
Regressão Não Bayesiana

Renato Martins Assunção

Departamento de Estatística e Laboratório de Estatística Espacial
Universidade Federal de Minas Gerais - UFMG

Sumário

1	Regressão não bayesiana	1
1.1	SAR model	1
1.2	O modelo autoregressivo condicional (CAR)	3
1.2.1	Equivalence between SAR and CAR models	4
1.3	Old Stuff - from SBE course	5
1.4	Maximum likelihood estimation	11
1.5	Confidence intervals and tests	18
1.6	Difficulties with non-normal data	21

Capítulo 1

Regressão não bayesiana

Although "a large number of model specifications for spatial processes have been suggested in the literature(...), a fundamental distinction, with a significant impact on estimation and testing strategies, is the difference between a simultaneous and a conditional spatial process.(...) Both models are most often presented in a autoregressive form, i.e., where the value of a variable at one point is related to its values in the rest of the spatial system"(Anselin, 1988, pp 32-33).

The **simultaneous autoregressive (SAR) model** dates back to the work of Whittle (1954).

1.1 SAR model

Vamos considerar uma grade finita composta de n sítios $D = \{(x_i, y_i) : i = 1, \dots, n\}$ com observações $y = (z_1, \dots, y_n)$. O modelo mais comum é aquele em que as observações possuem distribuição normal. A idéia é descrever a observação y_i realizada no sítio i decompondo-a em partes:

$$(y_i = \mu_i) + \sum_{j \neq i} S_{ij}(y_j - \mu_j) + \epsilon_i, \quad (1.1.1)$$

onde $\epsilon_1, \dots, \epsilon_n$ são independentes com distribuição $N(0, \lambda_i)$, $\lambda_i > 0$, $i = 1, \dots, n$. Isto é, a observação y_i é a sua média μ_i mais seu desvio em relação a esta média. Por sua vez, este desvio é uma soma ponderada dos desvios

dos outros sítios e um choque não associado com explicação conhecida. Os pesos S_{ij} refletem a estrutura espacial dos sítios de modo que aqueles sítios j mais próximos de i tendem a ter maior peso. Note que o somatório 1.1.1 não envolve um termo S_{ii} .

A expressão 1.1.1 define y_i em função dos outros elementos do vetor y . Entretanto, esses outros elementos $j \neq i$ terão sua definição envolvendo y_i . Assim, não existe uma sequência recursiva que gere os y_i mas, pelo contrário, eles são determinados simultaneamente, dando origem ao nome do modelo. Esta característica do modelo SAR fica clara se visualizarmos 1.1.1 como um sistema de equações com $S_{ii} = 0, i = 1, \dots, n$. Isto é, se denotarmos $y_i - \mu_i$ por z_i , então:

$$\begin{aligned} w_1 &= S_{11}z_1 + S_{12}z_2 + S_{13}z_3 + \cdots + S_{1n}z_n + \epsilon_1 \\ z_2 &= S_{21}z_1 + S_{22}z_2 + S_{23}z_3 + \cdots + S_{2n}z_n + \epsilon_2 \\ z_3 &= S_{31}z_1 + S_{32}z_2 + S_{33}z_3 + \cdots + S_{3n}z_n + \epsilon_3 \\ &\dots && \dots \\ z_n &= S_{n1}z_1 + S_{n2}z_2 + S_{n3}z_3 + \cdots + S_{nn}z_n + \epsilon_n \end{aligned}$$

Não é óbvio que o sistema de equações acima possua solução, isto é, que exista uma distribuição conjunta para o vetor z tal que cada coordenada possa ser expressa por 1.1.1. Assim, somos levados à questão: em que condições 1.1.1 define uma distribuição de probabilidade para o vetor z ? Caso exista uma distribuição satisfazendo todas as n equações, esta distribuição é única? Para responder a estas perguntas, vamos reescrever o sistema de equações simultâneas em notação matricial.

Em notação matricial, o modelo SAR é expresso da seguinte forma:

$$Z - \mu = S(Z - \mu) + \epsilon \quad (1.1.2)$$

onde a matriz S é $n \times n$. Note que S não precisa ser simétrica mas $S_{ii} = 0, i = 1, \dots, n$. A partir de 1.1.2, obtemos

$$(I - S)(Z - \mu) = \epsilon$$

donde, se $(I - S)$ é invertível,

$$Z = \mu + (I - S)^{-1} \epsilon \quad (1.1.3)$$

concluindo-se que

$$Z \sim N(\mu, (I - S)^{-1} \Lambda (I - S^t)^{-1}) \quad (1.1.4)$$

Assim, temos o seguinte resultado: **Resultado** Suponha que **Exemplo 9** sítios na forma de grade 3x3 com simetria do toro e movimento do peão. Então, se

$$B = c \times A = 1/5 \times \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}$$

temos que

$$\text{diag}((I-B)^{-1} (I-B^t)^{-1}) = \text{diag}((I-B)^{-2}) = (4.2, 3.8, 3.8, 3.7, 3.7, 3.7, 3.1, 3.6, 3.5).$$

Este vetor mostra as variâncias dos y_i 's e é surpreendente que, mesmo com esse grau extremo de simetria, elas não sejam iguais. Note que, se $c \geq 1/4$ na definição de B acima, a matriz $I - B$ é singular já que então $(I - B) \mathbf{x} = \mathbf{0}$.

Aumentando para 25 sítios na forma de grade 5x5 temos as variâncias indo de 2.1 a 2.7.

1.2 O modelo autoregressivo condicional (CAR)

$$y_i | \{y_j, j \neq i\} = \sum_j B_{ij} y_j + \eta_i \quad (1.2.5)$$

onde $\eta_i \sim N(0, \tau_i^2)$.

1.2.1 Equivalence between SAR and CAR models

Os modelos CAR e SAR são equivalentes no seguinte sentido: dado um vetor \mathbf{Z} com distribuição de probabilidade seguindo o modelo SAR, existe um modelo CAR correspondente que possui a mesma distribuição. O contrário também é válido.

Resultado: Os modelos SAR e CAR são equivalentes.

Prova: Vamos provar inicialmente que todo modelo SAR implica num modelo CAR. Suponha que \mathbf{B} e $\mathbf{\Lambda}$ são tais que existe $(\mathbf{I} - \mathbf{B})^{-1}$ com $B_{ii} = 0$ e $\lambda_i > 0, i = 1, \dots, n$. Queremos encontrar \mathbf{C} e \mathbf{M} tais que existe $(\mathbf{I} - \mathbf{C})^{-1}$ com $[M]_{ii} > 0, C_{ii} = 0$ e a matriz \mathbf{CM} sendo simétrica.

Denote $\mathbf{Q}^{-1} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda} (\mathbf{I} - \mathbf{B}^t)^{-1}$. A matriz \mathbf{Q}^{-1} é simétrica e definida positiva e, por consequência, a matriz $\mathbf{Q} = (\mathbf{I} - \mathbf{B}^t) \mathbf{\Lambda}^{-1} (\mathbf{I} - \mathbf{B})$ também é simétrica e definida positiva. Queremos $(\mathbf{I} - \mathbf{C})^{-1} \mathbf{M} = \mathbf{Q}^{-1}$, ou seja, $\mathbf{M}^{-1} (\mathbf{I} - \mathbf{C}) = \mathbf{Q}$, ou ainda, $\mathbf{C} = \mathbf{I} - \mathbf{M} \mathbf{Q}$. Entretanto, precisamos de ter $C_{ii} = 0$. Tome então $\mathbf{M} = [\text{diag}(\mathbf{Q})]^{-1}$. Esta matriz está bem definida pois $M_{ii} > 0$ já que, como \mathbf{Q}^{-1} é p.d., \mathbf{Q} também é p.d. e assim, $0 < \mathbf{x}^t \mathbf{Q}^{-1} \mathbf{x}$ para todo vetor \mathbf{x} . Em particular, se \mathbf{x} é um vetor de zeros, exceto pela presença de um único 1 na i -ésima posição, então $0 < \mathbf{x}^t \mathbf{Q}^{-1} \mathbf{x} = [\mathbf{Q}^{-1}]_{ii}$. Dessa forma,

$$[C]_{ij} = [\mathbf{I} - \mathbf{M} \mathbf{Q}]_{ij} = \begin{cases} -Q_{ij}/Q_{ii}, & \text{se } i \neq j \\ 0, & \text{se } i = j \end{cases}$$

Além disso, $(\mathbf{I} - \mathbf{C})^{-1} \mathbf{M} = \mathbf{Q}^{-1}$, o que implica que \mathbf{CM} é simétrica.

Vamos provar agora que todo modelo CAR implica num modelo SAR. Suponha que \mathbf{C} e \mathbf{M} são tais que existe $(\mathbf{I} - \mathbf{C})^{-1}$ com $M_{ii} > 0, C_{ii} = 0$ e \mathbf{CM} sendo simétrica. Vamos mostrar que existem \mathbf{B} e $\mathbf{\Lambda}$ tais que existe $(\mathbf{I} - \mathbf{B})^{-1}$ com

$$(\mathbf{I} - \mathbf{C})^{-1} \mathbf{M} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda} (\mathbf{I} - \mathbf{B}^t)^{-1},$$

$B_{ii} = 0$ e $\lambda_{ii} > 0, i = 1, \dots, n$.

A decomposição de Cholesky diz que se A é p.d. então existe uma única matriz U triangular superior com elementos da diagonal positivos tal que $A = U^t U$. Como $\mathbf{M}^{-1} (\mathbf{I} - \mathbf{C})$ é p.d., seja $U^t U = \mathbf{M}^{-1} (\mathbf{I} - \mathbf{C})$ a sua decomposição de Cholesky. Então U é triangular superior com $U_{ii} > 0$. Assim, queremos encontrar \mathbf{B} e $\mathbf{\Lambda}$ tais que

$$(U^t U)^{-1} = (\mathbf{I} - \mathbf{B})^{-1} \mathbf{\Lambda} (\mathbf{I} - \mathbf{B}^t)^{-1}$$

Isto é, queremos

$$U^t U = [\Lambda^{-0.5} (I - B)]^t [\Lambda^{-0.5} (I - B)]$$

Tome $U = \Lambda^{-0.5} (I - B)$, ou seja, $B = I - \Lambda^{0.5} U$. Como B_{ii} deve ser zero, escolha $[\Lambda^{0.5}]_{ii} = [U]_{ii}^{-1} > 0$. Todas as condições sobre as matrizes são válidas com esta escolha e temos assim o SAR correspondente a esse CAR.

Usando basicamente as mesmas idéias acima, podemos mostrar que:

Resultado: Os modelos equivalentes SAR e CAR geram a classe das distribuições normais n -variadas.

Prova: Basta mostrar que toda matriz de covariância Σ definida positiva pode ser escrita como as matrizes de um SAR e de um CAR repetindo os passos acima.

Desse modo, qualquer vetor com distribuição normal multivariada pode ser representado na forma de uma distribuição com estrutura espacial de um SAR ou de um CAR.

1.3 Old Stuff - from SBE course

- Suponha que $E(\mathbf{Y}) = \mathbf{0}$.
- O modelo SAR (Simultaneous AutoRegression) supõe que:

$$\mathbf{Y} = \rho W \mathbf{Y} + \boldsymbol{\epsilon}$$

$\boldsymbol{\epsilon}$ é vetor $N_n(\mathbf{0}, \Lambda)$ e Λ é matriz diagonal $(\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2)$.

- Distribuição de \mathbf{Y} é determinada pelas soluções simultâneas do sistema de equações estocásticas lineares acima
- Assim, $(I - \rho W) \mathbf{Y} = \boldsymbol{\epsilon}$
- Se $(I - \rho W)$ é inversível,

$$\mathbf{Y} = (I - \rho W)^{-1} \boldsymbol{\epsilon}$$

- Modelo SAR - Distribuição de \mathbf{Y}
- Se existe $(I - \rho W)^{-1}$, o modelo SAR está bem definido e

$$\mathbf{Y} \sim N_n(\mathbf{0}, V)$$

- onde

$$V = (I - \rho W)^{-1} \Lambda (I - \rho W^t)^{-1}$$

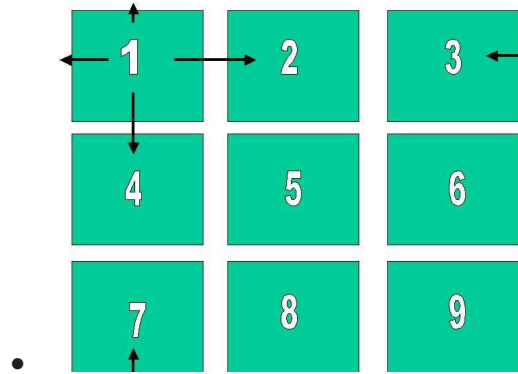
- Restrição em ρ e na matriz de vizinhança W no caso de modelo SAR: é necessário apenas que exista $(I - \rho W)^{-1}$
- ρ está restrito ao intervalo $[1/\lambda_{\min}, 1/\lambda_{\max}]$ onde λ_{\min} e λ_{\max} são os autovalores mínimo e máximo da matriz W
- W não precisa ser simétrica: i pode ser mais vizinho de j do que j de i .
- Modelo SAR - Covariância de \mathbf{Y}

-

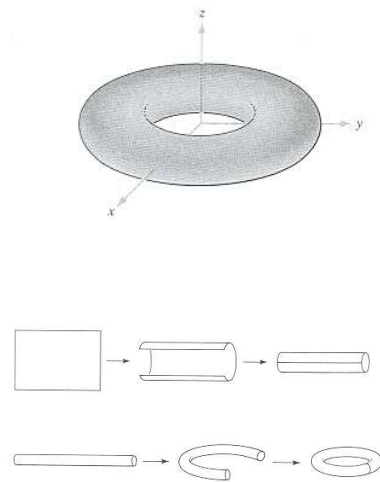
$$Cov(\mathbf{Y} = V = (I - \rho W)^{-1} \Lambda (I - \rho W^t)^{-1}$$

- Em geral, a matriz de vizinhança espacial W é *esparsa*: a maioria dos seus elementos é igual a zero
- No entanto, a matriz V será *cheia*. Isto implica em correlações $\neq 0$ entre localizações distantes. Em geral, decai com distância mas decaimento não é homogêneo (depende da posição e da direção)
- Em geral, a diagonal de V não será constante, mesmo que $Cov(\epsilon = \Lambda = \sigma^2 I_n$, um múltiplo da identidade.
- Assim, as observações não vêm de um processo espacial estacionário.
- Modelo SAR - Mais não-estacionariedade espacial
- Outra consequência desagradável: $Cov(y_i, y_j)$ não será invariante com a distância (vetorial) entre as áreas i e j
- Isto ocorre mesmo que a matriz W_{ij} seja constante com a distância.
- Isto também ocorre mesmo que $\Lambda = \sigma^2 I_n$.
- A principal razão para a não estacionariedade é que, num mapa típico de áreas, não existe homogeneidade de número de vizinhos ou de ou de padrão direcional de vizinhos.

- Exemplo
- Suponha uma grade regular com vizinhança dada pelo movimento do peão.
- Considere 9 áreas na forma de grade 3x3 com *simetria do toro* e movimento do peão para a estrutura de vizinhança.



- Transformando um retângulo num toro



- Exemplo - 2

- Se $|\rho| < 1/4$ então existe $(I - \rho W)^{-1}$

-

$$\rho W = 1/5 \times \begin{bmatrix} 0 & 1 & 1 & 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \end{bmatrix}$$

- A diagonal de $V = Cov(\mathbf{Y})$ é constante e igual a 3.70. A covariância dos vizinhos (de primeira ordem) é 2.84 e, para áreas não-vizinhas (ou de segunda ordem), é igual a 2.49.

- Exemplo - 3

-

- Com esse grau extremo de simetria, a matriz de covariância possui o padrão estacionário.

- Isto só acontece se o mapa for convertido em um toro

- Sem a conversão para toro, as variâncias não são constantes: 1.37, 1.60, 1.37, 1.60, 1.93, 1.60, 1.37, 1.60

- Covariância entre vizinhos é 0.67 a menos que uma das áreas seja a área 5 (central) quando então é 0.87

- E daí? Dificuldades para usar argumentos assintóticos baseados em uma única realização de um processo estocástico sem poder usar estacionariedade

- Mapa de MG dividido em n municípios: O que significa $n \rightarrow \infty$?

- Esta dificuldade assintótica não aparece em outros tipos de dados espaciais (superfície contínua e processos pontuais)

- Dificuldades para a estimação de ρ

- Numa série AR(1) $y_t = \rho y_{t-1} + \epsilon_t$ com distribuição normal, o estimador OLS $\hat{\rho} = \sum_t y_t y_{t-1} / \sum_t y_t^2$ é consistente e eficiente.
- No caso espacial, considere o modelo SAR sem variáveis explicativas

$$= \rho W + \epsilon$$

- Então o estimador OLS

$$\hat{\rho} = \sum_t W / (\sum_t 1)$$

é inconsistente.

- Estimação deve ser por máxima verossimilhança: veremos mais tarde
- Caso $E(\mathbf{Y}) \neq \mathbf{0}$
- Seja $E(\mathbf{Y}) = \mathbf{0}$. Então modelo SAR fica

$$\mathbf{Y} = \rho W(\mathbf{Y}) + \epsilon \text{ ou seja } \mathbf{Y} + (I - \rho W)^{-1} \epsilon$$

- Premultiplicando por $(I - \rho W)$ temos

$$\mathbf{Y} + \rho W \mathbf{Y} = \epsilon \text{ onde}$$

$(I - \rho W)$ é uma transformação 1-1.

- No modelo de regressão com covariáveis, faremos =
- Isto é,

$$\mathbf{Y} + \rho W \mathbf{Y} = \mathbf{X} + \epsilon$$

- Modelos SAR e CAR: reta \neq plano
- Em séries temporais, as duas especificações abaixo são equivalentes:

– a especificação simultânea

$$y_t = \rho y_{t-1} + \epsilon_t$$

com ϵ_t i.i.d. com $Var(\epsilon_t) = \sigma^2$

–

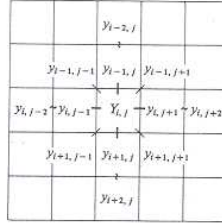
– e a especificação condicional

$$E(y_t | y_{t-1}, y_{t-2}, \dots) = \rho y_{t-1}$$

e

$$Var(y_t | y_{t-1}, y_{t-2}, \dots) = \sigma^2$$

- Esta equivalência *não ocorre* em modelos espaciais.
- Suponha um mapa na forma de uma grade regular
- Suponha um modelo SAR para os dados \mathbf{Y}
- Suponha uma matriz W binária simétrica determinada pelo movimento de peão
- Isto implica num modelo CAR equivalente
- Isto é, distribuição de \mathbf{Y} *podeseerescritacomoumCAR*.
- Entretanto, o modelo CAR equivalente ao SAR possui uma matriz W *diferente* da matriz W do modelo SAR.
- Matriz W de CAR é muito mais cheia que a de SAR: não-vizinhos no modelo SAR (ou vizinhos de segunda e terceira ordem) são vizinhos no modelo CAR.
- SAR: Covariância para W simétrica é $[(I - \rho W)^t (I - \rho W)]^{-1} = [I - 2\rho W + \rho^2 W^2]^{-1}$
- CAR: covariância baseada numa matriz de vizinhança W^* é igual a $[I - \rho W^*]^{-1}$
- Representação SAR corresponde a modelo CAR com vizinhos de primeira e segunda ordem.



- Os modelos CAR e SAR são equivalentes no seguinte sentido:
- Dado um vetor \mathbf{Z} com distribuição conjunta de probabilidade seguindo o modelo SAR, existe um modelo CAR correspondente que possui a mesma distribuição conjunta.
- O contrário também é válido: modelos CAR são expressos como modelos SAR.
- **Teorema:** Os modelos equivalentes SAR e CAR geram a classe de *todas* as distribuições normais n -variadas.
- No caso normal, decisão sobre qual modelo adotar (CAR vs SAR) passa a ser ditada pela teoria subjacente ao problema, ou é uma questão de gosto, ou de capacidade de interpretação dos parâmetros.

1.4 Maximum likelihood estimation

- Modelos de Regressão Espacial
- SAR - Lag espacial

$$= \rho W + X + \epsilon \text{ onde}$$

ϵ i.i.d $N(0, \sigma^2)$ e independente de X .

- SAR - Spatial Error Model

$$= X + \epsilon \epsilon \theta W \epsilon$$

com i.i.d $N(0, \sigma^2)$ e independente de X .

– No segundo modelo, espaço é meramente ruído (nuisance), Interesse principal é na estimação de β . Processo espacial de ϵ é apenas para levar em conta erro

- Uma variação de SAR com lag
- SAR - Spatial Durbin Model
- introduz defasagens espaciais nas variáveis exógenas

$$Y = \rho W Y + (I - \rho W) X \beta + \epsilon$$

- onde $\epsilon \sim i.i.d.N(0, \sigma^2)$ e independente de X .
- Caso Particular do modelo SAR de lag espacial
- Mais Modelos de Regressão Espacial
- CAR: Conditional Auto-Regressive Model.
- Seja $E(Y_i) = \mu_i$. Então a média condicional $\mu_{i|y} = E(Y_i | Y_1 = y_1, \dots, Y_n = y_n)$ e variância condicional $V_{i|y} = Var(Y_i | Y_1 = y_1, \dots, Y_n = y_n)$ são dadas por

$$\mu_{i|y} = \mu_i + \rho (W_{i1}(y_1 - \mu_1) + \dots + W_{in}(y_n - \mu_n)) \quad (1.4.6)$$

$$V_{i|y} = \sigma_i^2 \quad (1.4.7)$$

- Temos

$$Y \sim N_n(X, (I - \rho W)^{-1} \Lambda)$$

onde Λ é a matriz diagonal com os σ_i^2

- Estimação de β e ρ e λ
- OLS: inconsistência, vício, ineficiência
- 2SLS: inconsistência (Kelejian and Prucha, 1997)
- EMV: Condições de regularidade? Mas o que é assintótico?

- Duas abordagens:
 - * Infill asymptotics: região geográfica fica fixa mas unidades amostrais aumentam ficando mais densas.
 - * Expanding Domain Asymptotic: Áreas "internas" ficam fixas e aumenta-se o número de áreas na fronteira em todas as direções
- Tratamento para reticulados regulares (imagens)
 - * Neste caso, processos estacionários podem ser assumidos.
 - * Condições suficientes: limitam número de vizinhos na matriz de pesos assegurando que não crescem com o tamanho do reticulado
- Condições de regularidade para resultados assintóticos
- Caso de reticulados irregulares (grafos - mapa usual) é mais complicado
- Tratamento unificado apenas recentemente: Keleijian e Prucha (1998, 1999, 2001),
- Definição: Uma matriz A possui linhas e colunas uniformemente limitadas em valor absoluto se $\sum_i |A_{ij}|$ e $\sum_j |A_{ij}|$ são menores que constante C para todo i e j
- Exemplo de condições:
 - * Os elementos da matriz de pesos W são constantes conhecidas e o posto de $I - \rho W$ é n para todo $|\rho| < 1$.
 - * As matrizes W e $(I - \rho W)^{-1} (I - \rho W^t)^{-1}$ são linha-coluna uniformemente limitadas em valor absoluto.
 - * X é não-estocástica e de posto completo.
 - * $\lim_n n^{-1} X^t W^t W X$ é finita e não-singular
- Estimação OLS no modelo de lag espacial

- Considere o modelo

$$Y = \rho W + X + \epsilon \text{ onde}$$

$\epsilon_1, \dots, \epsilon_n$ i.i.d $N(0, \sigma^2)$ e independentes de X .

- Regressão com a variável dependente defasada y_{t-1} sendo usada como regressor: não temos problemas com OLS se o erro ϵ_t não é correlacionado com y_{t-1} .
- Isto ocorre se os erros ϵ_t são não correlacionados.

– Na situação espacial, a presença de WY como regressor induz endogeneidade independentemente de ρ .

– Como resultado, OLS será inconsistente (Anselin e Bera, 1998).

– Log-verossimilhança do modelo de lag espacial

- Considere o modelo

$$Y = \rho W + X + \epsilon \text{ onde}$$

ϵ i.i.d $N(0, \sigma^2)$ e independente de X .

- Então, existindo a inversa, temos

$$Y = (I - \rho W)^{-1} X + (I - \rho W)^{-1} \epsilon$$

- e a log-verossimilhança é

$$l(\rho, \sigma^2) = -\frac{n}{2} \log(2\pi\sigma^2) + \log |I - \rho W| - \frac{1}{2\sigma^2} Q$$

onde

$$Q = (-X)^t (-X)$$

- EMV do modelo no modelo de lag espacial
- O máximo único da log-verossimilhança é dado por:

$$\hat{\rho} = (X^t X)^{-1} X^t (I - \rho W)$$

$$\widehat{\sigma}^2 = \frac{1}{n} (I - \widehat{\rho}W)^t (I - X(X^t X)^{-1} X^t) (I - \widehat{\rho}W)$$

e $\widehat{\rho}$ minimiza a expressão

$$-\frac{2}{n} \sum_i \log(1 - \rho\lambda_i) + \log({}^t M - 2\rho^t MW + \rho^2(W)^t MW)$$

onde $M = I - X(X^t X)^{-1} X^t$ e λ_i são os autovalores de W

- Estimação de ρ no modelo de lag espacial
- A solução ρ pode ser encontrada por meio de um algoritmo de busca direta: uma grade fina é colocada entre $1/\lambda_{(1)} < 0$ e $1/\lambda_{(n)} > 0$ e percorrida exaustivamente. Em cada valor da grade calcule a função que envolve ρ e ache o valor que a maximiza.
- Pace e Barry (1997 e seguintes) mostraram que técnicas de matrizes esparsas podem ser usadas para estimar ρ rapidamente em matrizes da ordem de $10^4 \times 10^4$ elementos.
- Seus métodos estão implementados num pacote do MATLAB chamado *spatial econometrics*
- Pacote contém muitas outras funcionalidades: ver página de James LeSage
- Matriz de informação
- A matriz de covariância dos estimadores é dada pela inversa da matriz I de informação de Fisher avaliada no ponto de máximo.
- Seja $B = W(I - \widehat{\rho}W)^{-1}$

e

$$\alpha = - \sum_i \left(\frac{\lambda_i}{1 - \widehat{\rho}\lambda_i} \right)^2$$

-

$$I = \frac{1}{\widehat{\sigma}^2} \begin{bmatrix} n/2 & \widehat{\sigma}^2 \widehat{\rho} \text{tr}(B) & \widehat{\sigma}^2 \widehat{\rho} \text{tr}(B^t B) + \widehat{\sigma}^2 X^t B^t B X - \alpha \widehat{\sigma}^2 & \mathbf{0}^t \\ \widehat{\sigma}^2 \widehat{\rho} \text{tr}(B) & \widehat{\sigma}^2 \widehat{\rho} \text{tr}(B^t B) + \widehat{\sigma}^2 X^t B^t B X - \alpha \widehat{\sigma}^2 & \widehat{\sigma}^2 X^t B X & \mathbf{0}^t \\ \mathbf{0} \widehat{\sigma}^2 X^t B X & \widehat{\sigma}^2 X^t B X & \widehat{\sigma}^2 X^t X & \mathbf{0}^t \end{bmatrix}$$

– Estimação OLS no modelo de erros espaciais

– Considere o modelo

$$y = X\beta + \epsilon + \rho W\epsilon$$

com ϵ i.i.d $N(0, \sigma^2)$ e independente de X .

– Temos então

$$\hat{\beta} = X + (I - \rho W)^{-1}$$

– Suponha que $\hat{\beta}$ é o estimador OLS usual de β .

– Então, como em séries temporais, temos :

* $\hat{\beta}$ é não-viciado, consistente mas ineficiente (Anselin, 1988, p.59)

* OLS em 2 estágios (imitando procedimento de Cochrane-Orcutt, 1949) leva a estimadores inconsistentes (Kelejian e Prucha, 1997)

* Eles mostraram que a função resposta do modelo espacial autoregressivo associado com a transformação de Cochrane-Orcutt viola a condição do posto de Amemiya (1985, p.246) levando a problemas de identificação para o parâmetro espacial.

– Log-verossimilhança do modelo de erros espaciais

– Considere o modelo

$$y = X\beta + \epsilon + \rho W\epsilon$$

com ϵ i.i.d $N(0, \sigma^2)$ e independente de X .

– Temos então

$$\hat{\beta} = X + (I - \rho W)^{-1}$$

– A densidade conjunta de $(y, \hat{\beta})$ é obtida a partir daquela de y

$$f(y) = (1/2\pi\sigma^2)^{n/2} \exp(-y'y/(2\sigma^2))$$

e como $\hat{\beta} = (I - \rho W)^{-1}y$ então a verossimilhança de $(y, \hat{\beta})$ é dada por

$$f(y, \hat{\beta}) = |I - \rho W| / (2\pi\sigma^2)^{n/2} \exp(-\hat{\beta}'(I - \rho W)^t(I - \rho W)\hat{\beta} / (2\sigma^2))$$

- EMV no modelo de erros espaciais
- O máximo único da log-verossimilhança é dado por:

$$\hat{\gamma} = (X^t (I - \hat{\rho}W)^t (I - \hat{\rho}W) X)^{-1} X^t (I - \hat{\rho}W)^t (I - \hat{\rho}W)$$

$$\hat{\sigma}^2 = \frac{1}{n} (-X^t (I - \hat{\rho}W)^t (I - \hat{\rho}W) (-X))$$

e $\hat{\rho}$ minimiza a expressão

$$\log(\hat{\sigma}) - \frac{2}{n} \sum_i \log(1 - \rho\lambda_i)$$

onde λ_i são os autovalores de W .

- $\hat{\rho}$ é encontrado por busca direta em $(1/\lambda_{\min}, 1/\lambda_{\max})$.
- Matriz de informação do modelo de erros espaciais
- A matriz de covariância dos estimadores é dada pela inversa da matriz I de informação de Fisher avaliada no ponto de máximo.
- Seja $B = W(I - \hat{\rho}W)^{-1}$

e

$$\alpha = - \sum_i \left(\frac{\lambda_i}{1 - \hat{\rho}\lambda_i} \right)^2$$

- Então

$$I = \frac{1}{\hat{\sigma}^2} \begin{bmatrix} n/2 & \hat{\sigma} \operatorname{tr}(B) & \mathbf{0}t \\ \hat{\sigma} \operatorname{tr}(B) & \hat{\sigma}^2 (\operatorname{tr}(B^t B) - \alpha) & \mathbf{0}t \\ \mathbf{00} & \hat{\sigma} X^t (I - \hat{\rho}W)^t (I - \hat{\rho}W) X & \end{bmatrix}$$

1.5 Confidence intervals and tests

- * Teste de dependência espacial
- * Testes de presença de auto-correlação espacial em
 - Dados normais: I de Moran ou variantes (c de Geary, etc.)
 - Dados de contagem ou proporções: Moran Bayesiano empírico
- * Testes de autocorrelação espacial em regressão
 - I de Moran para resíduos de regressão
 - testes de métodos de momentos
- * Testes baseados em verossimilhança
 - Testes de multiplicadores de Lagrange/escore
 - testes de Wald e razão de verossimilhança baseados na estimação de MV do modelo espacial
- * Teste de presença de auto-correlação espacial em by
- * Sem covariáveis, mapa com com *. Hipótese nula: i.i.d.
- * Existe evidência de auto-correlação espacial em * ?
- * Defina $=^* -\bar{*}$.
- * I de Moran: estatística de teste é

$$I = \frac{n}{S_0} {}^tW / (by^t by)$$

onde

$$S_0 = \sum_{i \neq j} W_{ij}$$

- * Se i.i.d. pode-se obter distribuição assintótica: $I \sim N(-1/(n-1), V)$ onde V é dada por expressão envolvendo W e n
- * Expressão de V depende do tipo de modelo amostral
 - modelo de superpopulação para : dados são i.i.d. normais
 - modelo de permutação: ρ é fixo nos valores observados e distribuição de I é obtida assumindo que todas as alocações espaciais dos valores de Y são igualmente prováveis.
- * dados de taxas e proporções: Moran bayesiano empírico: Assunção e Reis (1999)
- * Testes de Hipótese em Modelos de Regressão
- * Existem dois tipos de testes de diagnóstico para dependência espacial em modelos de regressão:
 - testes para dependência espacial nos erros
 - testes para dependência substantiva ??
- * Abordagens comuns são aplicações baseadas no índice I de Moran e testes de multiplicadores de Lagrange baseados nos resíduos de uma regressão OLS.
- * Testes de Hipótese
- * Duas hipóteses: nula e alternativa
- * Hipótese Nula: regressão linear usual.
- * Hipóteses alternativas possíveis:
 - modelo SAR de lag espacial
 - modelo SAR de erro espacial
 - SMA, SEC, representação direta (não falarei sobre esses)

· modelo CAR

- * Modelo restrito (nulo) é o mesmo para todas as alternativas consideradas.
- * Problema: vários testes de má-especificação contra heterocedasticidade, forma funcional, etc.
- * Teste para dependência espacial nos erros: I de Moran
- * Sugerido por Moran (1950), esquecido e reavivado por Cliff and Ord (1972).
- * Rode regressão linear simples $\mathbf{Y} = \mathbf{X}\beta + \epsilon$ obtenha o vetor dos resíduos. A seguir, calcule

*

$$I = \frac{n}{S_0} \mathbf{1}' \mathbf{W} \mathbf{1} / \mathbf{1}' \mathbf{1}$$

onde $\mathbf{1}$ é o vetor de resíduos OLS e

$$S_0 = \sum_{i \neq j} W_{ij}$$

- * Cliff e Ord (1972): Sob hipótese de normalidade dos erros ϵ sob hipótese nula, assintoticamente $I \sim N(m_I, V)$ onde

$$m_I = \frac{-n \text{tr}(M W)}{(n - k) S_0}$$

com $M = I_n - X(X^t X)^{-1} X^t$. Existe longa expressão para V dependendo de X e W .

- * Similaridade com Durbin-Watson: $DW = \sqrt{A} / \sqrt{\mathbf{1}' \mathbf{1}}$ onde

$$A = \begin{bmatrix} 1 & -1 & 0 & 0 & \dots & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & \dots & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & \dots & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \dots & -1 & 2 & -1 \\ 0 & 0 & 0 & 0 & \dots & 0 & -1 & 1 \end{bmatrix}$$

- * DW é UMP para alternativas unilaterais com distribuição para o erro dada por $\epsilon_{t-1} = \lambda\epsilon_{t-1} + u_t$ (King, 1987, book Specification analysis in the linear model)
- * Propriedades ótimas de I de Moran como teste LM (localmente mais poderoso): Burrige (1981) mostrou que I é equivalente a estatística de teste LM para alternativa em que erros de regressão são $\epsilon\lambda W\epsilon$.
- * Distribuição EXATA de I de Moran
- * Tiefelsdorf e Boots (1995): distribuição exata de I de Moran !
- * Sejam $\lambda_1, \dots, \lambda_{n-k}$ os autovalores não nulos da matriz $n \times n$ dada por $MW = (I_n - X(X^t X)^{-1} X^t) W$.

- * Eles mostraram que

$$I \sim \frac{\sum_{i=1}^{n-k} \lambda_i Z_i^2}{\sum_{i=1}^{n-k} Z_i^2}$$

onde Z_1, \dots, Z_{n-k} são i.i.d. $N(0, 1)$

- * Valores críticos exatos para a distribuição nula de I podem ser obtidos por (difícil) integração numérica
- * Teste LM para dependência espacial nos erros
- * Sugerido por Burrige (1980)

- *

$$LM = \left(\sum W/\widehat{\sigma^2} \right)^2 / \text{tr}(W'W + W^2) \chi^2(1)$$

onde ϵ é o vetor de resíduos OLS e $\widehat{\sigma^2} = \epsilon' \epsilon / n$ é o EMV de σ^2 (SOB HO???)

1.6 Difficulties with non-normal data