Lecture 5: Asymptotic Properties Spatial Maximum Likelihood Estimator

Mauricio Sarrias

Universidad Católica del Norte

October 18, 2017

Reading

- Lee, L. F. (2004). Asymptotic Distributions of Quasi-Maximum Likelihood Estimators for Spatial Autoregressive Models. Econometrica, 72(6), 1899-1925.

Asymptotics in a spatial context

What is the meaning of $n \to \infty$ in a spatial context?

Asymptotics in a spatial context

- Two approaches: Increasing domain and infill asymptotics.

Definition (Increasing domain)

It consists of a sampling structure where new observations (spatial units) are added at the edges (boundary points). That is, it refers to more and more observations being sampled over an increasing domain.

Definition (Infill asymptotics)

Spatial domain is bounded, and new observations (points) are added in between existing ones, generating a denser surface.

Classical approach: Increasing domain.

Think about the following: What would happen with $\mathbf{W}$ if we adopt the increasing domain approach?

- $\mathbf{W}$ will change as $n \to \infty$. New spatial units will change the structure for the existing spatial units.
- Since the "equilibrium vector" is:

$$\mathbf{y}_n = \mathbf{A}_n^{-1}(\mathbf{X}_n\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}_n) \tag{1}$$

where $\mathbf{A}_n = \mathbf{I}_n - \rho_0\mathbf{W}_n$ is nonsingular; then $\mathbf{y}$ will change as $n \to \infty$.

- Since $\boldsymbol{\varepsilon}_n$ depends on $\mathbf{W}$, it will also change as $n \to \infty$.

So, $\mathbf{y}$ will change as $n \to \infty$.

Implication: the outcome for the first spatial unit, y_1, will be different if we consider a total $n = 10$ or $n = 15$ observations because of the changing nature of $\mathbf{W}$ as n changes and given the DGP in Equation (1). This implies that these elements and the vector $\mathbf{y}$ **should be indexed by** n:

$$\mathbf{y}_n = (y_{11}, y_{21}, y_{22}, ..., y_{nn})$$

For example, for $n = 1, 2, 3$, then (by row):

$$\begin{aligned} n = 1 &\implies y_{11} \\ n = 2 &\implies y_{12}\ y_{22} \\ n = 3 &\implies y_{13}\ y_{23}\ y_{33} \\ &\vdots \\ n = n &\implies y_{13}\ y_{23}\ y_{33} \cdots y_{3n} \end{aligned}$$

where $y_{11} \neq y_{12} \neq y_{13}$ and $y_{22} \neq y_{23}$. **This structure is called a Triangular Array**.

The **increasing domain** framework requires the knowledge of **triangular arrays**. The following definition give us a simply definition of Triangular Arrays.

Definition (Triangular Array of Random Variables)

The ordered collection of random variables

$$\{X_{11}, X_{21}, X_{22}, X_{31}, X_{32}, X_{33}, ..., X_{nn}, ...\},$$

or

$$\begin{pmatrix} X_{11} & & & & & & \\ X_{21} & X_{22} & & & & & \\ X_{31} & X_{32} & X_{33} & & & & \\ \vdots & \vdots & \vdots & \ddots & & & \\ X_{n1} & X_{n2} & X_{n3} & X_{n4} & \dots & X_{nn} & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is called a triangular array of random variables, and will be denoted by $\{X_{nn}\}$.

- We need new LLN and CLT for triangular arrays.
- What is the difference? We are now concerned with the asymptotic behaviour of:

$$S_n = n^{-1} \sum_{i=1}^{n} X_{ni}$$

For $n = 3$, we have:

$$S_3 = (1/3)(X_{31} + X_{32} + X_{33})$$

The traditional CLTs deal with functions of averages of the type:

$$n^{-1} \sum_{i=1}^{n} X_i$$

the X_i's being elements of the sequence $\{X_n\}$.
- However, the triangular array $\{X_{nn}\}$ is more general than a sequence $\{X_n\}$ in the sense that the random variables in a row of the array need not be the same as random variables in other rows.

- Both the LLN and CLT require **slightly stronger conditions** than the LLN and CLT for iid sequence of random variables.
- What are the conditions on the random variables so that a properly S_n converges to a normal distribution as $n \to \infty$. In a nutshell, assume:
  - Independence: assume all random variables in the array are independent.
  - Centering: assume $\mathbb{E}(X_{j,i}) = 0$ for all j, i.
  - Variances converge: assume $\sum_{i=1}^{n} \mathbb{E}(X_{n,i}^2) \to \sigma^2 > 0$ as $n \to \infty$
  - No single variance is too large.

  Then $S_n \xrightarrow{d} N(0, \sigma^2)$ as $n \to \infty$.
- Note that the dependent variable in the same row are mutually independent (spatial units are independent) and have the same distribution. But the distribution of the random variable y (and ϵ) in different rows are allowed to be different.

QMLE

What is the QMLE?

QMLE

- **Quasi-maximum likelihood Estimate (QMLE):** Estimate of a parameter that is formed by maximizing a function that is related to the logarithm of the likelihood function, but is not equal to it.
- The ML method assumes that the specified density function is the true density function.
- QMLE is more conservative when the actual error distribution does not match the assumed error distribution.

Assumption 1: Errors

Assume the following

1. The disturbances $\{\epsilon_{i,n} : 1 \leq i \leq n, n \geq 1\}$ are identically distributed. Furthermore, for each sample size n, they are jointly independent distributed with mean $\mathbb{E}(\epsilon_{i,n}) = 0$ and $\mathbb{E}(\epsilon_{i,n}^2) = \sigma_{\epsilon,n}^2$ where $0 < \sigma_{\epsilon,n}^2 < b$.
2. Its moments $\mathbb{E}\left(|\epsilon_{i,n}|^{4+\gamma}\right)$ for some $\gamma > 0$ exits.

- Error forms **triangular array**: they depend on n.
- Statistics will involve quadratic forms of ϵ_n, so we need the fourth moment of the error term to exist!
- Note that we are not assuming normality, though we are using ML approach (QML).

- $\mathbf{W}_n$ forms a triangular array of constants: $\mathbf{W}_n$ is assumed to be fixed (non stochastic).
  - the element w_{ij} is not the same when $n = 50$ or $n = 55$.
- Are the elements $w_{n,ij}$ bounded (limited) as $n \to \infty$?
- Recall that in econometric we are often interested in the asymptotic behaviour of sequences. For example we say that

$$X_n = O(b_n) \implies \lim_{n\to\infty} \frac{X_n}{b_n} = -\infty < c < \infty.$$

This implies that X_n is a bounded sequence of rate b_n.

Assumption 2: Weight Matrix

The elements $w_{n,ij}$ of $\mathbf{W}_n$ are at most of order h_n^{-1}, denoted by $O(1/h_n)$, uniformly in all i, j, where the rate sequence h_n can be **bounded** or **divergent**. As a normalization, $w_{n,ii} = 0$ for all i.

- The elements of $\mathbf{W}_n$ are sequences that might be bounded or divergent at rate h_n.
- That is, we do not know if $h_n w_{n,ij}$ is bounded or divergent.

Assumptions: What about W?

Assumption 3: Weight Matrix

The ratio $h_n/n \to 0$ as n goes to infinity

- Intuition: as $n \to \infty$, the row sum of $\mathbf{W}$ will also tend to increase (more neighbors!).
- The rate at which $w_{n,ij}$ increases as n increase can be:
  - bounded: limit on the number of neighbors.
  - divergent: not limit the number of neighbors.
- However, Asumption 3 excludes cases where the row sums diverges to infinity at a rate equal or faster of the sample size n.
- Therefore, assumptions 2 and 3 are intended to cover weight matrices whose elments are not restricted to be nonnegative and those that might not be row-standardized.
- They limit the cross-sectional correlation to a manageable degre:
  - the correlation between two spatial units should converge to zero as the distance separating them increases to ∞

Assumptions: Applied Implications

- k-nearest neighbors: Lee's assumption is satisfied since the number of neighbors is limited: $\{h_n\}$ is a bounded sequence.
- Distance: Consider this setting:

$$\mathrm{R1} \xleftarrow[2d]{\quad} \mathrm{R2} \xleftarrow[d]{\quad} \mathrm{R3} \xrightarrow[d]{\quad} \mathrm{R4} \xrightarrow[2d]{\quad} \mathrm{R5}$$

  When $\mathbf{W}$ is an inverse distance matrix and its off-diagonal elements are of the form $1/d_{ij}$, where d_{ij} is the distance between two spatial units i and j, each row sum is

$$1/d + 1/d + 1/2d + 1/2d + ... = 2 \times (1/d + 1/2d + 1/3d + ...)$$

  representing a series that is **not finite**.
- This is perhaps the main motivation of why some empirical applications introduce a cut-off point d^* such that $w_{ij} = 0$ if $d_{ij} > d^*$. However, since the ratio $2 \times (1/d + 1/2d + 1/3d + ...)/n \to 0$ as $n \to \infty$, Lee (2004)'s condition is satisfied, which implies that an inverse distance matrix without cut-off point does not necessarily have to be excluded in an empirical study for reasons of consistency.
- For this last case, Lee (2004)'s assumptions are met. However Kelejian and Prucha (1998, 1999) are not.

Assumptions: Applied Implications

- What if h_n is unbounded? Under this case $\sum_{j=1}^{n} d_{ij}$ is uniformly bounded away from zero at the rate h_n, where $\lim_{n\to\infty} h_n = \infty$
- In which cases $h_n \to \infty$? This case requires that each unit in the limit has infinitely many neighbors. As stated by Lee (2002), in economic applications where either the neighbors of any unit are dense in a relevant space or each unit is influenced by many of its neighboring units, which represents a significant proportion of the total population units, it is likely that $\sum_{j=1}^{n} d_{ij}$ will diverge and $(1/n)\sum_{j=1}^{n} d_{ij}$ will converge as n becomes large.

Assumptions: Applied Implications

In summary

- When $\{h_n\}$ is a bounded sequence, it implies a cross sectional unit has only a small number of neighbors, where the spatial dependence is usually defined based on geographical implications.
- When $\{h_n\}$ is divergent, it corresponds to the scenario where each unit has a large number of neighbors that often emerges in empirical studies of social interactions or cluster sampling data.

Assumptions: Applied Implications

Remark:
Whether $\{h_n\}$ is a bounded or divergent sequence has interesting implications on the OLS estimation. The OLS estimators of $\boldsymbol{\beta}$ and ρ are inconsistent when $\{h_n\}$ is bounded, but they can be consistent when $\{h_n\}$ is divergent (Lee, 2002).

Assumptions

Nonsingularity

The matrix $\mathbf{A}_n$ is nonsingular.

Under this assumption, the SLM model (system) has the reduced form (equilibrium) given by Equation (1), and:

$$\mathbb{E}(\mathbf{y}_n) = (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1} \mathbf{X}_n \boldsymbol{\beta} = \mathbf{A}_n^{-1} \mathbf{X}_n \boldsymbol{\beta}_0 \tag{2}$$

$$\operatorname{Var}(\mathbf{y}_n) = \sigma_0^2 (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1} (\mathbf{I}_n - \rho_0 \mathbf{W}_n)^{-1\top} = \sigma_0^2 \mathbf{A}_n^{-1} \mathbf{A}_n^{-1\top} \tag{3}$$

Bounded Matrices

Before explaining the rest of assumption, we need the notion of bounded matrices.

Definition (Bounded Matrices)

Let $\{\mathbf{A}_n\}$ be a sequence of n-dimensional square matrices, where $\mathbf{A}_n = [a_{n,ij}]$,

1. The column sums of $\{\mathbf{A}_n\}$ are uniformly bounded (in absolute value) if there exists a finite constant c that does not depend on n such that

$$\|\mathbf{A}_n\|_\infty = \max_{1 \leq j \leq n} \sum_{i=1}^{n} |a_{n,ij}| \leq c$$

2. The row sums of $\{\mathbf{A}_n\}$ are uniformly bounded (in absolute value) if there exists a finite constant c that does not depend on n such that

$$\|\mathbf{A}_n\|_1 = \max_{1 \leq i \leq n} \sum_{j=1}^{n} |a_{n,ij}| \leq c$$

Assumptions

Assumption 4

The sequences of matrices $\{\mathbf{W}_n\}$ and $\left\{\mathbf{A}_n^{-1}\right\}$ are uniformly bounded in both row and column sums.

- The uniform boundedness of the matrices is a condition to limit the spatial correlation to a manageable degree. For example, it guarantees that the variances of $\mathbf{y}_n$ are bounded as n goes to infinity.

Assumption 5:

The elements of $\mathbf{X}_n$ are uniformly bounded constants for all n. The $\lim_{n\to\infty} \mathbf{X}_n^\top \mathbf{X}_n / n$ exists and is nonsingular.

This rules out multicollinearity among the regressors. Note also that we are assuming that $\mathbf{X}_n$ is **nonstochastic**. If $\mathbf{X}_n$ were stochastic, then we will require:

$$\text{plim}_{n\to\infty} \mathbf{X}_n^\top \mathbf{X}_n / n,$$

to exists.

Assumptions

Assumption 6:

$\mathbf{A}_n^{-1}(\rho)$ are uniformly bounded in either row or column sums, uniformly in ρ in a compact parameter space $\mathbf{P}$. The true parameter ρ_0 is in the interior of $\mathbf{P}$

- $-1/\omega_{min} < \rho < 1/\omega_{max}$, where ω_{min} and ω_{max} are the minimum and maximum eigenvalues of $\mathbf{W}$, and $\mathbf{P}$ will be a closed interval contained in $(-1/\omega_{min}, 1/\omega_{max})$ for all n

Assumptions

Assumption 7:

The

$$\lim_{n\to\infty}\frac{1}{n}\left(\mathbf{X}_n,\mathbf{C}_n\mathbf{X}_n\boldsymbol{\beta}_0\right)'\left(\mathbf{X}_n,\mathbf{C}_n\mathbf{X}_n\boldsymbol{\beta}_0\right)$$

exists and is nonsingular.

This is a sufficient condition for global identification of $\boldsymbol{\theta}_0$

Consistency

Theorem (Consistency)

Under assumption 1-7, $\boldsymbol{\theta}_0$ is globally identifiable and $\widehat{\boldsymbol{\theta}}_n$ is a consistent estimator of $\boldsymbol{\theta}_0$.

The proof is given in Lee (2004). Identification of ρ_0 can be based on the maximum values of the concentrated log-likelihood function $Q_n(\rho)/n$. With identification and uniform convergence of $\left[\log L_n(\rho) - Q_n(\rho)\right]/n$ to zero on $\mathbf{P}$, consistency of the QMLE $\widehat{\boldsymbol{\theta}}_n$ follows.

Asymptotic Normality

Theorem (Asymptotic Normality)

Under Assumptions 1-7,

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathrm{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1} + \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\boldsymbol{\Omega}_{\boldsymbol{\theta}}\boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right), \tag{4}$$

where $\boldsymbol{\Omega}_{\boldsymbol{\theta}} = \lim_{n\to\infty} \boldsymbol{\Omega}_{\boldsymbol{\theta},n}$ *and*

$$\boldsymbol{\Sigma}_{\boldsymbol{\theta}} = -\lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n}\frac{\partial^2 \log L_n(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}\partial\boldsymbol{\theta}^\top}\right], \tag{5}$$

which are assumed to exists. If the ϵ_i*'s are* ***normally distributed****, then:*

$$\sqrt{n}\left(\widehat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0\right) \xrightarrow{d} \mathrm{N}\left(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\boldsymbol{\theta}}^{-1}\right). \tag{6}$$

Proof on blackboard!