errors are identically and independently distributed across i and t with zero mean and variance. Moreover $E \mid u_{it} \mid^{4+\eta} < \infty$ for some $\eta > 0$.

Assumption 3 The regressor matrix X has full column rank, and its elements are non-stochastic and uniformly bounded in N and T.

Assumption 4 $S_N(\lambda) = I_N - \lambda W$, $R_N(\lambda) = I_N - \rho W$ and $Q_N(\theta) = I_N + \theta W$ are invertible for all $\lambda \in \Lambda$, $\rho \in P$ and $\theta \in H$, where Λ, P and H are compact intervals.

Assumption 5 $\lim_{N \to \infty} \frac{1}{NT} \sum_{t=1}^{T} X_t 'X_t$ and $\lim_{N \to \infty} \frac{1}{NT} \sum_{t=1}^{T} X_t 'R_N '(\rho)R_N(\rho)X_t$ exists and are nonsingular.

Assumption 6 The disturbance u_{it} and the independent variable X_{it} are not correlation.

Assumption 7 *N* is large, where *T* can be finite or large.

3.2 Consistency

Let $\Phi(x)$ be the cumulative distribution function (CDF) for a standard normal random variable x, $F_{NT}(x)$ be the empirical distribution function (EDF) of the residuals. According to Glivenko-Cantelli theorem, as $NT \to \infty$, $F_{NT}(x)$ is almost surely convergence to $\Phi(x)$ owing to $\Phi(x)$ be a continuous function. To define the sample cumulative distribution functions of LM-lag test is as follows, $G_{NT}(x) = P(LM_{NT} \le x)$, where LM_{NT} denotes LM-lag test in equation (2).

Theorem 1: Under assumptions 1-7,

$$\sup_{x \in R} |G_{NT}(x) - \Phi(x)| = r_{NT}$$

$$\sup_{x \in R} |G_{NT}^*(x) - \Phi(x)| = r_{NT}^*$$

Where $G_{NT}^*(x) = P(LM_{NT}^* \le x)$ is the probability distribution function of bootstrap LM-lag test , $r_{NT} = O((NT)^{-1/2})$. The population moments of r_{NT} is replaced by the corresponding sample moment, and obtains r_{NT}^*.

The left of the above first equation is the distance of Kolmogorov-Smirnov of two CDFs, and the right is the convergence rate of the CDF of LM-lag test in finite sample. Its bootstrap version is given in the latter two. The derivation of theorem 1 is similar to Jin and Yu (2015) (Page 298), and then no repeat here. According to Horowitz (2001), Jin and Yu (2015), Yang (2015) and Hall (1992), if all the above four equation are established, the CDFs of bootstrap LM-lag test uniformly converges to that of the corresponding LM-lag test.

According to Horowitz (2001, Page 3166), bootstrap LM-lag test is defined by $\exists \varepsilon > 0$, if $\lim_{NT \to \infty} P[\sup_x |G_{NT}^*(x) - \Phi(x)| > \varepsilon] = 0$, bootstrap LM-lag test is consistent.

Proposition 1: Under assumption 1-7, LM-lag test in equation (2) is asymptotically zero mean and 1 variance. Then the CDF $G_{NT}^*(x)$ of bootstrap LM-lag test is a uniformly consistent estimator of its sample cumulative distribution function $G_{NT}(x)$. Bootstrap LM-lag test is consistency.

Mammen (1992) and Horowitz (2001, Page 3167) indicate that given $\{X_i : i = 1, \cdots, N\}$ be a random sample from a population, for a sequence of functions g_n and sequences of numbers t_n and σ_n, define $\bar{g}_n = n^{-1} \sum_{i=1}^{n} g_n(X_i)$ and $T_n = (\bar{g}_n - t_n)/\sigma_n$. For the bootstrap sample $\{X_i^* : i = 1, \cdots, N\}$, define $\bar{g}_n^* = n^{-1} \sum_{i=1}^{n} g_n(X_i^*)$ and $T_n = (\bar{g}_n - t_n)/\sigma_n$. Let $G_n(\tau) = P(T_n \le \tau)$ and $G_n^*(\tau) = P^*(T_n^* \le \tau)$, where P^* is the probability distribution induced by bootstrap sampling. Then $G_n^*(\cdot)$ consistently estimates G_n if and only if $O((NT)^{-2})$.

In this paper, as *N* tends to large, LM-lag test for spatial dependence in panel data models is asymptotically distributed as *N*(0, 1). Therefore the sufficient and necessary conditions are satisfied, which the CDF $G_{NT}^*(x)$ of bootstrap LM-lag test consistently converges.

3.3 Asymptotic refinements

As we all known, when the statistic is pivotal or asymptotic pivotal, the bootstrap is often more accurate in finite samples than first order asymptotic approximations, that is asymptotic refinements (Horowitz, 2001). LM-lag test is asymptotically distributed as the standard normal distribution, which does not depend on any unknown parameter, and then it is asymptotic pivotal.

Theorem 2: Under assumptions 1-7 and Cramer conditions which $\lim_{|t| \to \infty} \sup |\chi_1(t)| < 1$ is satisfied, the convergence rate of the asymptotic distribution of LM-lag test is $O((NT)^{-1})$, and that of

bootstrap LM-lag test is $O((NT)^{-2})$. Then bootstrap LM-lag test asymptotically obtains refinements, where $\chi_1(t)$ is the characteristic function of LM-lag test.

Due to the limitation of length, to repeat is no more. Because LM-lag test is asymptotic pivotal, the convergence rate of bootstrap LM-lag test is, and its first order asymptotic distribution converges as $O((NT)^{-1})$. The error made by the bootstrap LM-lag test approximation to the asymptotic distribution is $O((NT)^{-3/2})$. Compared to asymptotic LM-lag test, bootstrap LM-lag test gets asymptotic refinements.

4. Conclusion

In this paper, we apply bootstrap methods to LM-lag test for spatial dependence in panel data models. The performance of bootstrap LM-lag test in spatial panel data models is investigated. We establish that the CDFs of bootstrap LM-lag test uniformly converge to that of LM-lag test. It is found that the CDF of bootstrap LM-lag test is a uniformly consistent estimator of its sample cumulative distribution function. In other words, bootstrap LM-lag test is consistent. Furthermore, based on Edgeworth expansion methods, we obtain that the convergence rate of the asymptotic distribution of LM-lag test is $O((NT)^{-1})$, and that of bootstrap LM-lag test is $O((NT)^{-2})$. Obviously, bootstrap LM-lag test obtains asymptotic refinements.

References

[1] L. Anselin, et al., Simple diagnostic tests for spatial dependence, Regional Science Urban Economics. 26(1996) 77-104.
[2] L. Anselin, et al., Spatial panel econometrics. In: Matyas L., Sevestre P.(eds.) The Econometrics of panel data, Springer-Verlag, 2008.
[3] R. Beran, Prepivoting test statistics: a bootstrap view of asymptotic refinements, Journal of the American Statistical Association. 83(1988) 687-697.
[4] A.C. Cameron, P.K. Trivedi, Microeconometrics Methods and Applications, Cambridge University Press, 2015.
[5] J.P. Elhorst, Spatial panel data models. In Fischer M.M., Getis A. (eds.) Handbook of Applied Spatial Analysis, Springer: Berlin Heidelberg New York, 2010.
[6] L. Godfrey, Bootstrap Tests for Regression Models. Palgrave Macmillan, 2009.
[7] P. Hall, The Bootstrap and Edgeworth Expansion. Springer-verlag, 1992.
[8] P. Hall, J.L. Horowitz, Bootstrap critical values for tests based on generalized methods of moments estimators, Econometrica. 64(1996), 891-916.
[9] J.L. Horowitz, Bootstrap-based critical values for the information matrix test, Journal of Econometrics. 61(1994) 395-411.
[10] Horowitz J.L., 2001. The bootstrap. In: Heckman J.J., and Leamer E. (Eds.), Handbook of Econometrics, Elsevier Science B.V.
[11] F. Jin, L.F. Lee, On the bootstrap for Moran's I test for spatial dependence, Journal of Econometrics. 184(2015), 295-314.
[12] K.P. Lin, Z.H. Long, B.L. Ou, The size and power of bootstrap tests for spatial dependence in a linear regression model, Computational Economics. 38(2011) 153-171.
[13] Z.H. Long, B.L. Ou, K.P. Lin, Tests for spatial dependence applying bootstrap methods: mathematics derivation and simulation analysis. Beijing: Science Press (in Chinese), 2012.
[14] Z.L. Yang, LM tests of spatial dependence based on bootstrap critical values, Journal of Econometrics. 185(2015) 33-59.

Acknowledgments