

ELIAS TEIXEIRA KRAINSKI

**ESTRUTURA DE COVARIÂNCIA DE MODELOS
ESPACIAIS PARA DADOS DE ÁREAS**

Belo Horizonte
07 de março de 2008

ELIAS TEIXEIRA KRAINSKI
ORIENTADOR: RENATO MARTINS ASSUNÇÃO

**ESTRUTURA DE COVARIÂNCIA DE MODELOS
ESPACIAIS PARA DADOS DE ÁREAS**

Dissertação apresentada ao Programa de Pós-Graduação em Estatística da Universidade Federal de Minas Gerais como requisito parcial para a obtenção do grau de Mestre em Estatística.

ELIAS TEIXEIRA KRAINSKI

Belo Horizonte
07 de março de 2008



UNIVERSIDADE FEDERAL DE MINAS GERAIS

FOLHA DE APROVAÇÃO

Estrutura de covariância de modelos espaciais para dados de áreas

ELIAS TEIXEIRA KRAINSKI

Dissertação defendida e aprovada pela banca examinadora constituída por:

Ph. D. RENATO MARTINS ASSUNÇÃO – Orientador
Universidade Federal de Minas Gerais

Ph. D. PAULO JUSTINIANO RIBEIRO JUNIOR
Universidade Federal do Paraná

Ph. D. LOURDES CORAL CONTRERAS MONTENEGRO
Universidade Federal de Minas Gerais

Belo Horizonte, 07 de março de 2008

Agradecimentos

Primeiramente a Deus pela vida e por me guiar até aqui.

Ao prof. Renato, pela orientação segura e objetiva. Pelos conhecimentos passados. Pela motivação. Pelas grandes e pequenas idéias.

Aos professores da pós-graduação da UFMG que tive contato e pude extrair um pouquinho de experiência.

Ao prof. Paulo Justiniano, pelo *background* computacional e sugestões para a dissertação.

À profa. Lourdes pelas sugestões feitas para a dissertação.

Aos colegas da pós-graduação pela amizade e “companheirismo”. Não vou citar nomes para não correr o risco de esquecer algum.

Aos colegas do Leste, pelo ambiente amigável de trabalho.

Aos meus pais pelo apoio que deram para eu vir a Belo Horizonte.

À Raquel, o amor da minha vida, por todo o seu amor e compreensão e por me deixar dedicar tempo para a dissertação.

A todos que de alguma forma contribuíram para o bom período que passei em Belo Horizonte até aqui.

Resumo

Este trabalho está dividido em três Capítulos. Em todo o trabalho nós utilizamos o mapa dos EUA para facilitar a referência ao trabalho de Wall (2004), que foi o principal motivador deste trabalho. No primeiro Capítulo nós introduzimos os modelos SAR e CAR e fazemos uma análise de dados. Nós consideramos dados de Renda per Capita, Expectativa de vida e Percentual de Graduados nos EUA.

O capítulo 2 é um artigo submetido. Nesse Capítulo, nós mostramos mais detalhadamente os resultados não intuitivos. Nós consideramos resultados de álgebra linear e obtemos uma expressão simples e intuitiva para a matriz de covariância que explica os resultados não intuitivos. Nós obtemos termos para aproximações da matriz de covariância e estudamos algumas aproximações para a matriz de covariância. Nós também estudamos o segundo autovalor da matriz de vizinhança utilizada pelos modelos SAR e CAR e sua relação com os termos da expressão obtida para a matriz de covariância. Também estudamos o impacto da conectividade e do tamanho do mapa no segundo autovalor.

No Capítulo 3 nós consideramos um modelo espacial Bayesiano para dados gaussianos. Nesse modelo, consideramos um efeito aleatório com distribuição a priori CAR. Obtemos a distribuição a posteriori e uma expressão simples e intuitiva para a matriz de covariância a posteriori dos efeitos aleatórios. Nós avaliamos o impacto da informação a priori em relação a informação dos dados em termos da precisão da priori e da precisão dos dados. Obtemos também a expressão da covariância a distribuição posteriori dos efeitos aleatórios quando a distribuição à priori é CAR intrínseca. Nesse Capítulo, fazemos referência ao Capítulo 2 como um artigo submetido.

No Capítulo 4 nós tiramos algumas conclusões e colocamos algumas linhas de pesquisa para trabalhos futuros.

Sumário

1	Introdução aos modelos SAR e CAR e uma aplicação	1
1.1	Os modelos SAR e CAR	2
1.2	Um exemplo de aplicação	3
1.3	Covariâncias implicadas	6
2	Another close look at spatial structure of CAR and SAR models	9
2.1	Introduction	9
2.2	The SAR and CAR models	10
2.3	The puzzling results	11
2.4	Some preliminary definitions and results	13
2.4.1	Random graphs and the matrix W	13
2.4.2	A matrix identity	14
2.4.3	The powers of the W matrix	14
2.5	Revisiting the puzzling results	15
2.5.1	The CAR model with $\rho_c > 0$	15
2.5.2	The CAR model with $\rho_c < 0$	18
2.5.3	The SAR model	19
2.5.4	The role of $ \lambda_2 $	21
3	Covariância posteriori de efeitos aleatórios com priori CAR	24
3.1	Introdução	24
3.2	Modelos espaciais Bayesianos	25
3.3	Covariância à posteriori	27
3.4	Resultados não-intuitivos de Wall	29
3.5	Alguns resultados algébricos	32
3.6	Analisando a covariância a posteriori	35
3.6.1	Uma expressão simples	35
3.6.2	O caso distribuição a priori CAR intrínseca	37
3.6.3	Convergência e segundo autovalor	37
4	Conclusão	39
	Referências Bibliográficas	40

Lista de Figuras

1.1	Mapa dos 48 estados continentais dos EUA com grafo associado á vizinhança . . .	3
1.2	Mapas dos dados de Renda per Capita, Expectativa de vida e Percentual de Graduados	4
1.3	Variâncias implicadas pelo modelos SAR (a) e CAR (b) pelo número de vizinhos e correlações entre pares de vizinhos pelo número de vizinhos pelos modelos SAR (c) e CAR (d)	7
1.4	Correlações implicadas entre áreas vizinhas por ρ_s (esquerda) e ρ_c (direita). . . .	8
2.1	The graph of USA states by neighborhood (upper left), SAR correlations implied by number o neighbors if $\rho_s = 0.6$ (upper right) and the correlations implied by SAR and CAR models for all possible ρ_s (bottom left) and ρ_c (bottom right) values	12
2.2	Successively approximating the correlation between Vermont and Massachusetts as we increase the number of terms in the finite sums of (??), with $\rho_s = -0.99999$.	19
2.3	Edges of US states neighborhood graph drawn according to the pairwise correlation when ρ_c approaches its lower bound -1.3923 . Solid line: positive correlation, $+1$; dashed line: negative correlation, -1	20
2.4	Sucessively adding or pruning the adjacency neighborhood graph of four graphs. The geographical regions are the US states map, the the counties of Wyoming and Iowa, and the municipalities of Minas Gerais, a Brazilian state. The second column of plots shows five realizations of the of the addition-pruning process in each graph. The third column of plots shows 95% confidence envelopes based on the simulations in dashed lines, as well as the mean $ \lambda_2(j) $ value in solid line. . .	22
3.1	Correlação a posteriori entre vizinhos por ρ , para vários valores de τ_b	30
3.2	Número de correlações positivas entre pares de vizinhos em função de τ_b	31
3.3	Coefficiente de correlação de Kendall entre os postos de correlação entre pares de vizinhos implicada por diferentes pares de valores de ρ , em função de τ_b	31
3.4	Coefficiente de correlação de Kendall entre os postos de correlação entre pares de vizinhos implicada por pares de valores seqüenciais de ρ , em função de τ_b	32
3.5	Amplitude por τ_b para alguns valores de ρ	33
3.6	Segundo autovalor de \mathbf{W} (vertical) pela a razão entre os dois primeiros autovalores de \mathbf{W}^* (horizontal), considerando diferentes valores de τ_b	38

Lista de Tabelas

1.1	Estimativas obtidas para os parâmetros dos modelos	5
1.2	Intervalos de 95% de confiança para as estimativas da tabela 1.1	6
2.1	Values of the entries of \mathbf{W}^k , $\rho^k \mathbf{W}^k$, and the cumulative sum $\sum_{j=0}^k \rho^j \mathbf{W}^j$ for the pairs of neighboring states (Alabama, Florida) and (Alabama, Geórgia). We consider the values $\rho = 0.97$ and $\rho = 0.49$	16
2.2	Values of the entries of \mathbf{W}^k , $\rho^k \mathbf{W}^k$, and the cumulative sum $\sum_{j=0}^k \rho^j \mathbf{W}^j$ for the pair Vermont and Massachusetts. We consider $\rho = -0.99999$	18

Capítulo 1

Introdução aos modelos SAR e CAR e uma aplicação

Neste trabalho nós analisamos detalhadamente a estrutura de covariância de modelos espaciais para dados de áreas, tais como casos de doença agregados por municípios, o IDH em cada estado de um país, etc. Há dois modelos propostos na literatura que são mais utilizados. Um é especificado por um conjunto de equações de regressão da observação de uma área nas observações das áreas vizinhas. Esse sistema de equações é resolvido de forma simultânea, por isso esse modelo é chamado de *Simultaneous Autoregressive* - SAR. O outro modelo é especificado por um conjunto de distribuições condicionais, onde se propõe que a observação de uma área tem distribuição gaussiana com o parâmetro de média sendo a média ponderada das observações das áreas vizinhas e a variância sendo inversamente proporcional ao número de áreas vizinhas. Esse modelo é chamado de *Conditional Autoregressive* - CAR.

O parâmetro de regressão do modelo SAR e o parâmetro de ponderação da média condicional no modelo CAR são parâmetros de autoregressão. Esses parâmetros medem a força da dependência da observação de uma área nas observações das áreas vizinhas. Para ambos os modelos, é possível se obter a distribuição conjunta do vetor ao qual se propõe o modelo.

Na literatura, alguns trabalhos chamam a atenção para resultados não intuitivos da estrutura de covariância desses modelos, mais especificamente da correlação entre pares de áreas vizinhas. No modelo CAR a covariância entre áreas com mesmo número de vizinhos é diferente, Besag and Kooperberg (1995). Wall (2004) apontou detalhadamente três resultados não intuitivos nas correlações implicadas entre áreas vizinhas.

Neste trabalho, nós vamos mostrar como esses resultados não intuitivos podem ser explicados. Fazemos isso a partir do uso de alguns resultados de álgebra linear que usamos para analisar a matriz de covariância implicada pelos modelos SAR e CAR. Nós também usamos esses resultados para estudar o comportamento da covariância a posteriori num modelo Bayesiano com efeito aleatório espacial com distribuição a priori CAR.

Neste capítulo introduzimos os modelos SAR e CAR na Seção ???. Na Seção 1.2 fazemos uma aplicação ao conjunto de dados para ilustrar o uso dos modelos CAR e SAR. Na Seção 1.3 olhamos detalhadamente para as correlações implicadas por esses modelos e introduzimos os

resultados não intuitivos implicados.

1.1 Os modelos SAR e CAR

Seja uma região D , particionada em n áreas disjuntas, A_1, \dots, A_n com $A_i \cup A_j = \emptyset$ e $\cup_{i=1}^n A_i = D$. Os 853 municípios que fazem parte do estado de Minas Gerais, por exemplo, formam uma partição do estado de Minas Gerais. Seja y_i o valor observado de um determinado fenômeno na área A_i . O interesse é modelar o processo estocástico $Y(A_i)$, $i = 1, \dots, n$ ou simplesmente $\mathbf{y} = (y_1, \dots, y_n)$. Dois modelos para esse processo foram propostos e são muito utilizados: o modelo autoregressivo simultâneo - SAR Whittle (1954) e o modelo autoregressivo condicional - CAR Besag (1974).

O modelo SAR é determinado pela solução simultânea do sistema de equações dado por:

$$y_i = \mu_i + \sum_{j=1}^n b_{ij}(y_j - \mu_j) + \epsilon_i \quad (1.1)$$

onde $\epsilon = (\epsilon_1, \dots, \epsilon_n)' \sim N(0, \Lambda)$ com Λ diagonal, $E(y_i) = \mu_i$, e b_{ij} são constantes conhecidas ou desconhecidas e $b_{ii} = 0$, $i = 1, \dots, n$. A distribuição conjunta de \mathbf{y} é

$$\mathbf{y} \sim N(\boldsymbol{\mu}, (I_n - B)^{-1} \Lambda (I_n - B)^{-1'}) , \quad (1.2)$$

em que $B_{ij} = (b_{ij})$.

O modelo CAR é determinado por um conjunto de distribuições condicionais

$$y_i | \mathbf{y}_{-i} \sim N(\mu_i + \sum_{j=1}^n c_{ij}(y_j - \mu_j), \tau/d_i) \quad (1.3)$$

em que $\mathbf{y}_{-i} = \{y_j : j \leq i\}$ são os valores de \mathbf{y} nas áreas vizinhas de i , $E(y_i) = \mu_i$, τ e d_i c_{ij} são constantes conhecidas ou desconhecidas $c_{ij} = 0$, $i = 0, \dots, n$. Este modelo é *condicional* por que τ/d_i é a variância condicional.

$$\mathbf{Z} \sim N(\boldsymbol{\mu}, (I_n - C)^{-1} T^{-1}) \quad (1.4)$$

em que $C_{ij} = c_{ij}$ e $T = \tau \text{diag}\{d_1, \dots, d_n\}$.

Geralmente adota-se $B = \rho_s \mathbf{W}$ e $C = \rho_c \mathbf{W}$, em que ρ_s e ρ_c são parâmetros de correlação espacial dos modelos CAR e SAR, respectivamente, e \mathbf{W} reflete a estrutura de vizinhança entre as áreas. Uma definição bastante comum de \mathbf{W} é feita a partir da matriz de adjacência \mathbf{A} . A matriz de adjacência é definida fazendo $\mathbf{A}_{ij} = 1$ se as áreas i e j tem borda comum e $\mathbf{A}_{ij} = 0$ caso contrario. Por definição $\mathbf{A}_{ii} = 0$. Neste trabalho consideramos que \mathbf{W} é definida de forma que suas linhas somem 1, fazendo $\mathbf{W}_{ij} = (a_{ij}/a_i)$, em que $a_i = d_i$ é o número de vizinhos da área i .

1.2 Um exemplo de aplicação

Nesse trabalho, consideramos o mapa formado pelos 48 estados continentais dos EUA, para fazer comparações com os resultados apontados em Wall (2004). O mapa com as divisões políticas desses 48 estados e o grafo associado à estrutura de vizinhança, pode ser visualizado na Figura 1.1. Maine, localizado ao norte da costa leste, é um estado com apenas um estado vizinho, New Hampshire.

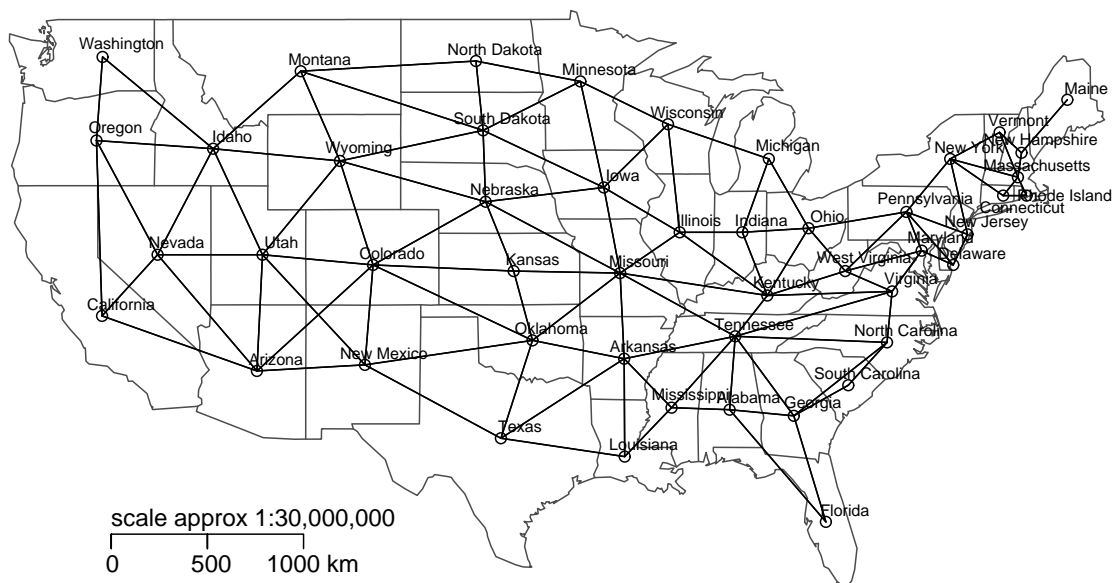


Figura 1.1: Mapa dos 48 estados continentais dos EUA com grafo associado à vizinhança

Em seguida analisaremos dados de renda per capita, expectativa de vida e percentual de graduados para exemplificar as correlações implicadas pelos modelos SAR e CAR. Esses dados estão disponíveis no pacote 'datasets' do R, R Development Core Team (2007), sob o nome de **state**. Na Figura 1.2 podemos visualizar os mapas correspondentes aos dados mencionados.

Na análise vamos considerar dois modelos, um para o logaritmo da renda per capita e outro para a expectativa de vida. Em ambos modelos, consideramos o percentual de graduados como variável explicativa. Para efeito de comparação, ajustamos os modelos utilizando três abordagens. Na primeira, consideramos as observações nas diferentes áreas são independentes e utilizamos um modelo de regressão linear simples. Na segunda, modelamos a dependência espacial como um processo contínuo no espaço, com a correlação dependendo da distância entre os centróides dos estados, e utilizamos um modelo geoestatístico (Geo), Diggle and Ribeiro Jr. (2007). Na terceira abordagem, consideramos a dependência espacial como um campo aleatório markoviano gaussiano, com a média de cada estado dependendo dos seus vizinhos, e utilizamos o modelo SAR e o modelo CAR.

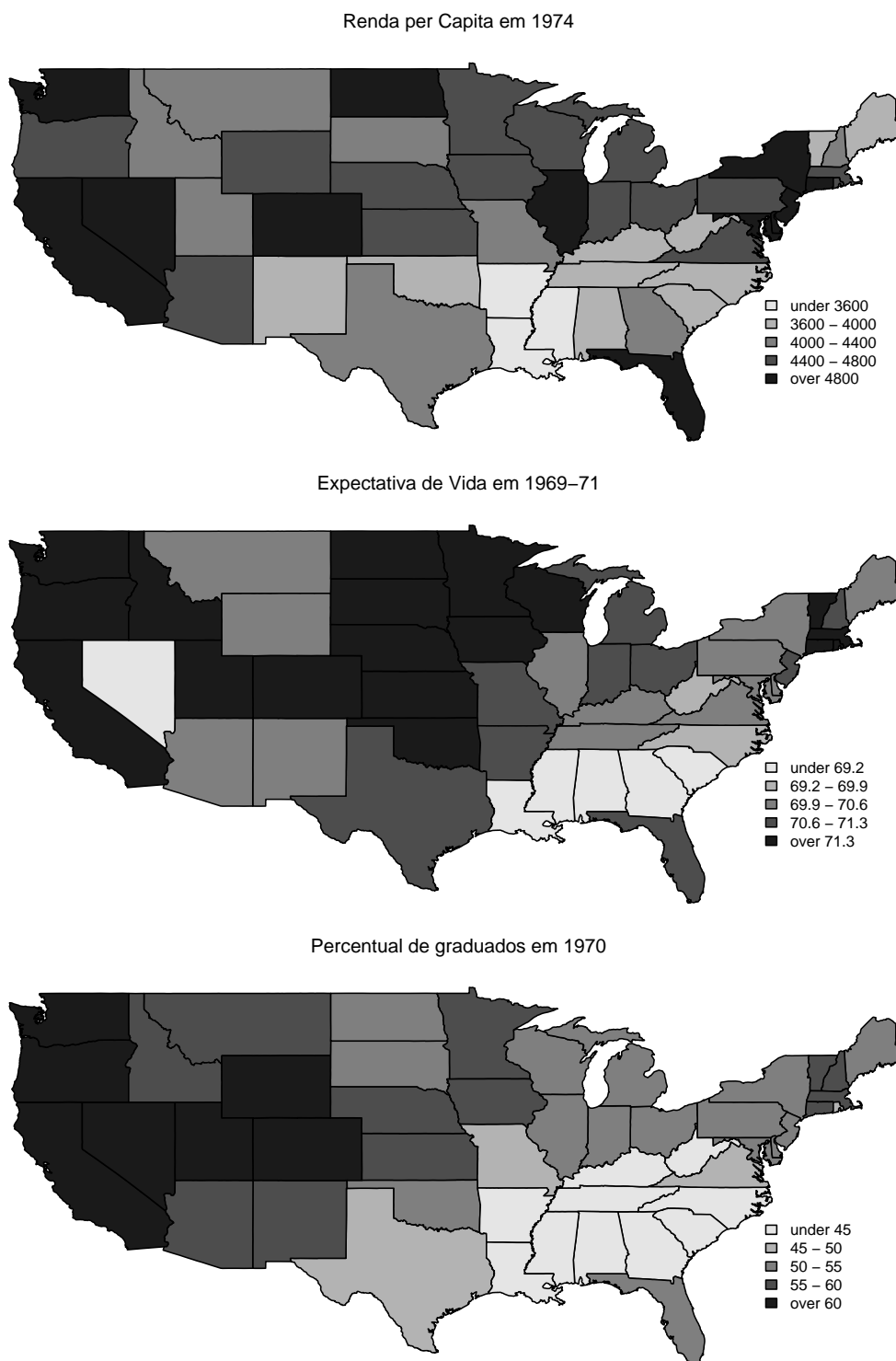


Figura 1.2: Mapas dos dados de Renda per Capita, Expectativa de vida e Percentual de Graduados

O modelo geoestatístico é um modelo para campo aleatório em espaço contínuo. Nós vamos considerá-lo apenas para comparação com os outros modelos. Esse modelo é definido por

$$y_i = \mathbf{X}_i\beta + s_i + e_i \quad (1.5)$$

onde \mathbf{X}_i é o vetor de covariáveis observadas na área i , s_i é o efeito aleatório espacial e e_i é um erro aleatório com e_i independente de e_j , chamado de efeito pepita. s_i é espacialmente estruturado de forma que $Cov(s_i, s_j)$ é uma função da distância entre as áreas i e j , por exemplo, a função de correlação exponencial, que usamos neste trabalho é tal que $Cov(s_i, s_j) = Var(e_i) + Var(s_i)(1 - exp(-dist_{ij}/\phi))$, onde $dist_{ij}$ é a distância entre os centróides das áreas i e j e ϕ é um parâmetro que mede o alcance da correlação ou dependência espacial. Para mais detalhes ver Diggle and Ribeiro Jr. (2007).

Nós utilizamos o **R** para realizar todas as análises. Utilizamos a função `lm()` para o modelo de regressão linear (IID), implementamos os modelos SAR e CAR, e utilizamos o pacote **geoR**, Ribeiro Jr and Diggle (2001) para ajustar o modelo geoestatístico. Na tabela 1.1 obtemos as estimativas dos parâmetros em cada modelo utilizando o método de máxima verossimilhança. Nós não incluímos o erro-padrão de cada estimativa nessa Tabela e fazemos inferência utilizando o intervalo de confiança, da Tabela 1.2. No modelo geoestatístico o efeito pepita, variância de e_i , também foi estimado, porém a estimativa foi igual a zero em ambos os modelos.

Tabela 1.1: Estimativas obtidas para os parâmetros dos modelos

Logaritmo da Renda				
	IID	Geo	SAR	CAR
β_0	7.8586	7.8198	7.7451	7.7525
β_1	0.0099	0.0104	0.0119	0.0118
σ^2	0.0111	0.0106	0.0334	0.0325
ϕ, ρ_s, ρ_c		3.0220	0.5527	0.8278
log(Veros.)	40.3206	45.6752	45.5102	45.2606
Expectativa de Vida				
	IID	Geo	SAR	CAR
β_0	65.3427	67.0368	66.2299	66.2774
β_1	0.1047	0.0725	0.0884	0.0875
σ^2	0.9861	1.0398	3.8152	3.7236
ϕ, ρ_s, ρ_c		3.6359	0.3052	0.5687
log(Veros.)	-67.2729	-62.5750	-66.7247	-66.6877

Notamos que a verossimilhança do modelo IID é menor que a dos modelos onde se considera a dependência espacial, para ambas variáveis. Fazendo um teste de razão de verossimilhanças, obtemos que a dependência espacial deve ser levada em conta na modelagem da renda per capita. Isso implica que um estado tende a ter comportamento semelhante ao dos seus vizinhos, quanto à renda per capita, ao nível de 5% de significância. Isso porque a diferença entre o logaritmo da verossimilhança dos modelos com dependência espacial e do modelo IID é maior

que 3,84, o valor crítico baseado na distribuição qui-quadrado com 1 grau de liberdade. No caso da expectativa de vida, a dependência é significativa quando modelada utilizando o modelo geoestatístico, com alcance estimado de 3.64, mas não é significativa quando utilizado os modelos SAR e CAR, com coeficientes de correlação estimados de 0.31 e 0.57 respectivamente. Na tabela 1.2 temos os intervalos de confiança dos parâmetros e podemos tirar melhores conclusões, como por exemplo, considerar se o zero está contido nos intervalos. Notamos, por exemplo, que os intervalos para ρ_s e ρ_c contêm zero nos modelos para expectativa de vida.

Tabela 1.2: Intervalos de 95% de confiança para as estimativas da tabela 1.1

Logaritmo da Renda				
	IID	Geo	SAR	CAR
β_0	(7.6528; 8.0644)	(7.5670; 8.0726)	(7.4710; 8.0192)	(7.4806; 8.0244)
β_1	(0.0060; 0.0137)	(0.0058; 0.0150)	(0.0068; 0.0170)	(0.0067; 0.0168)
σ^2	(0.0075; 0.0168)	(0.0070; 0.0204)	(0.0229; 0.0512)	(0.0223; 0.0498)
ϕ, ρ_s, ρ_c		(1.3924; 7.5448)	(0.2606; 0.8018)	(0.4505; 0.9917)
Expectativa de vida				
β_0	(63.4066; 67.2788)	(5.2854; 69.5712)	(63.9347; 68.5250)	(63.9140; 68.6409)
β_1	(0.0683; 0.1411)	(-0.0356; 0.1185)	(0.0454; 0.1315)	(0.0432; 0.1318)
σ^2	(0.6644; 1.4820)	(0.6669; 2.2475)	(2.6224; 5.8580)	(2.5590; 5.7163)
ϕ, ρ_s, ρ_c		(1.6134; 10.0071)	(-0.0544; 0.6327)	(-0.0831; 0.9395)

Os intervalos de confiança de β_0 e β_1 foram calculados considerando normalidade assintótica dos estimadores. Para os demais parâmetros, nós utilizamos o intervalo de confiança baseado na função de verossimilhança perfilhada. Considerando os modelos para o logaritmo da renda per capita, o percentual de graduados é significativo, para todos os modelos, pois o intervalo de confiança para β_1 contém o zero. Para expectativa de vida, observamos um resultado interessante: No modelo IID o percentual de graduados é significativo; no modelo geoestatístico isso não acontece mas a dependência espacial é significativa; e, nos modelos SAR e CAR essa covariável é significativa, porém a dependência espacial não é!

Esse resultado pode ter ocorrido devido ao fato de ambas as variáveis, expectativa de vida e percentual de graduados, terem padrão espacial parecido, Figura 1.2. No modelo IID houve significância por ambas estarem correlacionadas. No modelo geoestatístico, a dependência espacial considerada fez com que não houvesse significância, enquanto que nos modelos SAR e CAR, a correlação entre as variáveis fez com que o efeito espacial não fosse significativo.

1.3 Covariâncias implicadas

Nesta Seção nós olhamos mais detalhadamente para a estrutura de covariância implicada pelos modelos SAR e CAR e introduzimos os resultados não intuitivos apontados em Wall (2004). Na Figura 1.3, observamos as variâncias em cada área e as correlações entre áreas vizinhas implicadas pelos modelos SAR e CAR, considerando as estimativas dos parâmetros dos modelos ajustados para o logaritmo da renda per capita, ou seja, $\rho_s = 0.5527$ e $\rho_c =$

0.8278. Nos gráficos das correlações implicadas por número de vizinhos, ambos os pares $(Corr(i, j), d_i)$ e $(Corr(i, j), d_j)$ estão plotados.

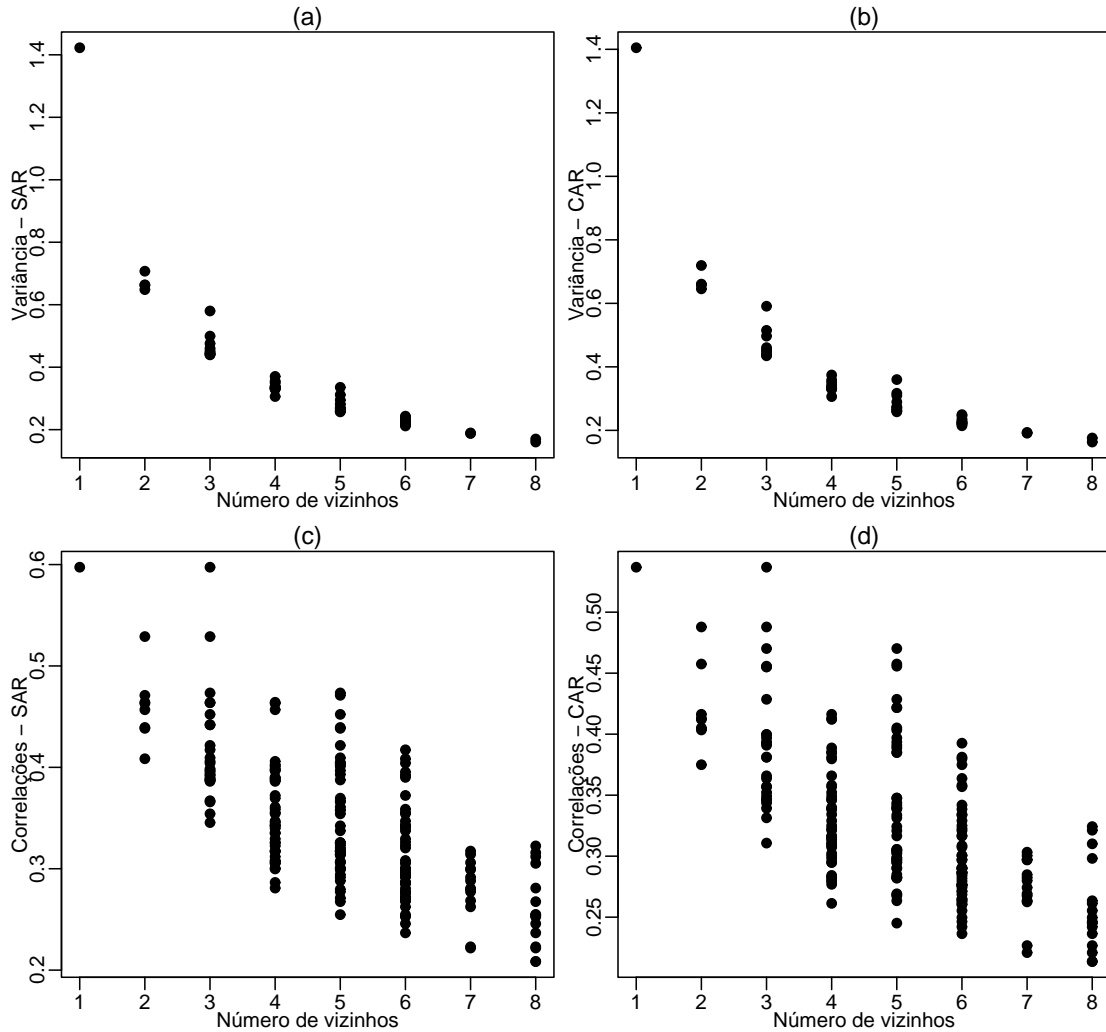


Figura 1.3: Variâncias implicadas pelo modelos SAR (a) e CAR (b) pelo número de vizinhos e correlações entre pares de vizinhos pelo número de vizinhos pelos modelos SAR (c) e CAR (d)

Observando os gráficos de dispersão entre a variância implicada em cada área e o número de vizinhos, na Figura 1.3, notamos que a variância tende a ser inversamente proporcional ao número de vizinhos, porém não é uma relação linear. Quanto aos gráficos da correlação implicada entre pares de vizinhos pelo número de vizinhos de cada um, notamos que a correlação tende a ser menor quando uma área do par tem menos vizinhos, porém essa relação também não é perfeita.

Agora, nós usaremos alguns valores para ρ_s e ρ_c para apresentar três resultados não intuitivos apontados em Wall (2004). Esses resultados ocorrem para ambos os modelos.

1. A variabilidade nas correlações entre vizinhos aumenta em função de ρ_s ou ρ_c . Considerando $\rho_s = 0.1$, as correlações entre os estados vizinhos variam de 0.026 a 0.115 e para

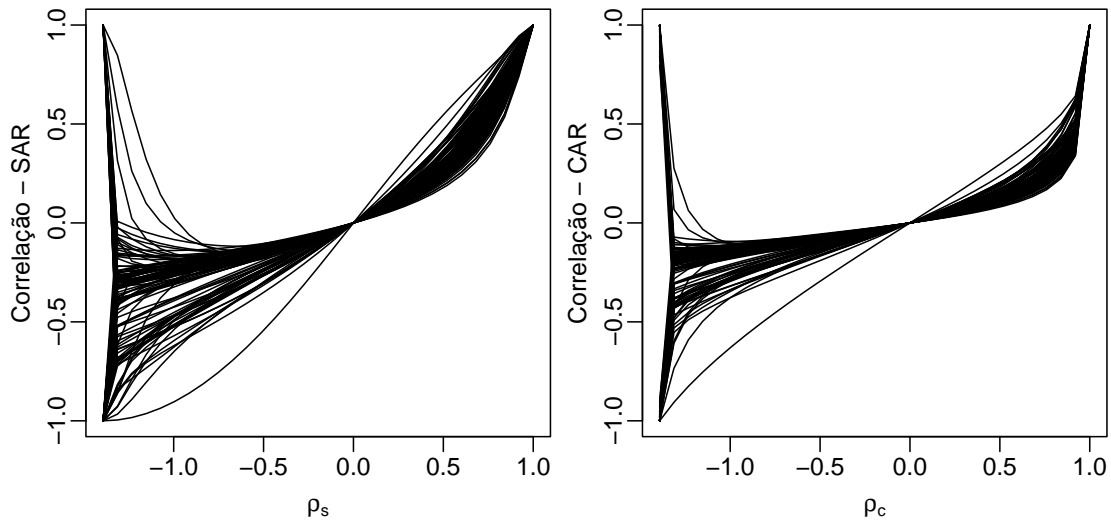


Figura 1.4: Correlações implicadas entre áreas vizinhas por ρ_s (esquerda) e ρ_c (direita).

$\rho_s = 0.6$, de 0.241 a 0.642.

2. Inconsistência no ranking das correlações. Considerando $\rho_c = 0.49$, a correlação entre Alabama e Flórida é 0.200 enquanto a correlação entre Alabama e Geórgia é 0.156. Entretanto, se $\rho_c = 0.97$, a correlação entre Alabama e Flórida é 0.650, menor que a correlação entre Alabama e Geórgia, que neste caso é 0.671.
3. A correlação implicada por ρ_s ou ρ_c negativos, pode ser positiva para alguns pares de vizinhos. Com $\rho_s = -0.7$ a correlação entre Maryland e Pennsylvania é -0.180 e entre Vermont e Massachussets é -0.157 . Mas se $\rho_s = -1.37$, passam a ser 0.463 e 0.985, respectivamente.

Se consideramos todos os valores possíveis de ρ_s e ρ_c ou seja, entre $1/\min\{\lambda_i\}$ e $1/\max\{\lambda_i\}$, podemos visualizar esses resultados fazendo o gráfico da correlação implicada entre pares de vizinhos por ρ_s e ρ_c . Na Figura 1.4, temos as correlações entre estados vizinhos implicadas pelos modelos SAR e CAR em função de ρ_s e ρ_c . Nessa Figura podemos visualizar os três resultados não intuitivos para ambos os modelos.

Esses resultados foram apontados em Wall (2004) e no próximo Capítulo nós os estudaremos em mais detalhes e usamos resultados de álgebra linear para compreendê-los. No Capítulo 3 nós estudamos esses resultados não intuitivos para um modelo espacial Bayesiano. No Capítulo 4 nós fazemos algumas conclusões e recomendações para trabalhos futuros.

Capítulo 2

Another close look at spatial structure of CAR and SAR models

2.1 Introduction

Lattice data refer to statistical data observed at spatial locations or areas in a given geographical region. It is common to assume that observations at sites near each other tend to have similar values. The Conditional Autoregressive (CAR) and the Simultaneous Autoregressive (SAR) models are widely used to analyze these lattice data. The SAR model is preferred in likelihood inference, while the CAR model is more common in Bayesian inference as a prior distribution for spatially structured random effects.

Despite their popularity, these models bring uneasy consequences for the implied correlation structure of the variables. Several authors have pointed out that the SAR and CAR models yield non constant variances at each site as well as unequal covariances between regions separated by the same number of neighbors (Haining (1990), page 82; Besag and Kooperberg (1995)).

Wall (2004) extensively studied the covariance structure entailed by these models. She found that the implied correlation between a pair of neighboring areas is negatively associated with the number of neighbors of each region. However, she also showed that this relationship is not simple and much variability remains unexplained. For example, considering the three neighboring US states Missouri, Arkansas, and Tennessee, she showed that, although Missouri and Tennessee have the same neighboring structure, their correlation with Arkansas differs. She also showed that sites with equal number of neighbors can have different variances implied.

In addition to these uncomfortable results, Wall (2004) pointed out a series of puzzling results from these two spatial models. One of them is that correlations between areas switch their ranks depending on ρ , the spatial dependence parameter. Suppose that a pair (i, j) of sites are more correlated than another pair (k, l) when $\rho = 0.5$. It is not uncommon that for another value of the correlation parameter, $\rho = 0.7$ say, the pair (k, l) becomes more correlated than (i, j) . One would expect, perhaps naively, that the order should be the same, irrespective of the spatial dependence parameter value. Even more puzzling are the results

concerning negative values for ρ . She found that when ρ is negative, correlations between the neighboring areas are also negative but, as ρ decreases further, some pairs of areas start to be positively correlated, even approaching +1 at times.

Wall (2004) concluded that the implied spatial correlation between the different sites using the SAR and CAR models does not seem to follow an intuitive or practical scheme and called for more research to be carried out to clarify these problems. This is the main purpose of this paper. We explain the apparently counterintuitive or impractical consequences of the model specification by using the complete neighborhood graph structure, not only the immediate neighborhood. In accounting for the complete neighborhood structure, we see that a crucial role is played by the second largest eigenvalue modulus of the neighborhood matrix used in the SAR and CAR models. We use a simple matrix algebra identity to write the covariance matrices of the SAR and CAR models as a matrix power series. This enables us to express the correlation between any two pairs of areas i and j as an infinite series with exponential decay given by the spatial dependence parameter ρ . Moreover, the k -th term coefficient of this series is proportional to a weighted sum of the different paths to move from area i to area j in k steps.

2.2 The SAR and CAR models

Let a region D be partitioned into n areas $\{A_1, \dots, A_n\}$ such that $D = A_1 \cup \dots \cup A_n$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$. Let y_i be a random variable measured at area i and $\mathbf{y} = (y_1, \dots, y_n)^t$. We denote by \mathbf{y}_{-i} the $(n-1)$ -dimensional vector without the i -th coordinate of \mathbf{y} . The conditional autoregressive model (CAR) is given by a set of n conditional distributions

$$y_i | \mathbf{y}_{-i} \sim N \left(\mu_i + \sum_{j=1}^n c_{ij} (y_j - \mu_j), v_i^2 \right) \quad (2.1)$$

where $c_{ii} = 0$ and $v_i^2 > 0$ for $i = 1, \dots, n$. It is not any set of n conditional distributions that determine uniquely a joint distribution for the vector \mathbf{y} . However, a very popular choice in spatial studies for the constants c_{ij} and v_i defines a valid joint model, and we adopt this choice in the rest of this paper.

The choice of the $n \times n$ matrix $\mathbf{C} = (c_{ij})$ is related to the degree of spatial proximity between areas i and j . Let $\mathbf{A} = (a_{ij})$ be an $n \times n$ binary neighborhood matrix such that $a_{ij} = 1$ if, and only if, areas i and j are neighbors (denoted by $i \sim j$). We let $a_{ii} = 0$. Define $\mathbf{W} = (w_{ij})$ such that $w_{ij} = a_{ij}/a_i$, where $a_i = \sum_j a_{ij} = d_i$, the number of neighboring areas of region i . Finally, define $\mathbf{C} = \rho_c \mathbf{W}$ and $v_i = \sigma_c/d_i$. Under a restriction on the value of ρ_c , the CAR model (2.1) with these options defines a valid joint distribution for the vector \mathbf{y} given by a multivariate normal distribution:

$$\mathbf{y} \sim N_n \left(\boldsymbol{\mu}, (\mathbf{I} - \rho_c \mathbf{W})^{-1} \mathbf{K} \right) \quad (2.2)$$

where $\boldsymbol{\mu} = (\mu_1, \dots, \mu_n)'$, \mathbf{I} is the identity matrix and \mathbf{K} is the diagonal matrix $\text{diag}(v_1, \dots, v_n)$

which is equal to $\sigma_c^2 \text{diag}(d_1^{-1}, \dots, d_n^{-1})$. The restriction on ρ_c is necessary to ensure that $(\mathbf{I} - \rho_c \mathbf{W})^{-1} \mathbf{K}$ is positive definite and it suffices to take ρ_c such that ρ_c is between $1/\min_i \lambda_i$ and $1/\max_i \lambda_i$ where $\lambda_i, i = 1, \dots, n$, are the eigenvalues of \mathbf{W} (Haining, 1990, page 82).

This choice also implies that (2.1) reduces to

$$y_i | \mathbf{y}_{-i} \sim N\left(\mu_i + \rho_c \overline{(y - \mu)}_i, \sigma_c^2/d_i\right) \quad (2.3)$$

where $\overline{(y - \mu)}_i = \sum_j w_{ij}(y_j - \mu_j)$ is the average of the deviations $y_j - \mu_j$ among $j \sim i$, i.e. among the neighboring areas of i .

The SAR model is defined by n simultaneous equations

$$y_i = \mu_i + \sum_{j=1}^n s_{ij}(y_j - \mu_j) + \epsilon_i \quad (2.4)$$

where $\epsilon = (\epsilon_1, \dots, \epsilon_n)' \sim N(0, \mathbf{\Lambda})$ with $\mathbf{\Lambda}$ diagonal with $\Lambda_{ii} = \sigma_s/d_i$, $E(y_i) = \mu_i$, and s_{ij} are known constants with $s_{ii} = 0, i = 1, \dots, n$. This model is *simultaneous* because the random variables are simultaneously determined by the n equations in 2.4. Provided that the inverse of the matrix $I_n - \mathbf{S}$ exists, the distribution of $\mathbf{y} = (y_1, \dots, y_n)'$ is

$$\mathbf{y} \sim N(\boldsymbol{\mu}, (I_n - \mathbf{S})^{-1} \mathbf{\Lambda} (I_n - \mathbf{S})^{-1'}) , \quad (2.5)$$

where $S_{ij} = s_{ij}$. A popular choice for \mathbf{S} is to take $\mathbf{S} = \rho_s \mathbf{W}$, where $\rho_s \in (-1, 1)$. Following Wall (2004), we will constrain ρ_s to the same interval as ρ_c in order to allow for comparisons between the models.

With these choices for the SAR and CAR model, the correlation matrix entries are functions of only \mathbf{W} and ρ_c or ρ_s . For the SAR model, we have

$$\text{Cor}(i, j) = \frac{\sigma_s^2 (I - \rho_s \mathbf{W})_{ij}^{-1} d_j^{-1} (I - \rho_s \mathbf{W})_{ji}^{-1}}{\sqrt{\sigma_s^2 (I - \rho_s \mathbf{W})_{ii}^{-2} d_i^{-1}} \sqrt{\sigma_s^2 (I - \rho_s \mathbf{W})_{jj}^{-2} d_j^{-1}}},$$

and σ_s^2 is canceled out. For the CAR model, we have

$$\text{Cor}(i, j) = \frac{\sigma_c^2 (I - \rho_c \mathbf{W})_{ij}^{-1} d_j^{-1}}{\sqrt{\sigma_c^2 (I - \rho_c \mathbf{W})_{ii}^{-1} d_i^{-1}} \sqrt{\sigma_c^2 (I - \rho_c \mathbf{W})_{jj}^{-1} d_j^{-1}}},$$

and σ_c^2 is canceled out.

2.3 The puzzling results

We summarize the main puzzling results concerning the correlations implied by the SAR and CAR models and described by Wall (2004). She used the United States map to illustrate the implications that the CAR and SAR models entail for the covariance between pairs of areas. Consider the graph composed by the 48 contiguous continental states. Two states i

and j are connected by an edge (meaning that $w_{ij} > 0$) if they share borders. This graph is in Figure 2.1, with the underlying US map. The upper right plot in Figure 2.1 shows the correlations $\text{Cor}(i, j)$ between pairs of neighboring states by the number of neighbors. Every pair (i, j) of neighboring areas contribute two points in this plot depending on each area's number of neighbors, the pair $(d_i, \text{Cor}(i, j))$ and the pair $(d_j, \text{Cor}(i, j))$. We can see that, for a given number of neighbors, there is a large variation in the correlations.

The lower row of plots in Figure 2.1 shows how the correlations $\text{Cor}(i, j)$ varies with the spatial dependence parameter ρ . Each line represents the correlation between two neighboring areas and the horizontal axis corresponds to the spatial dependence parameter ρ_s of the SAR model (left hand side plot), or ρ_c for the CAR model (right hand side plot). Based on the eigenvalues of \mathbf{W} for the US lattice, the spatial parameter space is restricted to $(-1.392, 1)$.

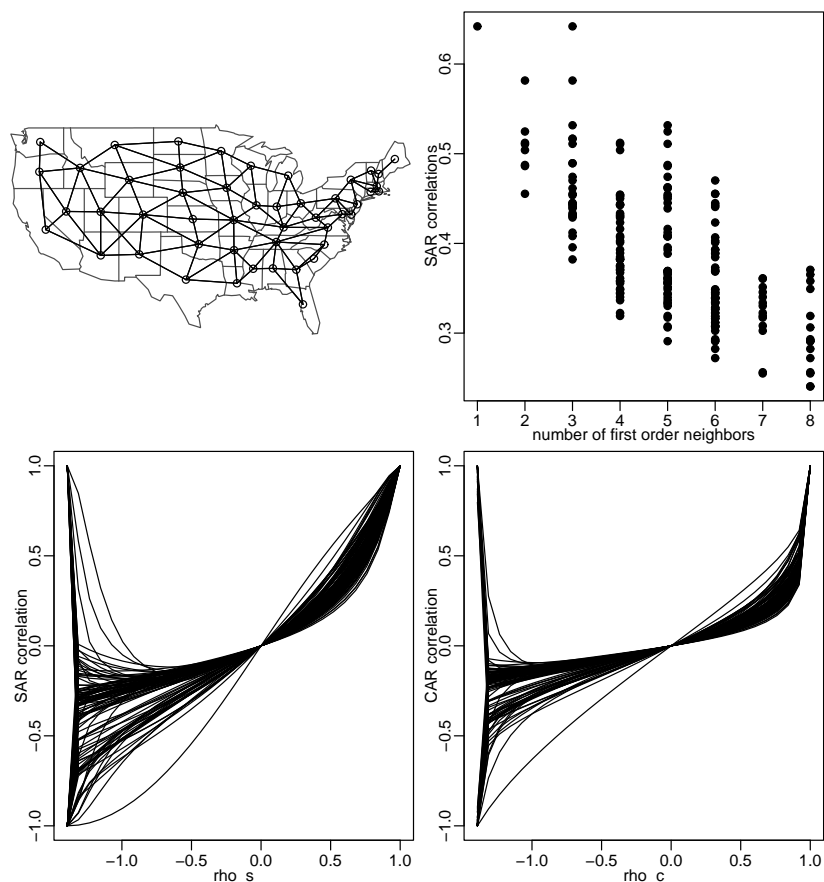


Figure 2.1: The graph of USA states by neighborhood (upper left), SAR correlations implied by number of neighbors if $\rho_s = 0.6$ (upper right) and the correlations implied by SAR and CAR models for all possible ρ_s (bottom left) and ρ_c (bottom right) values

Most of the puzzling results appear in these plots. We can see that lines cross each other as ρ varies, irrespective of the model adopted. This means that, if we increase the spatial correlations between all pairs of areas by increasing ρ , states which are less correlated than others can become more correlated after varying ρ . For example, when $\rho_c = 0.49$, the

correlation between Alabama and Florida is 0.1993 while the correlation between Alabama and Georgia is 0.1561. However, when $\rho_c = 0.97$, the correlation between Alabama and Florida is 0.6311, smaller than the correlation between Alabama and Georgia, which is equal to 0.6490. This seems odd as it means that the effect of changing ρ is not uniquely defined.

Consider the behavior of $\text{Cor}(i, j)$ when ρ approaches its lower bound -1.392. the pairwise correlations approach either -1 or $+1$. The latter limit value is counter-intuitive: some pairs tend to be perfectly positively correlated when we expect they to be the opposite of their neighboring values according to the SAR or CAR models.

Some results are reassuring. The correlations increase monotonically with ρ when the spatial dependence parameter is positive. However, the range of the correlations depends on the value of ρ . For instance, when $\rho_s = 0.1$, correlations between neighboring states vary between 0.026 and 0.115, while this variation lies between 0.241 and 0.642 when $\rho_s = 0.6$.

2.4 Some preliminary definitions and results

To explain the puzzling consequences, we use linear algebra and graph theory results.

2.4.1 Random graphs and the matrix \mathbf{W}

The \mathbf{W} matrix can be seen as the transition matrix of a Markov chain defined on a graph. Assume that n nodes or vertices, represented by the areas A_i , are connected by undirected edges such that there is an edge between areas i and j if $w_{ij} \neq 0$. Define a discrete-time and finite Markov chain with transition matrix given by \mathbf{W} . That is, if a particle is in one vertex i at time t , it moves to a different vertex in the next moment choosing among the neighbors of A_i with equal probability. These type of Markov models are called random walks on graphs (Brémaud (2001), page 214), or random graph, for short. \mathbf{W}^k is the transition matrix for the chain movements in k steps.

The random walk on the neighborhood graph converges to a unique stationary distribution if the Markov chain defined by \mathbf{W} is ergodic and aperiodic. For this, the graph must be connected, i.e., from each node there exists a path of edges connecting successive nodes until any other arbitrarily chosen node is reached. If \mathbf{W} is the row normalized adjacency matrix of an undirected graph \mathbf{G} , then the stationary distribution of the Markov chain defined by \mathbf{W} is given by $\boldsymbol{\pi} = (\pi_1, \dots, \pi_n)$ where $\pi_i = d_i/D$, where d_i is the number of neighboring areas of i and $D = \sum_i d_i$ (Brémaud (2001), page 214). π_i is called the density of area i or node i .

This implies that the power \mathbf{W}^k converge to a matrix composed by identical rows, all of them equal to the stationary distribution vector $\boldsymbol{\pi}$. That is, $\mathbf{W}_{ij}^k \rightarrow d_j/D$, as $k \rightarrow \infty$. The convergence to this stationary distribution is geometric, with relative speed proportional to the second-largest eigenvalue modulus. This result is known as the Perron-Fröbenius theorem (Brémaud (2001), page 157) and it is important for our development. It can be shown that the eigenvalue of \mathbf{W} with the largest modulus has multiplicity 1 and it is $\lambda = 1$. Let λ_2, \dots ,

λ_n be the other eigenvalues of \mathbf{W} ordered in a such a way that

$$\lambda_1 = 1 \geq |\lambda_2| > \dots \geq |\lambda_n|.$$

Let m_2 be the multiplicity of λ_2 and $\mathbf{1} = (1, \dots, 1)^t$. Then, the Perron-Fröbenius theorem proves that

$$\mathbf{W}^k = \mathbf{1} \boldsymbol{\pi}^t + O(k^{m_2-1} |\lambda_2|^k),$$

where k is a positive constant. In particular, if $|\lambda_2| > |\lambda_3|$ then $m_2 = 1$ and the convergence speed decays exponentially with the second largest eigenvalue modulus $|\lambda_2|$.

2.4.2 A matrix identity

There is a matrix identity which is fundamental to understanding the behavior of the correlations implied by the models and described in Section 2.2. If \mathbf{M} is a square matrix such that each entry of the matrix \mathbf{M}^k goes to zero as k increases, then the inverse $(\mathbf{I} - \mathbf{M})^{-1}$ exists and is given by

$$(\mathbf{I} - \mathbf{M})^{-1} = \mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \mathbf{M}^3 + \dots \quad (2.6)$$

(Iosifescu (1980), page 45). Take $\mathbf{M} = \rho \mathbf{W}$ where $|\rho| < 1$. Since $0 \leq [\mathbf{W}^k]_{ij} \leq 1$ for all i, j and for all integer k , we can write

$$(\mathbf{I} - \rho \mathbf{W})^{-1} = \mathbf{I} + \rho \mathbf{W} + \rho^2 \mathbf{W}^2 + \rho^3 \mathbf{W}^3 + \dots \quad (2.7)$$

2.4.3 The powers of the W matrix

If $[\mathbf{W}^k]_{ij} > 0$, then the probability of going from i to j in k steps in the random graph is positive. This means that there exists at least one sequence of k edges connecting nodes such that the initial and final nodes are i and j , respectively. Let us call such a path of a k -th order path between areas i and j . In fact, the value of $[\mathbf{W}^k]_{ij}$ is a weighted sum of all the k -th order paths between i and j . For example, $[\mathbf{W}^2]_{ij}$ is given by

$$[\mathbf{W}^2]_{ij} = \sum_{k=1}^n W_{ik} W_{kj} = \sum_{k=1}^n \frac{a_{ik}}{d_i} \frac{a_{kj}}{d_k} = \frac{1}{d_i} \sum_{k=1}^n \frac{a_{ik} a_{kj}}{d_k}. \quad (2.8)$$

The binary product $a_{ik} a_{kj}$ is equal to 1 only if k connects both i and j . Therefore, $[\mathbf{W}^2]_{ij}$ is proportional to a weighted sum of all second-order paths $i \rightarrow k \rightarrow j$. Each path contributes a fraction inversely proportional to the number d_k of neighbors the intervening area k has. The more connected k is, the smaller the contribution of the path $i \rightarrow k \rightarrow j$ to $[\mathbf{W}^2]_{ij}$. Note that $[\mathbf{W}^2]_{ii} > 0$ because there is at least one path of the type $i \rightarrow k \rightarrow i$ since each area has at least one neighbor.

Similarly, $[\mathbf{W}^3]_{ij}$ is given by

$$[\mathbf{W}^3]_{ij} = \sum_{l=1}^n [\mathbf{W}^2]_{il} w_{lj} = \frac{1}{d_i} \sum_{l=1}^n \sum_{k=1}^n \frac{a_{ik} a_{kl} a_{lj}}{d_k d_l}. \quad (2.9)$$

Each path $i \rightarrow k \rightarrow l \rightarrow j$ is inversely weighted by how dense is the neighborhood graph at k and l . Note that paths such as $i \rightarrow j \rightarrow i \rightarrow j$ are also counted.

2.5 Revisiting the puzzling results

Putting together the results of Section 2.4, for $|\rho| < 1$, we can write

$$[(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij} = [\mathbf{I}]_{ij} + \rho[\mathbf{W}]_{ij} + \rho^2[\mathbf{W}^2]_{ij} + \rho^3[\mathbf{W}^3]_{ij} + \dots \quad (2.10)$$

As long as $|\rho| < 1$, the correlation between i and j is a convergent series decaying exponentially in the spatial dependence parameter ρ . The k -th coefficient in this series is given by $[\mathbf{W}^k]_{ij}$, which is proportional to the weighted sum of paths of size k between i and j . The weight of a path is the product of the number of neighbors of the intervening areas in the path.

Since the k -th coefficient $[\mathbf{W}^k]_{ij}$ can be interpreted as a probability, it lies between 0 and 1. Furthermore, $[\mathbf{W}^k]_{ij}$ approaches the limit d_j/D for all i and the speed of this convergence, for all i and j , is determined by the second largest eigenvalue modulus of \mathbf{W} . This means that, with a good approximation and for some value k , we can write

$$[(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij} \approx [\mathbf{I}]_{ij} + \rho[\mathbf{W}]_{ij} + \dots + \rho^{k-1}[\mathbf{W}^{k-1}]_{ij} + \frac{d_j\rho^k}{D(1-\rho)} \quad (2.11)$$

$$\approx [\mathbf{I}]_{ij} + \rho[\mathbf{W}]_{ij} + \dots + \rho^{k-1}[\mathbf{W}^{k-1}]_{ij} \quad (2.12)$$

With these facts, the results are less puzzling and easier to understand. Basically, when we naively try to understand the covariance structure focusing only on the first-order neighborhood structure, we are doomed from the start. For instance, if the third degree approximation in (2.12) suffices, we have the CAR model covariance between areas i and j , $i \sim j$, given approximately by

$$\frac{v^2}{d_j} \left(\frac{\rho a_{ij}}{d_i} + \frac{\rho^2}{d_i} \sum_{k=1}^n \frac{a_{ik}a_{kj}}{d_k} + \frac{\rho^3}{d_i} \sum_{l=1}^n \sum_{k=1}^n \frac{a_{ik}a_{kl}a_{lj}}{d_l d_k} \right)$$

Ignoring the neighborhood structure geographically more distant than the first order will produce a crude approximation to the true correlation coefficient. Giving due consideration to the longer paths from i to j , though with ever decreasing weight, we find the results described by Wall (2004) to be much less puzzling, as we discuss next.

2.5.1 The CAR model with $\rho_c > 0$

First, let us consider the CAR model and $\rho_c > 0$. Then, (2.10) shows that the correlation must increase monotonically with ρ_c , since all the coefficients in that series expansion are nonnegative. This is one of the empirical results from Wall (2004). Although it is what one

expects intuitively, now we understand the underlying reason for this monotone increase of $\text{Cor}(i, j)$.

However, correlations of different pairs can increase at different rates. This is because the series expansion coefficients in (2.10) are pair-specific. In fact, the derivative of $[(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij}$ is equal to

$$\frac{\partial}{\partial \rho} [(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij} = [\mathbf{W}]_{ij} + 2\rho[\mathbf{W}^2]_{ij} + 3\rho^2[\mathbf{W}^3]_{ij} + \dots \quad (2.13)$$

This implies that, for $\rho \in (0, 1)$, we have an increasing derivative with ρ . If ρ is not too close to 1, the rate of increase of this derivative depends mostly on the second-order neighborhood $[\mathbf{W}^2]_{ij}$.

Different pairs can exchange their relative positions as $\rho_c > 0$ increases and it is clear now why and when this happens. The derivative on (2.13) depends of the specific pair i, j under consideration. For example, assuming that the second degree polynomial approximation in (2.12) is good enough, then

$$\frac{\partial}{\partial \rho} [(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij} \approx [\mathbf{W}]_{ij} + 2\rho[\mathbf{W}^2]_{ij} \quad (2.14)$$

Therefore, the larger ρ_c , the greater the positive contribution of the second-order neighborhoods. Hence, when ρ_c is small, a pair (i, j) can have a small correlation that may increase faster than the correlation in other areas simply because its second order coefficient $[\mathbf{W}^2]_{ij}$ is relatively large.

Tabela 2.1: Values of the entries of \mathbf{W}^k , $\rho^k\mathbf{W}^k$, and the cumulative sum $\sum_{j=0}^k \rho^j\mathbf{W}^j$ for the pairs of neighboring states (Alabama, Florida) and (Alabama, Geórgia). We consider the values $\rho = 0.97$ and $\rho = 0.49$.

k	1	2	3	4	5	10	30	50	100
Alabama and Florida, $\rho_c = 0.97$									
\mathbf{W}^k	0.2500	0.0500	0.0984	0.0498	0.0588	0.0317	0.0127	0.0100	0.0094
$\rho^k\mathbf{W}^k$	0.2425	0.0470	0.0898	0.0441	0.0505	0.0233	0.0051	0.0022	0.0004
<i>Cum.Sum</i>	0.2425	0.2895	0.3794	0.4235	0.4740	0.6246	0.8345	0.8997	0.9526
Alabama and Georgia, $\rho_c = 0.97$									
\mathbf{W}^k	0.2500	0.1562	0.1516	0.1333	0.1179	0.0754	0.0312	0.0249	0.0234
$\rho^k\mathbf{W}^k$	0.2425	0.1470	0.1383	0.1180	0.1012	0.0556	0.0125	0.0054	0.0011
<i>Cum.Sum</i>	0.2425	0.3895	0.5278	0.6458	0.7470	1.1026	1.6092	1.7711	1.9030
Alabama and Florida, $\rho_c = 0.49$									
$\rho^k\mathbf{W}^k$	0.1225	0.0120	0.0116	0.0029	0.0017	0.0000	0.0000	0.0000	0.0000
<i>Cum.Sum</i>	0.1225	0.1345	0.1461	0.1490	0.1506	0.1517	0.1517	0.1517	0.1517
Alabama and Georgia, $\rho_c = 0.49$									
$\rho^k\mathbf{W}^k$	0.1225	0.0375	0.0178	0.0077	0.0033	0.0001	0.0000	0.0000	0.0000
<i>Cum.Sum</i>	0.1225	0.1600	0.1778	0.1855	0.1889	0.1915	0.1915	0.1915	0.1915

This is the explanation for the apparently strange behavior of the switching ranks between the correlations of Alabama and Florida and Alabama and Georgia. We use Table 2.1 to illustrate our arguments focusing on the CAR model with $\rho_c = 0.97$ and $\rho_c = 0.49$. For

Alabama and Florida,

$$[(\mathbf{I} - \rho\mathbf{W})^{-1}]_{\text{Al, Fl}} \approx 0.25\rho + 0.05\rho^2 + 0.10\rho^3 + 0.05\rho^4 + \dots$$

while, for Alabama and Georgia, we have

$$[(\mathbf{I} - \rho\mathbf{W})^{-1}]_{\text{Al, Ge}} \approx 0.25\rho + 0.16\rho^2 + 0.15\rho^3 + 0.13\rho^4 + \dots$$

The coefficients of this expansion has a slower decline for the more fully connected pair (Alabama, Georgia) than for the pair (Alabama, Florida). When $\rho = 0.49$, this difference is not relevant because the diminishing ρ^k quickly shrinks the term $\rho^k[\mathbf{W}^k]_{ij}$ towards zero for both pairs. The consequence is that the first few terms, with small k , dominate the series. Considering only the first order approximation with k , we are within 64% and 81% of their limiting values, equal to 0.1915 for the pair (Alabama, Georgia), and equal to 0.1517 for the pair (Alabama, Florida), respectively. Using a third degree approximation with $k = 3$, we get very close to these limits, within 93% and 96%, respectively.

This picture changes substantially when $\rho = 0.97$. Now, even relatively large k -th order neighborhoods contribute a fair amount to the series sum. As a consequence, the convergence of $[(\mathbf{I} - \rho\mathbf{W})^{-1}]$ is slow. With $k = 1$, we are within only 13% and 25% from their limiting values, equal to 1.9030 for the pair (Alabama, Georgia), and equal to 0.9526 for the pair (Alabama, Florida), respectively. Increasing to $k = 10$ we are still away from the limiting values, 58% from (Alabama, Georgia), and 66% from (Alabama, Florida). This means that more geographically distant neighborhood structures, reflected in the k steps paths from i to j in the \mathbf{W}^k entries, have a non-negligible impact on the series' limits. Since these paths are different for the two pairs of areas, the end result is that an initial ordering of correlations when $\rho = 0.49$ is switched as ρ increases to 0.97.

Let us turn our attention to the relationship between variances $\text{Var}(y_i)$ and the number d_i of first order neighbors. Wall (2004) noticed that there is a typical negative relationship between these two quantities but that there is also variation of $\text{Var}(y_i)$ among areas with equal d_i . We use again the approximation in (2.12) to clarify this in the case of the CAR model.

Suppose that the \mathbf{W}^k converge fast enough such that

$$\begin{aligned} \text{Var}(y_i) &= \frac{\sigma_c^2}{d_i} [(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ii} \\ &\approx \frac{\sigma_c^2}{d_i} \left(1 + \rho[\mathbf{W}]_{ii} + \rho^2[\mathbf{W}^2]_{ii} + \frac{d_i\rho^3}{D(1-\rho)} \right) \\ &= \frac{\sigma_c^2}{d_i} \left(1 + \rho^2[\mathbf{W}^2]_{ii} \right) + \frac{\sigma_c^2\rho^3}{D(1-\rho)} \\ &\approx \frac{\sigma_c^2}{d_i} \left(1 + \frac{\rho^2}{d_i} \sum_k \frac{a_{ik}a_{ki}}{d_k} \right) \end{aligned}$$

where, in the last approximation, we ignored the last term and used (2.12). Therefore, the

declining value of $\text{Var}(y_i)$ with d_i is obvious but we also need to recognize the effect of the second (and higher) neighborhood order. The sum $\sum_k(a_{ik}a_{ki})/d_k$ depends on its number of terms. That is, it depends on the number d_i of first order neighbors $k \sim i$. It also depends on the connectedness degree of these neighbors through their d_k values.

To illustrate with an extreme case, suppose that area i has a single neighbor, area k . Then

$$\text{Var}(y_i) \approx \sigma_c \left(1 + \frac{\rho^2}{d_k} \right)$$

Two areas in this same single-neighbor situation have different variances if their single neighbors have different number of neighbors. The more connected is the single neighbor k , the smaller the variance of i .

2.5.2 The CAR model with $\rho_c < 0$

Concerning the negative pairwise correlations, again the spatial dependence parameter ρ_c and the higher order neighboring areas are crucial to understand their behavior. For $-1 < \rho_c < 0$, the terms in the series (2.10) alternate signs and this explains the counter intuitive behavior of some pairs of areas. If ρ is close to its lower bound -1 , the decay ρ^k is slow and more distant neighborhood patterns impact on the correlation value with alternating signs. The first term $\rho[\mathbf{W}]_{i,j}$ in the covariance expansion (2.10) is obviously negative. However, since $[\mathbf{W}^k]_{i,j}$ is not a monotone decreasing function of k , it is possible that the sum of the first two brings the covariance closer to zero or even positive. This happens if an increase in $[\mathbf{W}^2]_{i,j}$ with respect to $[\mathbf{W}]_{i,j}$ more than compensates the decrease from $|\rho|$ to ρ^2 . This argument is valid with higher order of k .

As an example, consider Vermont and Massachusetts. When $\rho_c = -0.99999$, the correlation between these two areas is equal to -0.1051 . The convergence of $[(\mathbf{I} - \rho\mathbf{W})^{-1}]_{ij}$ for this pair is very slow. Table 2 shows the values \mathbf{W}^k , $\rho^k\mathbf{W}^k$, and the cumulative sum $\sum_{j=0}^k \rho^j\mathbf{W}^j$ for Vermont and Massachusetts. We can see that the cumulative sum alternates widely. The difference between $k = 100$ and $k = 101$ for the cumulative sum is in the second devimal place, a substantial value for such a large order k .

Tabela 2.2: Values of the entries of \mathbf{W}^k , $\rho^k\mathbf{W}^k$, and the cumulative sum $\sum_{j=0}^k \rho^j\mathbf{W}^j$ for the pair Vermont and Massachusetts. We consider $\rho = -0.99999$.

k	1	2	3	4	5	10	100	101
$\rho^k\mathbf{W}^k$	-0.3300	0.1778	-0.1948	0.2061	-0.1729	0.1590	0.0315	-0.0313
<i>CumSum</i>	-0.3300	-0.1556	-0.3504	-0.1442	-0.3172	-0.1151	-0.1671	-0.1984

All pairs of neighboring areas have negative correlation in the CAR model when $-1 < \rho_c < 0$. However, in the SAR model with $\rho_s = -0.99999$, Vermont and Massachusetts has correlation equal to 0.0293 . We discuss the SAR model in more detail in section 4.3 but it is appropriate to anticipate some its results here. Similarly to the CAR model, using the power expansion of $[(\mathbf{I} - \rho\mathbf{W})^{-1}]$ for the SAR covariance in (2.2), we can express $\text{Cor}(i, j)$ as a power

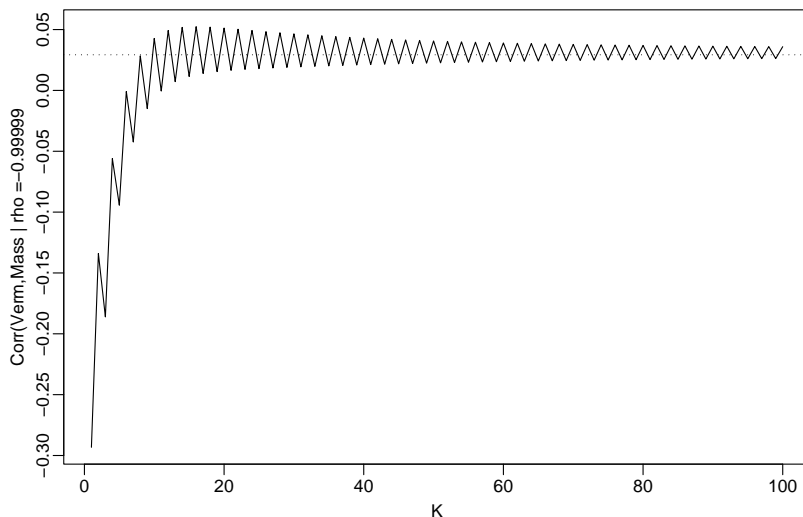


Figure 2.2: Successively approximating the correlation between Vermont and Massachusetts as we increase the number of terms in the finite sums of (??), with $\rho_s = -0.99999$.

series in ρ_s . Figure 2.2 shows the approximation as successively larger finite sums are used to approximate the eventually positive correlation. Considering only the first neighborhood orders, the approximation is negative.

The behavior for $\rho \leq -1$ is less simple to explain with our results. The series expansion (2.10) is no longer valid and our interpretations can not be put into use. When an extremely negative spatial parameter is used in the US states graph, the pairwise correlations approach either to -1 or to $+1$. In Figure ?? we draw the edges according to the limiting behavior of the pairwise correlation as ρ approaches its lower bound $-1.3923 = \min_i \{\lambda_i\}^{-1}$. A virtually identical figure is obtained for the SAR model. Solid lines are used for those pairs in which the correlation approach -1 while the dashed lines represent the pairs with limiting correlation approaching $+1$. It is not clear what the pattern means but we present a conjecture. It seems as if areas which act as the center of star shaped local neighborhoods have their connecting edges mostly positive. See, for example, Idaho, Colorado, and South Dakota. The edges composing the outer rings of these star-shaped local neighborhoods have negative correlations.

To consider an intuitive explanation for this conjecture, imagine that we are going to assign the value $+1$ to approximately one half of the areas and -1 to the remaining areas. In this way we keep the global mean close to zero. Let B the number of neighboring edges connecting areas with different values. If we want to maximize B , it seems that assigning $+1$ to the center of star shaped areas and -1 to the areas in the outer rings may be near optimal. We are investigating the truth of our conjecture at the moment.

2.5.3 The SAR model

The arguments for the SAR model are very similar to those presented for the CAR model but the formulas are more convoluted. Using the power series expansion (2.10) in the SAR

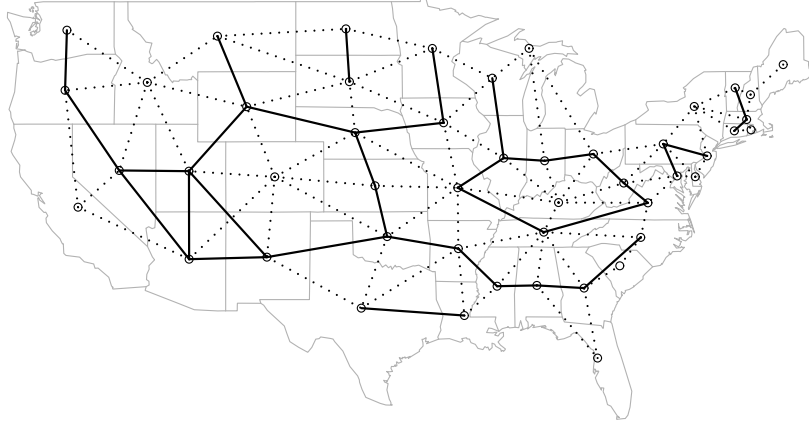


Figura 2.3: Edges of US states neighborhood graph drawn according to the pairwise correlation when ρ_c approaches its lower bound -1.3923 . Solid line: positive correlation, $+1$; dashed line: negative correlation, -1 .

covariance (2.2), we can write

$$\begin{aligned}\Sigma_s &= (I - \rho\mathbf{W})^{-1}\Lambda((I - \rho\mathbf{W})^{-1})' \\ &= (I + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3 + \dots)\Lambda(I + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3 + \dots)' \\ &= \sum_{n=0}^{\infty} \left[\rho^n \sum_{k=0}^n \mathbf{W}^k \Lambda(\mathbf{W}^{n-k})' \right]\end{aligned}$$

which has elements given by

$$(\Sigma_s)_{ij} = \frac{\sigma_s}{d_j} \sum_{n=0}^{\infty} \left[\rho^n \sum_{k=0}^n (\mathbf{W}^k)_{ij} (\mathbf{W}^{n-k})_{ji} \right]. \quad (2.15)$$

If the third degree approximation for $(I - \rho\mathbf{W})^{-1}$ suffices then

$$(\Sigma_s)_{ij} = \sum_{k=1}^n (I - \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3)_{ik} \frac{\sigma_s^2}{d_k} (I - \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3)_{jk} \quad (2.16)$$

where the element $(I - \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \rho^3\mathbf{W}^3)_{ik}$ is equal to

$$(I_{\{i=k\}} - \rho \frac{a_{ik}}{d_i} + \frac{\rho^2}{d_i} \sum_{p=1}^n n \frac{a_{ip}a_{pk}}{d_p} + \frac{\rho^3}{d_i} \sum_{p=1}^n \sum_{q=1}^n \frac{a_{ip}a_{pq}a_{qk}}{d_p d_q})_{ik} \quad (2.17)$$

The main difference between SAR and CAR is that a third degree approximation for $(I - \rho\mathbf{W})^{-1}$ imply in up to a sixth degree polynomial in ρ_s for each entry of the covariance matrix, the coefficients involving elements of \mathbf{W} . Therefore, we get the same type of polynomial approximation as for the CAR model and our qualitative conclusions remains unchanged for the SAR model. Incidentally, note that this higher approximating polynomial degree in the

SAR model as compared to the CAR model explains why, bottom plots in Figure 2.1, the first order neighbor correlations increase at a slower rates as a function of positive ρ_c in the CAR model compared with positive ρ_s in the SAR model. For a given approximating polynomial in ρ for $(\mathbf{I} - \rho\mathbf{W})^{-1}$, the implied SAR correlation polynomial has more positive terms than the corresponding CAR polynomial, as can be seen in (2.5.3), for example.

2.5.4 The role of $|\lambda_2|$

The second largest eigenvalue modulus $|\lambda_2|$ is in the interval $[0, 1)$ and it is responsible for the speed at which $[\mathbf{W}^k]_{ij}$ converges to its limiting value, $\pi_i = d_j / \sum_i d_i$. That is, the smaller $|\lambda_2|$, the smaller the degree k required in the approximation (2.12). Regular graphs are those with d_i constant. For a highly irregular neighborhood graph it is difficult to obtain exact results analytically. However, on regular graphs, these results are available and they highlight the interplay between the neighborhood structure and the approximation speed (Chung (1997), Chapter 1). Basically, the more connected the graph is, the larger the value of $|\lambda_2|$. Hence, $|\lambda_2|$ is a measure of overall connectedness of a graph.

In order to illustrate these points, we computed $|\lambda_2|$ for some regular graphs. Consider a ring graph with nodes $\{(u, u + 1) : 1 \leq u < n^2\} \cup \{(1, n^2)\}$. Then, $|\lambda_2| = \cos(2\pi/n^2) \approx 1$ if n is large (Chung, 1997, page 6). This decreases substantially when we pass to a grid graph with n^2 vertices symmetrically wrapped into a torus. In this case, each vertex has four neighboring vertices and $|\lambda_2| = (1 + \cos(2\pi/n))/2$, the midpoint between 1 and $\cos(2\pi/n) < \cos(2\pi/n^2)$ for $n \geq 2$. Finally, consider the most dense graph possible with n^2 vertices, the complete graph in which every area is a neighbor of every other area. Then, $|\lambda_2| = 1/(n^2 - 1) \approx 0$, if n^2 is large.

Admittedly, these graphs are highly artificial and do not represent the typical maps found in practice. To have a better idea of the effect of the average density of connections on $|\lambda_2|$, and hence on the speed of the convergence $[\mathbf{W}^k]_{ij} \rightarrow d_j/D$, we successively pruned a real map while keeping the entire graph connected. The objective is to show how the value of $|\lambda_2|$ tends towards 1 as we prune the graph.

The usual US states map is not the best choice for this demonstration. The reason is that $|\lambda_2| = 0.9714$ for this graph, a large initial value. This large value indicates that there are parts of the map (such as the NE region) that are hard to reach in a random walk, implying in long paths or a nearly disconnected graph. Even more regularly connected graphs have large eigenvalues. Therefore, in addition to pruning the usual adjacency neighborhood graph, we also added edges between second-order neighbors. That is, we increased the density of connections in the graph by adding edges between areas that are separated from each other by at most a third area.

We randomly selected an edge to be deleted while this was possible until only $n - 1$ edges remained (that is, until we reached a spanning tree). We also randomly created edges between second-order neighbors. To keep the balance on the two directions, we added edges until we reached the same number needed to generate the most pruned graph. We repeated this procedure one hundred times independently.

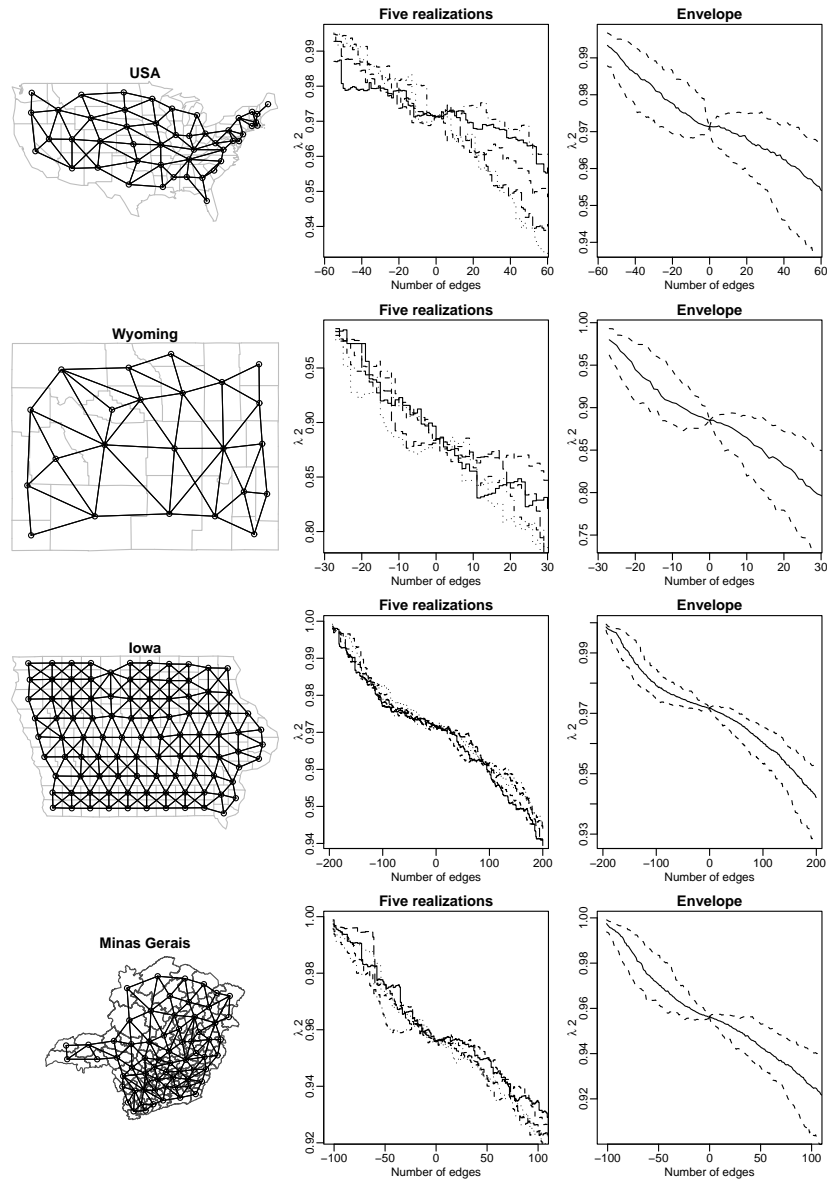


Figure 2.4: Successively adding or pruning the adjacency neighborhood graph of four graphs. The geographical regions are the US states map, the the counties of Wyoming and Iowa, and the municipalities of Minas Gerais, a Brazilian state. The second column of plots shows five realizations of the of the addition-pruning process in each graph. The third column of plots shows 95% confidence envelopes based on the simulations in dashed lines, as well as the mean $|\lambda_2(j)|$ value in solid line.

Figure 2.4 shows the graph of the second largest eigenvalue modulus $|\lambda_2(j)|$ where j is either the number of deleted edges from the original map (if $j < 0$) or the number of added edges (if $j > 0$). We used four geographical regions, shown in the first column of plots: the US states map, the counties of Wyoming and Iowa, and the municipalities of Minas Gerais, a Brazilian state with the approximately extension as France. Their $|\lambda_2(0)|$ values are 0.9714, 0.8850, 0.9717, and 0.9561, respectively. The second column of plots shows five realizations

of the addition-pruning process in each graph. Each line is the value of $|\lambda_2(j)|$ as j varies. The third column of plots shows in dashed lines 95% confidence envelopes based on the 100 simulations, as well as the mean $|\lambda_2(j)|$ value as a solid line.

Specific paths within the confidence envelope are not necessarily monotone. That is, the deletion (or addition) of a specific edge can decrease (or increase) the eigenvalue of the resulting \mathbf{W} matrix. However, the average behavior is that the denser the connections, the smaller the eigenvalue and hence, faster the convergence. In terms of the puzzling results discussed in Wall (2004), this means that the denser the graph, the less likely the change of ranks between different pairs of areas.

Capítulo 3

Covariância posteriori de efeitos aleatórios com priori CAR

Neste capítulo nós consideramos um modelo espacial Bayesiano para dados gaussianos. Nesse modelo, consideramos um efeito aleatório espacial com distribuição a priori CAR. Nós mostramos os resultados não intuitivos de Wall, Wall (2004), na covariância a posteriori do efeito aleatório espacial. Nós encontramos uma expressão simples para a matriz de covariância a posteriori que auxilia na compreensão da estrutura estocástica do efeito aleatório a posteriori. Essa expressão é também obtida se a distribuição a priori é CAR intrínseca.

3.1 Introdução

Muitos estudos envolvem dados de áreas geográficas, seja um país dividido em estados, um estado dividido em municípios, um município dividido em bairros ou qualquer região subdividida de forma discreta. Nessas subdivisões, são observados fenômenos que produzem medidas discretas ou contínuas. Um fenômeno pode ser observado mais de uma vez em cada área, caracterizando processos espaço-temporais. Também podem ser observados vários fenômenos ao mesmo tempo, caracterizando processos multivariados. Esses dados podem ser casos de doença agregados por municípios, o índice de desenvolvimento humano em cada estado de um país, etc.

Os modelos de regressão podem ser estendidos para analisar esses dados espaciais. As extensões são necessárias para modelar a estrutura de dependência espacial dos dados. Essa dependência espacial pode ser modelada diretamente utilizando modelos de regressão espacial para dados gaussianos, na abordagem frequentista Anselin et al. (2004), ou considerando efeitos aleatórios espaciais para dados gaussianos ou não, na abordagem bayesiana, Banerjee et al. (2004).

A abordagem bayesiana tem sido a mais utilizada, principalmente devido ao fato de ser mais facilmente estendida a modelos mais complexos do que a abordagem frequentista. Nessa abordagem, a componente espacial é geralmente modelada considerando uma distribuição a priori autogressiva condicional ou CAR, do inglês *Conditional AutoRegressive*, para os

efeitos aleatórios espaciais desconhecidos. Nós estudaremos o modelo CAR, Besag (1974), em detalhes na próxima seção. Geralmente o uso da distribuição CAR como priori requer o uso de técnicas de Monte Carlo via Cadeia de Markov, ou do inglês *Monte Carlo via Markov Chain - MCMC*, para obter amostras da distribuição à posteriori. Apenas em casos particulares podemos encontrar distribuições a posteriori conhecidas e isso não é necessário. Atualmente vários modelos considerando prioris CAR foram implementados e estão disponíveis em programas computacionais populares, tais como WinBUGS, Lunn et al. (2000), e BayesX, Lang (2000).

Devido à facilidade de simular amostras da distribuição a posteriori, os modelos espaciais bayesianos foram estendidos a modelos espaço-temporais, modelos multivariados e modelos de sobrevivência espaciais, Banerjee et al. (2004); modelos com parâmetros variando no espaço Assunção (2003); e modelos aditivos generalizados, Fahrmeir and Lang (2001). No entanto, existe pouco conhecimento analítico sobre a estrutura estocástica da distribuição a posteriori dos efeitos espaciais desses modelos. Em geral, atribui-se distribuições a priori para os parâmetros, simula-se da distribuição a posteriori, mas falta um entendimento mais detalhado e analítico das propriedades da distribuição a posteriori dos parâmetros espaciais.

Em modelos bem simples, tem sido apontados alguns resultados não intuitivos da estrutura estocástica implicada pelo modelo CAR. A covariância entre pares de áreas com mesmo número de vizinhos são diferentes, Besag and Kooperberg (1995). Em contexto, não bayesiano considerando o modelo CAR como verossimilhança para os dados, Wall (2004) mostrou alguns resultados não intuitivos na estrutura de covariância em função do parâmetro de auto-regressão espacial. Esses resultados foram estudados novamente e explicações para eles foram obtidas através do estudo dos autovalores da matriz de que reflete a estrutura de vizinhança, Assunção et al. (2008).

Neste artigo, nós fazemos um estudo detalhado de um modelo bayesiano espacial gaussiano e obtemos uma expressão para a estrutura de covariância a posteriori dos parâmetros modelados por uma priori CAR que permite entender melhor a sua estrutura estocástica. Na próxima Seção nós introduzimos os modelos bayesianos espaciais, na Seção 3.3 obtemos a expressão da covariância a posteriori. Na Seção 3.4 nós mostramos os resultados não intuitivos apontados em Wall (2004) para a covariância a posteriori. Na Seção 3.5 nós obtemos alguns resultados algébricos e na Seção 3.6 analisamos detalhadamente a matriz de covariância a posteriori.

3.2 Modelos espaciais Bayesianos

Nesta seção nós vamos considerar que temos observações y_i , $i = 1, \dots, n$, feitas em n áreas geográficas. Podemos considerar também um vetor de covariáveis \mathbf{X}_i observadas em cada área. Um modelo que pode ser considerado para esses dados é o modelo linear generalizado misto, ou *Generalized Linear Mixed Model - GLMM*. Esse modelo é utilizado para modelar a esperança da variável de interesse de forma linear através de uma função de ligação:

$$\mu_i = E(y_i) = g(\eta_i), \text{ onde, } \eta_i = \mathbf{X}_i\boldsymbol{\beta} + b_i$$

onde, μ_i é a esperança da média, $g(\cdot)$ é uma função da média da variável chamada de função de ligação, η_i é o preditor linear para a média da variável na área i , β são parâmetros de regressão, \mathbf{X}_i é a linha i da matriz $n \times p$ com o vetor $\mathbf{1}$ e as $p - 1$ covariáveis e b_i é o efeito aleatório da área i . Na forma vetorial temos

$$\boldsymbol{\mu} = E(\mathbf{y}) = g(\boldsymbol{\eta}), \text{ onde, } \boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{b}.$$

A priori CAR é usada para modelar \mathbf{b} , o vetor de efeito aleatório.

O modelo CAR para o vetor \mathbf{b} , é definido por autoregressões de b_i em seus vizinhos. Assim,

$$b_i | \mathbf{b}_{-i} \sim N \left(\sum_{j \sim i} \frac{\rho b_j}{d_i}, \frac{1}{d_i \tau_b} \right),$$

em que \mathbf{b}_{-i} são todos os elementos de \mathbf{b} exceto o elemento b_i , $j \sim i$ indica que a área j é vizinha da área i , d_i é o número de vizinhos da área i , ρ é um parâmetro de autocorrelação e mede a força da dependência de b_i nos seus vizinhos e $\tau_b > 0$ é o parâmetro de precisão. Esse modelo faz com que \mathbf{b} tenha uma variação suave no espaço.

É possível obter a distribuição conjunta de \mathbf{b} , Banerjee et al. (2004). Assumindo que $E(b_i) = 0$, a distribuição de \mathbf{b} é normal multivariada

$$\mathbf{b} \sim MVN(0, (\mathbf{I} - \rho \mathbf{W})^{-1} \mathbf{T}^{-1}),$$

em que \mathbf{T} é uma matriz diagonal com $T_{ii} = \tau_b d_i$ e \mathbf{W} é definida a partir da matriz de adjacência. A matriz de adjacência $\mathbf{A} = (a_{ij})$ de dimensão $n \times n$ é definida fazendo $a_{ij} = 1$ se $i \sim j$ e $a_{ij} = 0$ caso contrário ($a_{ii} = 0$). A matriz $\mathbf{W} = (w_{ij})$ é tal que $w_{ij} = a_{ij}/a_i$ onde $a_i = \sum_j a_{ij} = d_i$ é o número de áreas vizinhas da área i . É necessário fazer $\rho \in (\lambda_n^{-1}, \lambda_1^{-1})$ em que $\lambda_1, \dots, \lambda_n$ são os autovalores de \mathbf{W} ordenados em ordem decrescente, ou $\rho \in (\lambda_n^{-1}, 1)$, pois $\lambda_1 = 1$.

Nós vamos considerar que, y_1, \dots, y_n são observações feitas nas áreas $1, \dots, n$ de um processo gaussiano condicionalmente independente com média μ_i e variância τ_y^{-1} . Neste caso temos que o vetor n dimensional $\mathbf{y} | \mathbf{b}, \tau_y \sim MVN(\mathbf{b}, \frac{1}{\tau_y} \mathbf{I})$. Com \mathbf{y} gaussiano, $g(\cdot)$ costuma ser a identidade e $E(y_i) = \mu_i = \eta_i = \mathbf{X}_i \boldsymbol{\beta}$, pois $E(b_i) = 0$.

Esse modelo foi estendido em várias direções e para obter uma intuição da estrutura a posteriori desses modelos, nós vamos considerar \mathbf{y} gaussiano e estudar detalhadamente sua distribuição à posteriori. Como veremos, muito pode ser aprendido sobre as consequências da adoção de uma distribuição a priori CAR para o efeito aleatório \mathbf{b} neste caso. No caso de assumir \mathbf{y} gaussiano, podemos encontrar distribuições conhecidas para a distribuição a posteriori. Porém, para outras distribuições, tal como a de Poisson, mesmo nos casos mais simples a distribuição a posteriori não possui forma conhecida e os métodos MCMC são utilizados para fazer inferência.

3.3 Covariância à posteriori

Nós queremos obter a distribuição a posteriori de \mathbf{b} considerando uma distribuição a priori CAR para \mathbf{b} :

$$\mathbf{b}|\tau_b, \rho \propto \exp\left\{-\frac{1}{2}\mathbf{b}'[\mathbf{T}(\mathbf{I} - \rho\mathbf{W})]\mathbf{b}\right\}.$$

com ρ , \mathbf{T} e \mathbf{W} como definidos anteriormente. Para $\boldsymbol{\beta}$ nós adotamos uma distribuição a priori Normal p variada:

$$\boldsymbol{\beta} \propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{m})\mathbf{V}^{-1}(\boldsymbol{\beta} - \mathbf{m})\right\},$$

com \mathbf{m} o vetor de médias e \mathbf{V} a matriz de covariância. Para τ_y nós consideramos uma distribuição a priori $Gamma(\alpha, \beta)$, $\alpha > 0$ e $\beta > 0$:

$$\tau_y|\alpha, \beta \propto \tau_y^{\alpha-1} \exp(-\tau_y\beta).$$

A verossimilhança para os dados é o produto de distribuições normais:

$$y \propto \tau_y^{-n/2} \exp\left\{-\frac{\tau_y}{2}[\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b})]'[\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b})]\right\}.$$

Assumindo independência entre as prioris, temos então que

$$\begin{aligned} \boldsymbol{\beta}, \tau_y, \mathbf{b}|\mathbf{y}, \tau_b, \mathbf{m}, \mathbf{V}, \alpha, \beta &\propto \exp\left\{-\frac{1}{2}(\boldsymbol{\beta} - \mathbf{m})\mathbf{V}^{-1}(\boldsymbol{\beta} - \mathbf{m})\right\} \times \tau_y^{\alpha-1} \exp(-\tau_y\beta) \\ &\times \exp\left\{-\frac{1}{2}\mathbf{b}'[\mathbf{T}(\mathbf{I} - \rho\mathbf{W})]\mathbf{b}\right\} \\ &\times \tau_y^{n/2} \exp\left\{-\frac{\tau_y}{2}(\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))'(\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))\right\}. \end{aligned}$$

Podemos obter as distribuições condicionais para cada um dos parâmetros ($\boldsymbol{\beta}$, τ_y e \mathbf{b}).

Temos que $\boldsymbol{\beta}|\mathbf{y}, \mathbf{b}, \tau_y, \mathbf{m}, \mathbf{V}$

$$\begin{aligned} &\propto \exp\left\{-\frac{1}{2}[(\boldsymbol{\beta} - \mathbf{m})'\mathbf{V}^{-1}(\boldsymbol{\beta} - \mathbf{m}) + \tau_y(\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))'(\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))]\right\} \\ &\propto \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}'\mathbf{V}^{-1}\boldsymbol{\beta} - 2\boldsymbol{\beta}'\mathbf{V}^{-1}\mathbf{m} - 2\tau_y(\mathbf{y} - \mathbf{b})'(\mathbf{X}\boldsymbol{\beta}) + \tau_y(\mathbf{X}\boldsymbol{\beta})'(\mathbf{X}\boldsymbol{\beta})]\right\} \\ &= \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}'\mathbf{V}^{-1}\boldsymbol{\beta} - 2\mathbf{m}'\mathbf{V}^{-1}\boldsymbol{\beta} - 2\tau_y[\mathbf{X}'(\mathbf{y} - \mathbf{b})]'\boldsymbol{\beta} + \tau_y\boldsymbol{\beta}'(\mathbf{X}'\mathbf{X})\boldsymbol{\beta}]\right\} \\ &= \exp\left\{-\frac{1}{2}[\boldsymbol{\beta}'(\mathbf{V}^{-1} + \tau_y\mathbf{X}'\mathbf{X})\boldsymbol{\beta} - 2[\mathbf{m}'\mathbf{V}^{-1} + \tau_y(\mathbf{X}'(\mathbf{y} - \mathbf{b}))']\boldsymbol{\beta}]\right\} \end{aligned}$$

Vamos considerar um resultado de algebra linear, a *multivariate completion of squares* ou *ellipsoidal retification* onde

$$\mathbf{u}'\mathbf{A}\mathbf{u} - 2\boldsymbol{\alpha}'\mathbf{u} = (\mathbf{u} - \mathbf{A}^{-1}\boldsymbol{\alpha})'\mathbf{A}(\mathbf{u} - \mathbf{A}^{-1}\boldsymbol{\alpha}) - \boldsymbol{\alpha}'\mathbf{A}^{-1}\boldsymbol{\alpha}.$$

Portanto

$$\boldsymbol{\beta}|\mathbf{y}, \mathbf{b}, \tau_y, \mathbf{m}, \mathbf{V} \propto \exp \left\{ -\frac{1}{2} [(\boldsymbol{\beta} - \mathbf{m}^*)' (\mathbf{V}^{-1} + \tau_y \mathbf{X}' \mathbf{X}) (\boldsymbol{\beta} - \mathbf{m}^*)] \right\}$$

onde $\mathbf{m}^* = (\mathbf{V}^{-1} + \tau_y \mathbf{X}' \mathbf{X})^{-1} [\mathbf{V}^{-1} \mathbf{m} + \tau_y \mathbf{X}' (\mathbf{y} - \mathbf{b})]$ e $\boldsymbol{\beta}|\mathbf{y}, \mathbf{b}, \tau_y, \mathbf{m}, \mathbf{V}$ tem distribuição normal- p variada com média \mathbf{m}^* e variância $(\mathbf{V}^{-1} + \tau_y \mathbf{X}' \mathbf{X})^{-1}$.

A distribuição condicional de τ_y é obtida de

$$\tau_y|\mathbf{y}, \boldsymbol{\beta}, \mathbf{b}, \tau_b, \rho \propto \tau_y^{n/2+\alpha-1} \exp \left\{ -\tau_y \left[\beta + \frac{1}{2} (\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))' (\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b})) \right] \right\},$$

ou seja, que $\tau_y|\mathbf{y}, \boldsymbol{\beta}, \mathbf{b}, \alpha, \beta$ tem distribuição $Gamma(\alpha + n/2, \beta + \frac{1}{2} (\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))' (\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b})))$.

A distribuição a posteriori de \mathbf{b} condicionada aos demais parâmetros é obtida de

$$\begin{aligned} \mathbf{b}|\mathbf{y}, \boldsymbol{\beta}, \tau_y, \tau_b, \rho &\propto \exp \left\{ -\frac{1}{2} [\mathbf{b}' [\mathbf{T}(\mathbf{I} - \rho \mathbf{W})] \mathbf{b} + \tau_y (\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))' (\mathbf{y} - (\mathbf{X}\boldsymbol{\beta} + \mathbf{b}))] \right\} \\ &\propto \exp \left\{ -\frac{1}{2} [\mathbf{b}' [\mathbf{T}(\mathbf{I} - \rho \mathbf{W})] \mathbf{b} - 2\tau_y (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{b} + \mathbf{b}' (\tau_y \mathbf{I}) \mathbf{b}] \right\} \\ &= \exp \left\{ -\frac{1}{2} [\mathbf{b}' [\mathbf{T}(\mathbf{I} - \rho \mathbf{W}) + \tau_y \mathbf{I}] \mathbf{b} - 2\tau_y (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' \mathbf{b}] \right\} \end{aligned}$$

Completando quadrados e considerando $\mathbf{C} = [\tau_y \mathbf{I} + \mathbf{T}(\mathbf{I} - \rho \mathbf{W})]$ temos

$$\mathbf{b}|\mathbf{y}, \boldsymbol{\beta}, \tau_y, \tau_b, \rho \propto \exp \left\{ -\frac{1}{2} [(\mathbf{b} - \tau_y \mathbf{C}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))' \mathbf{C} (\mathbf{b} - \tau_y \mathbf{C}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}))] \right\}$$

Portanto, temos que

$$\mathbf{b}|\mathbf{y}, \boldsymbol{\beta}, \tau_y, \tau_b, \rho \sim MVN \left(\tau_y \mathbf{C}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}), \mathbf{C}^{-1} \right). \quad (3.1)$$

Se consideramos o caso particular em que não temos covariáveis no modelo e $E(y_i) = 0$, nós podemos utilizar um resultado geral de inferência bayesiana para dados normais multivariados. Neste caso, condicionando a τ_b e ρ conhecidos, temos que a distribuição marginal de $\boldsymbol{\mu}$ a posteriori, integrando sobre τ_y , é t -multivariada com 2α graus de liberdade, vetor de locação $\boldsymbol{\mu}$ e matriz de precisão $(\alpha/\beta)[\mathbf{T}(\mathbf{I} - \rho \mathbf{W})]$, DeGroot (1970). Neste caso particular, não é necessário utilizar MCMC para simular amostras da distribuição a posteriori de \mathbf{b} .

Na prática, é comum atribuir distribuições a priori para os hiperparâmetros $\mathbf{m}, \mathbf{V}, \tau_b$ e ρ , ou seja, utilizar hiperprioris para considerar a incerteza sobre esses hiperparâmetros. Porém, não é possível obter distribuições condicionais com forma conhecida para eles. Como temos condicionais completas para os demais parâmetros com forma conhecida, podemos utilizar o procedimento *Gibbs Sampler* com passos de Metropolis-Hastings para esses parâmetros.

3.4 Resultados não-intuitivos de Wall

Wall (2004) apontou três resultados não intuitivos implicados pelo modelo CAR. Esses resultados foram mostrados usando o mapa dos 48 estados continentais dos Estados Unidos da América. Nós também tomamos esse mapa para mostrar os resultados não intuitivos de Wall no modelo espacial Bayesiano gaussiano. Neste caso, nós temos que avaliar os resultados em função de ρ considerando diferentes valores de τ_y e τ_b . Nós usamos uma sequência de ρ no intervalo $(-1.392387, 1)$, o limite válido para esse mapa e consideramos valores para τ_y e τ_b , condicionando que a soma de ambos seja igual a 1. Usamos os valores $\{0, 0.01, 0.1, 0.3, 0.5, 0.7, 0.9, 0.99, 1\}$ para τ_b ($\tau_y = 1 - \tau_b$). Portanto, nós estudaremos a influência da distribuição a priori na ocorrência dos resultados não intuitivos.

Na Figura 3.1 mostramos a correlação a posteriori entre \mathbf{b} vizinhos. Considerando $\tau_b = 1$, no gráfico superior esquerdo, o resultado é a correlação implicada pelo modelo CAR, a correlação a priori. Observamos nos outros gráficos que os três resultados não intuitivos de Wall ocorrem. A amplitude da correlação decresce se ρ decresce em módulo, ocorrem troca de postos (linhas se cruzam) e ocorrem correlações positivas para valores negativos de ρ . Porém, a medida que τ_b decresce em relação a τ_y , o resultado mais intrigante, que é a ocorrência de correlação positiva quando $\rho < 0$ entre alguns pares de vizinhos, não ocorre. Para uma análise mais detalhada do que ocorre com os resultados não intuitivos, nós calculamos algumas estatísticas sobre esses resultados.

Nós contamos o número de pares com correlação positiva implicada por ρ negativo. Usamos ρ próximo do limite inferior e variamos τ_b entre 0.70 e 1. Esses resultados são mostrados na Figura 3.2. Se $\tau_b = 0.7$ não há correlação positiva. Uma correlação positiva ocorre com $\tau_b = 0.7281$. O número de pares aumenta rapidamente quando τ_b é próximo de 1, lembrando que no modelo CAR, equivalente a fazer $\tau_b = 1$ neste caso, ocorreram 34 pares de vizinhos com correlação positiva. Podemos considerar então que na prática é muito raro ocorrer um par de correlações negativas se $\rho < 0$, pois para isso ρ deve ser próximo do limite inferior e τ_b muito grande em relação a τ_y .

O segundo resultado não intuitivo é a mudança de ordem no rank das correlações entre pares de vizinhos se ρ muda. Para avaliar a influência de τ_b nesse resultado, nós calculamos o coeficiente de correlação de Kendall, ϱ . Nós consideramos os valores de τ_b da análise anterior. Para cada um desses valores, consideramos $\rho = 0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95$ e calculamos os postos da correlação entre os pares de vizinhos. A seguir, calculamos ϱ entre os vetores de postos obtidos por esses valores de ρ e os postos obtidos por $\rho = 0.05$, um valor de referência. Para valores negativos de ρ , nós usamos $\rho = -0.05, -0.15, -0.3, -0.5, -0.7, -1, -1.35$ e calculamos ϱ com $\rho = -0.05$ de referência. Para ambos os casos, $\rho < 0$ e $\rho > 0$ nós consideramos também ϱ entre os postos implicados por valores sequenciais de ρ .

Na Figura 3.3, nós mostramos os valores de ϱ calculados. Cada linha corresponde a um par de valores de ρ e a linha é em função de τ_b . Na Figura 3.3 (esquerda) consideramos valores negativos de ρ e $\rho = 0.05$ como referência. Na Figura 3.3 (direita) valores positivos e $\rho = 0.05$ como referência. Em ambos os gráficos se $\tau_b = 0$, $\varrho = 1$ para todos os pares de

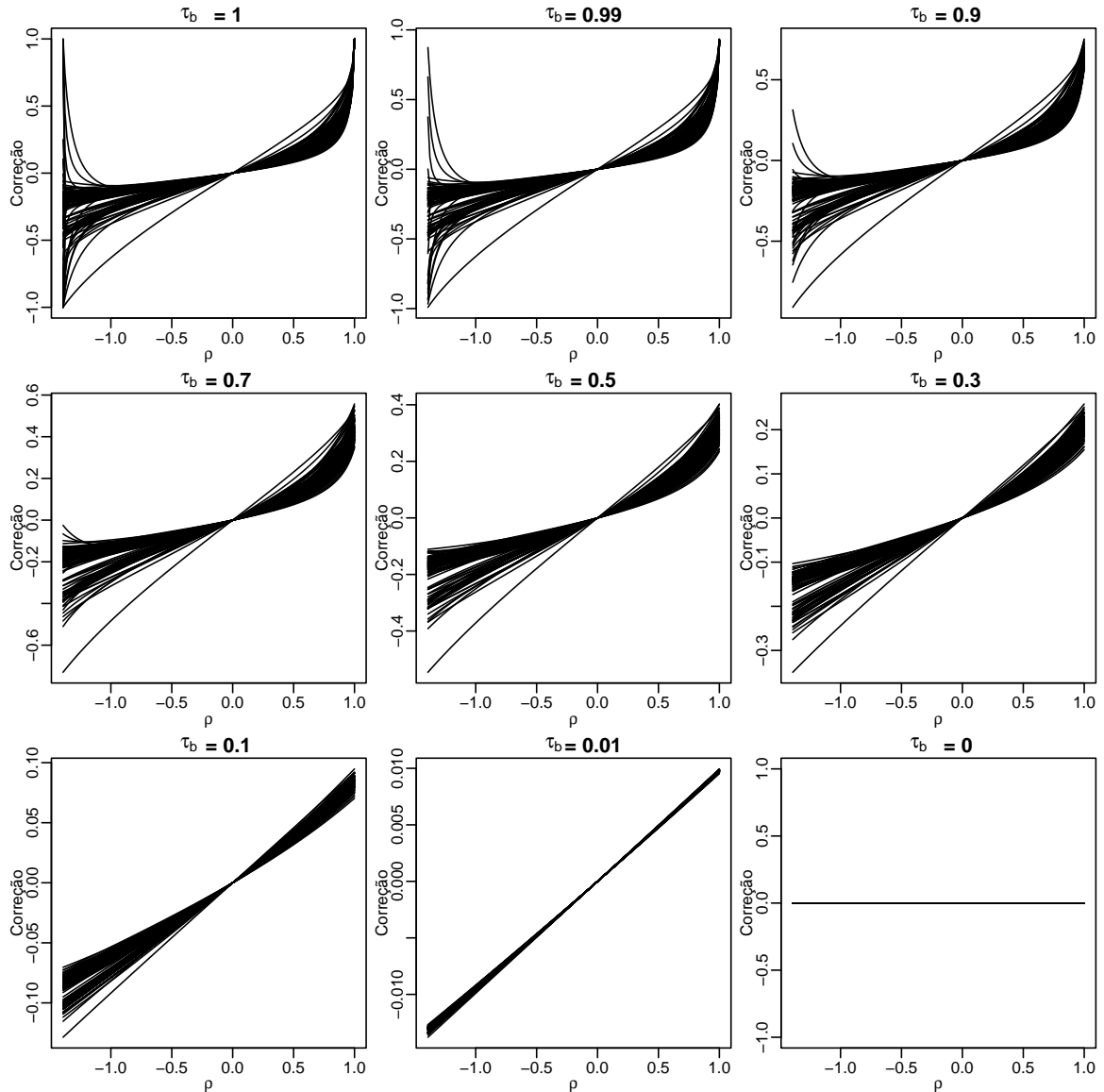


Figura 3.1: Correlação a posteriori entre vizinhos por ρ , para vários valores de τ_b .

valores de ρ , pois nesse caso a correlação à posteriori entre os pares de vizinhos não depende da estrutura do grafo. Se τ_b aumenta, ϱ decresce rapidamente e estabiliza num platô. Isso indica que um pouco de informação a priori já é suficiente para que haja variação nos postos. Notamos também que para valores de ρ mais distantes do valor de referência, o platô é mais baixo, o que é um resultado razoável. No caso dos pares de $\rho > 0$, há pouca mudança nos postos se ρ muda de 0.05 para 0.15. Notamos também que a linha de ϱ quando ρ muda de 0.05 para 0.25 é parecido com a linha de ϱ quando ρ muda de -0.05 para -0.15 .

Notamos também alguns resultados confusos. Um resultado confuso é que as curvas não são paralelas e cruzam em alguns casos. No caso de $\rho > 0$, por exemplo, o menor valor de ϱ é obtido com $\rho = 0.75$ se $\tau_b = 1$. Outro é que olhando a curva referente a $\rho = 0.95$, para $\tau_b = 0.1$ o valor de ϱ é o menor valor, mas a medida que τ_b aumenta, ϱ continua aproximadamente

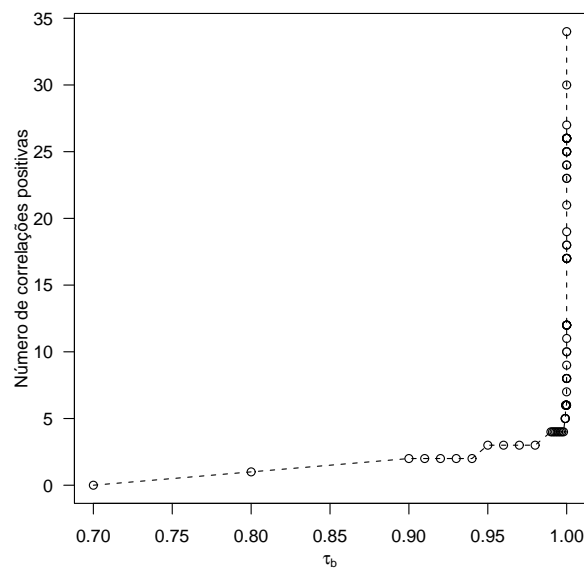


Figura 3.2: Número de correlações positivas entre pares de vizinhos em função de τ_b .

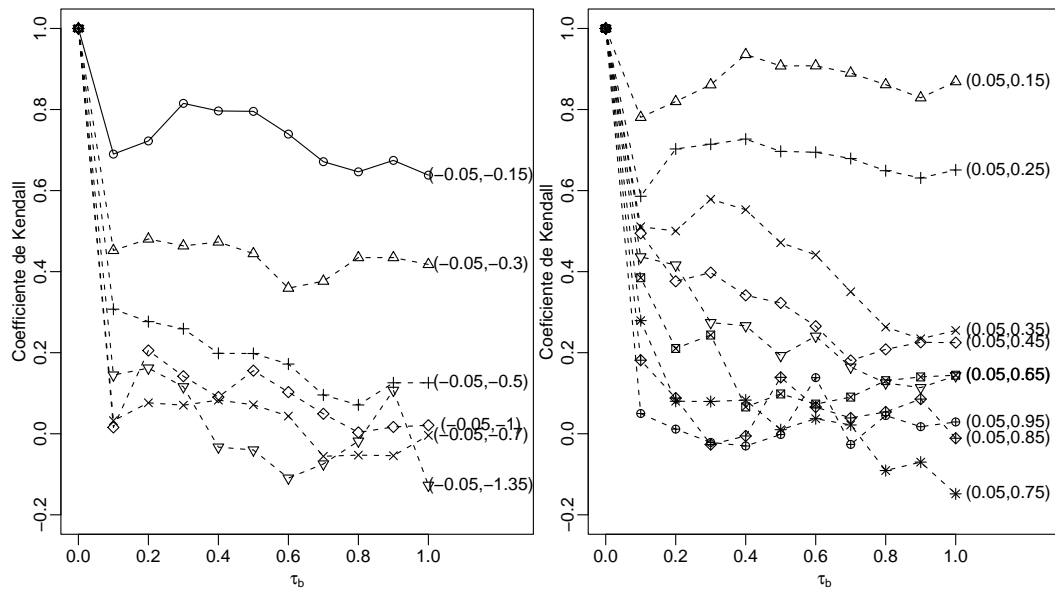


Figura 3.3: Coeficiente de correlação de Kendall entre os postos de correlação entre pares de vizinhos implicada por diferentes pares de valores de ρ , em função de τ_b .

constante, além disso, para $\tau_b = 1$ não é o menor valor.

Se a diferença entre os valores de ρ considerados cresce, podemos esperar valores menores de ϱ e esse resultado foi observado na maioria dos casos. Porém nós podemos estar interessados no caso em que os valores de ρ considerados são sequenciais com uma diferença aproximadamente constante entre ambos. Para verificar isso, nós calculamos ϱ entre os postos implicados por diferentes valores de ρ com diferença entre eles iguais a 0,1. Esses resultados são mostrados na Figura 3.4. Observando ambos os gráficos dessa Figura, notamos que em geral ϱ não decai tão rapidamente a medida que τ_b cresce, como na Figura 3.3. Também, notamos que em geral ϱ decai mais em função de τ_b se os valores de ρ são mais distantes de 0 (zero). Esse resultado indica que a mudança nos postos das correlações entre os pares tende a ser maior se ρ esta mais próximo dos extremos.

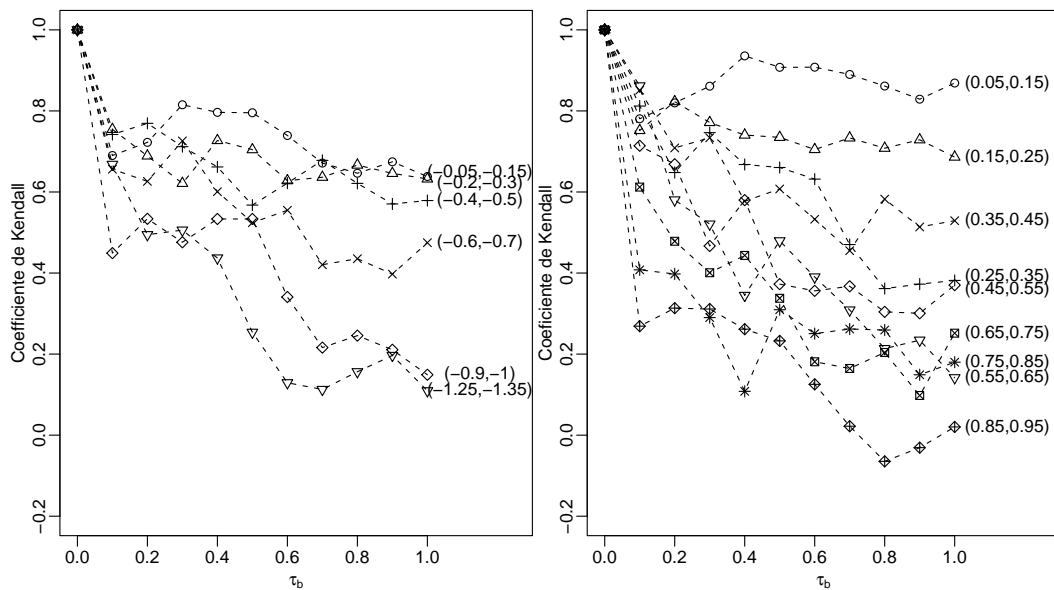


Figura 3.4: Coeficiente de correlação de Kendall entre os postos de correlação entre pares de vizinhos implicada por pares de valores sequenciais de ρ , em função de τ_b .

O terceiro resultado apontado em Wall (2004) é o aumento da amplitude das correlações marginais se $|\rho|$ cresce. Nós também olhamos a amplitude da correlação implicada à posteriori entre os pares de vizinhos em função de τ_b . Na Figura 3.5, obtemos a amplitude em função de τ_b para alguns valores de ρ . Esse resultado é muito claro, para um valor fixo de ρ , se τ_b cresce, a amplitude aumenta e se ρ decresce, a amplitude diminui.

3.5 Alguns resultados algébricos

Nesta seção nós estudamos a matriz de covariância de $\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho$

$$Cov(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho) = [\tau_y \mathbf{I} + \mathbf{T}(\mathbf{I} - \rho \mathbf{W})]^{-1}$$

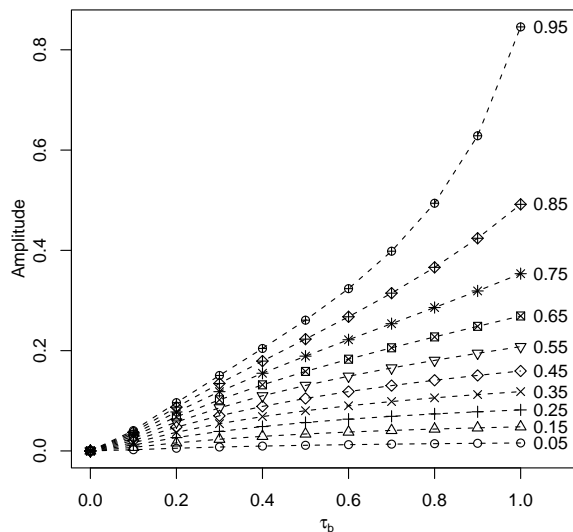


Figura 3.5: Amplitude por τ_b para alguns valores de ρ .

a partir de uma expressão mais simples obtida para essa matriz. Particularmente, queremos obter uma expressão para a covariância entre pares de áreas,

$$Cov(\mathbf{b}_i, \mathbf{b}_j | \mathbf{y}, \tau_y, \tau_b, \rho) = \left\{ [\tau_y \mathbf{I} + \mathbf{T}(\mathbf{I} - \rho \mathbf{W})]^{-1} \right\}_{ij}.$$

O estudo de $Cov(\mathbf{b} | \mathbf{y}, \tau_y, \tau_b, \rho)$ com essa expressão mais simples auxilia na compreensão da estrutura estocástica de \mathbf{b} a posteriori.

Inicialmente, vamos reescrever essa matriz de covariância de forma conveniente:

$$\begin{aligned} Cov(\mathbf{b} | \mathbf{y}, \tau_y, \tau_b, \rho) &= [\tau_y \mathbf{I} + \mathbf{T}(\mathbf{I} - \rho \mathbf{W})]^{-1} \\ &= (\tau_y \mathbf{I} + \mathbf{T} - \rho \mathbf{T} \mathbf{W})^{-1} \\ Cov(\mathbf{b} | \mathbf{y}, \tau_y, \tau_b, \rho)(\tau_y \mathbf{I} + \mathbf{T}) &= [(\tau_y \mathbf{I} + \mathbf{T}) - (\rho \mathbf{T} \mathbf{W})]^{-1} (\tau_y \mathbf{I} + \mathbf{T}) \\ &= \left\{ (\tau_y \mathbf{I} + \mathbf{T})^{-1} [(\tau_y \mathbf{I} + \mathbf{T}) - (\rho \mathbf{T} \mathbf{W})] \right\}^{-1} \\ &= (\mathbf{I} - \rho \mathbf{W}^*)^{-1} \end{aligned}$$

em que

$$\mathbf{W}^* = \mathbf{T}^* \mathbf{W} \text{ e } \mathbf{T}^* = \text{diag} \left\{ \frac{d_1 \tau_b}{\tau_y + d_1 \tau_b}, \dots, \frac{d_n \tau_b}{\tau_y + d_n \tau_b} \right\}.$$

Logo, temos que

$$Cov(\mathbf{b} | \mathbf{y}, \tau_y, \tau_b, \rho) = (\mathbf{I} - \rho \mathbf{W}^*)^{-1} (\tau_y \mathbf{I} + \mathbf{T})^{-1} = [(\tau_y \mathbf{I} + \mathbf{T})(\mathbf{I} - \rho \mathbf{W}^*)]^{-1} \quad (3.2)$$

Olhando para \mathbf{T}_{ii}^* , notamos que esse termo é a precisão à priori de y_i em relação à precisão total, com uma ponderação em função do número de vizinhos da área i . Esse termo é então uma medida da influência da precisão à priori em relação a precisão total, que é a precisão da distribuição a priori para cada área somada à precisão dos dados. Na seção 3.4 nós estudamos

os resultados não intuitivos de Wall em função de ρ considerando a influência da precisão a priori, utilizando o mapa dos Estados Unidos. Note que, se $\tau_y = 0$ a covariância à posteriori é exatamente a covariância à priori.

Nós podemos obter uma expressão mais simples e que fornece uma interpretação intuitiva para a covariância a posteriori de pares $\mathbf{b}_i, \mathbf{b}_j$. Para isso, usamos um resultado básico de álgebra linear. Esse resultado diz que se \mathbf{M} é uma matriz quadrada e cada entrada da matriz \mathbf{M}^k tende a zero se k aumenta, então a inversa $(\mathbf{I} - \mathbf{M})^{-1}$ existe e é dada por

$$(\mathbf{I} - \mathbf{M})^{-1} = \mathbf{I} + \mathbf{M} + \mathbf{M}^2 + \mathbf{M}^3 + \dots, \quad (3.3)$$

Iosifescu (1980), pag. 45.

Nós precisamos verificar se os elementos de $(\rho \mathbf{W}^*)^k$ tendem a zero se k aumenta. Acerca de \mathbf{W} , sabemos que é uma matriz estocástica, então $\sum_{j=1}^n \mathbf{W}_{ij}^k = 1$, $i = 1, \dots, n$ e $k > 0$ e da teoria de grafos temos que $\mathbf{W}_{ij}^k \rightarrow d_i/D$ se $k \rightarrow \infty$, em que $D = \sum_{r=1}^n d_r$. Sobre \mathbf{W}^* , temos que $\sum_{j=1}^n \mathbf{W}_{ij}^* = \mathbf{T}_{ii}^* = d_i \tau_b / (\tau_y + d_i \tau_b) \leq 1$. Portanto

$$[(\mathbf{W}^*)^k]_{ij} \rightarrow 0 \text{ se } k \rightarrow \infty. \quad (3.4)$$

Considerando $\rho \mathbf{W}^*$, seja $r = \max(|\rho \mathbf{T}^*|)$ para ρ tal que $|\rho \mathbf{T}^*| \in (0, 1)$. Então, para ρ neste caso, temos

$$[(|\rho \mathbf{W}^*|)^k]_{ij} < r^k [(\mathbf{W}^*)^k]_{ij} \rightarrow r^k \frac{d_i}{D} \rightarrow 0, \text{ se } k \rightarrow \infty.$$

Logo

$$(\rho \mathbf{W}^*)^k \rightarrow 0 \text{ se } \rho \mathbf{T}_{ii} \in (-1, 1), \text{ para } i = 1, \dots, n. \quad (3.5)$$

Este resultado nos permite utilizar o resultado (3.3) para um intervalo maior de valores de ρ , incluindo valores menores que -1 .

Nós podemos utilizar a interpretação de potências da matriz de adjacência \mathbf{A} para compreender melhor a estrutura de covariância à posteriori. O elemento ij de \mathbf{A}^k é o número de caminhos distintos para ir da área i para a área j em k passos. O elemento ij de \mathbf{W}^k é a probabilidade de sair da área i passear aleatoriamente pelo grafo e estar na área j em k passos. O elemento $[\mathbf{W}^k]_{ij}$ é a razão entre o número de caminhos de i para j em k passos e o produto entre número total de caminhos possíveis em cada passo. No limite essa probabilidade é d_i/D , onde $D = \sum_{i=1}^n d_i$, que é a densidade do nó. Então, quanto mais densa é a área, ou maior o seu número de vizinhos, maior é a probabilidade de se passar por essa área durante o passeio aleatório.

Inicialmente vamos encontrar os elementos de $(\mathbf{W}^*)^k$, para $k = \{0, 1, 2 \text{ e } 3\}$. Para $k = 0$, $(\mathbf{W}^*)^0 = \mathbf{I}$ por definição. O elemento ij para $k = 1$ é

$$(\mathbf{W}^*)_{ij} = \frac{w_{ij} d_i \tau_b}{\tau_y + d_i \tau_b} = \frac{a_{ij} \tau_b}{\tau_y + d_i \tau_b},$$

para $k = 2$ é

$$[(\mathbf{W}^*)^2]_{ij} = \sum_{r=0}^n \frac{a_{ir}\tau_b a_{rj}\tau_b}{(\tau_y + d_i\tau_b)(\tau_y + d_r\tau_b)} = \frac{\tau_b^2}{\tau_y + d_i\tau_b} \sum_{r=1}^n \frac{a_{ir}a_{rj}}{\tau_y + d_r\tau_b}$$

e para $k = 3$ é

$$[(\mathbf{W}^*)^3]_{ij} = \frac{\tau_b^3}{\tau_y + d_i\tau_b} \sum_{r=1}^n \sum_{s=1}^n \frac{a_{ir}a_{rs}a_{sj}}{(\tau_y + d_r\tau_b)(\tau_y + d_s\tau_b)}.$$

Notamos que de forma geral, $(\mathbf{W}^*)^k$ é o número de caminhos de tamanho k de i para j dividido por uma ponderação do produto do número total de caminhos em cada passo, ou seja,

$$[(\mathbf{W}^*)^k]_{ij} = \frac{\tau_b^k}{\tau_y + d_i\tau_b} \times \frac{\# \text{ caminhos tamanho } k \text{ de } i \text{ para } j}{\prod (\tau_y + \# \text{ alternativas a cada passo} \times \tau_b)}.$$

O termo de ponderação é $d_m\tau_b/(\tau_y + d_m\tau_b)$ que é a importância da precisão a priori $d_m\tau_b$ em relação a precisão total, com d_m sendo o número de caminhos possíveis de serem seguidos em cada passo. No caso particular em que $\tau_y = 0$, $\mathbf{W}^* = \mathbf{W}$.

3.6 Analisando a covariância a posteriori

Nós queremos estudar a distribuição a posteriori de \mathbf{b} . Então, precisamos estudar a matriz $\mathbf{C}^{-1} = [\tau_y\mathbf{I} + \mathbf{T}(\mathbf{I} - \rho\mathbf{W})]^{-1}$ para entender a estrutura estocástica de $\mathbf{b}|\mathbf{y}, \boldsymbol{\beta}, \tau_y, \tau_b, \rho$. Wall (2004) estudou a matriz $[\mathbf{T}(\mathbf{I} - \rho\mathbf{W})]^{-1}$, a matriz de covariância do modelo CAR, que é a distribuição a priori de \mathbf{b} . Nesse trabalho, foram apontados três resultados não intuitivos em função de ρ : 1) a ocorrência de correlação positiva entre pares de vizinhos quando $\rho < 0$; 2) troca de postos entre as correlações entre dois pares vizinhos diferentes quando ρ muda; 3) mudança na amplitude das correlações entre pares de vizinhos se ρ muda. Como veremos na seção 3.4, esses resultados também acontecem no modelo bayesiano adotado se τ_b é grande o suficiente em relação a τ_y .

3.6.1 Uma expressão simples

Considerando os resultados da seção 3.5, se $\rho\mathbf{T}_{ii}^* \in (-1, 1)$, $i = \{1, \dots, n\}$, podemos escrever

$$\text{Cov}(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho) = [\mathbf{I} + \rho\mathbf{W}^* + \rho^2[\mathbf{W}^*]^2 + \rho^3[\mathbf{W}^*]^3 + \dots](\tau_y\mathbf{I} + \mathbf{T})^{-1}. \quad (3.6)$$

A matriz \mathbf{W} não é simétrica e $(\rho\mathbf{W}^*)^k$ também não é e o termo $(\tau_y\mathbf{I} + \mathbf{T})$ faz com que $\text{Cov}(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho)$ seja simétrica. Para um par de áreas i, j temos

$$\text{Cov}(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho) = \frac{1}{\tau_y + d_j\tau_b} \sum_{k=0}^{\infty} \rho^k [(\mathbf{W}^*)^k]_{ij}, \quad (3.7)$$

Em que o termo $(\tau_y + d_j\tau_b)^{-1}$ faz com que $\text{Cov}(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho) = \text{Cov}(\mu_j, \mu_i|\mathbf{y}, \tau_y, \tau_b, \rho)$.

Como casos particulares temos: Se $\tau_y = 0$,

$$Cov(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho) = [\mathbf{I} + \rho\mathbf{W} + \rho^2\mathbf{W}^2 + \dots]\mathbf{T}^{-1}$$

e

$$Cov(b_i, b_j|\mathbf{y}, \tau_y, \tau_b, \rho) = (d_j\tau_b)^{-1} \sum_{k=0}^{\infty} \rho^k (\mathbf{W}^k)_{ij}.$$

Se $\tau_b = 0$ ou $\rho = 0$, $Cov(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho) = \tau_y\mathbf{I}$ e $Cov(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho) = 0$. Se fazemos τ_b tender a zero, a variância à posteriori tende a ser a variância dos dados, pois nesse caso não temos informação à priori. Se τ_y tende a zero, os dados não trazem muita informação sobre μ e a covariância a posteriori tende a ser igual a covariância à priori.

Nós podemos considerar aproximações para $Cov(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b, \rho)$, utilizando apenas os primeiros termos da expansão (3.7). Olhando para esses termos podemos buscar interpretações para $Cov(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho)$.

Considerando aproximação de ordem 1 para $Cov(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho)$ temos

$$Cov(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho) \approx \begin{cases} \frac{1}{\tau_y + d_i\tau_b} & \text{se } i = j \\ \frac{\rho\tau_b}{(\tau_y + d_j\tau_b)(\tau_y + d_i\tau_b)} & \text{se } i \sim j \\ 0 & \text{c.c.} \end{cases}. \quad (3.8)$$

Com esta aproximação, a variância da área i depende do número de vizinhos, e é o inverso da soma das precisões, à priori e dos dados. Se $i \sim j$, temos que a covariância é inversamente proporcional ao número de vizinhos de ambas as áreas. Então, um par de áreas com muitos vizinhos tem covariância menor que outro com poucos vizinhos com essa aproximação. Isso dá uma noção de que cada área deve “dividir” a sua “informação” com as áreas vizinhas.

Na aproximação de ordem 2, temos

$$Cov(\mu_i, \mu_j|\mathbf{y}, \tau_y, \tau_b, \rho) \approx \begin{cases} \frac{1}{\tau_y + d_i\tau_b} \left[1 + \frac{\rho^2\tau_b^2}{\tau_y + d_i\tau_b} \sum_{r=0}^n \frac{a_{ir}a_{rj}}{(\tau_y + d_r\tau_b)} \right] & \text{se } i = j \\ \frac{\rho\tau_b}{(\tau_y + d_i\tau_b)(\tau_y + d_j\tau_b)} \left[1 + \rho\tau_b \sum_{r=0}^n \frac{a_{ir}a_{rj}}{(\tau_y + d_r\tau_b)} \right] & \text{se } i \sim j \\ \frac{\rho^2\tau_b^2}{(\tau_y + d_j\tau_b)(\tau_y + d_i\tau_b)} \sum_{r=0}^n \frac{a_{ir}a_{rj}}{(\tau_y + d_r\tau_b)} & \text{c.c.} \end{cases}. \quad (3.9)$$

Notamos que neste caso a variância à posteriori da área i depende do seu número de vizinhos e do número de vizinhos de seus vizinhos. Notamos que o último termo, obtido de $(\rho\mathbf{W}^*)^2$, carrega a informação dos vizinhos comuns às áreas i e j . Quanto mais vizinhos comuns houverem em relação ao número de vizinhos de cada uma, maior é esse termo. Se, por exemplo, temos um mapa com $d_i + 1$ áreas e a área i tem d_i vizinhos, ou seja, a área i está ao centro das outras áreas e todas são vizinhos dela. O termo de ordem 2 para a variância nessa área é então $d_i[\rho\tau_b/(\tau_y + d_i\tau_b)]^2$.

A covariância entre um par de vizinhos depende do número de vizinhos comuns a ambas e do número de vizinhos de cada uma dessas áreas. Se um par de áreas não são vizinhos,

a covariância não é nula se ambas possuem algum vizinho comum, ou seja, há um caminho em dois passos para ir de uma área a outra. Esse termo será maior quanto maior o número de caminhos de tamanho 2 ligando ambas em relação ao número de vizinhos dos vizinhos de cada uma. Um extremo é obtido se, por exemplo, temos as áreas i e j e todas os vizinhos de i também são vizinhos de j , ou seja, ambas são extremos do mapa e as demais áreas estão entre ambas e são vizinhos de ambas.

A medida que aumentamos o grau da aproximação, a matriz de covariância tem menos termos nulos. O grau de aproximação necessário para preencher toda a matriz é o diâmetro do grafo de vizinhança. O diâmetro é o máximo do menor caminho que liga um par de áreas qualquer. A estrutura de covariância à posteriori depende não só dos vizinhos mas de toda a estrutura do grafo de vizinhança associado ao mapa. Porém, os caminhos de tamanhos muito grandes são menos importantes, pois $(\rho \mathbf{W}^*)_{ij}^k \rightarrow 0$ se $k \rightarrow \infty$. É importante então analisar a velocidade de convergência de $\rho \mathbf{W}^*$ para estudar a importância da vizinhança mais longinqua na covariância a posteriori.

3.6.2 O caso distribuição a priori CAR intrínscica

O modelo CAR intrínscico é obtido no caso particular de $\rho = 1$. A distribuição conjunta a priori de \mathbf{b} neste caso é imprópria. Porém, prova-se que a distribuição a posteriori de \mathbf{b} é própria. Então, analisar a distribuição de $\mathbf{b}|\mathbf{y}$ quando a priori é CAR intrínscica, é o caso em que analisamos apenas a influência da estrutura do mapa e das relações entre τ_y e τ_b .

Nesse caso, basta considerarmos a convergência de $[\mathbf{W}^*]^k$, demonstrada em (3.4). Portanto, podemos considerar que

$$\text{Cov}(\mathbf{b}|\mathbf{y}, \tau_y, \tau_b) = [\mathbf{I} + \mathbf{W}^* + [\mathbf{W}^*]^2 + [\mathbf{W}^*]^3 + \dots](\tau_y \mathbf{I} + \mathbf{T})^{-1}. \quad (3.10)$$

e para um par de áreas i, j , temos

$$\text{Cov}(\mathbf{b}_i, \mathbf{b}_j|\mathbf{y}, \tau_y, \tau_b) = \frac{1}{\tau_y + d_j \tau_b} \sum_{k=0}^{\infty} [(\mathbf{W}^*)^k]_{ij} \quad (3.11)$$

Assim podemos explicar o fato de que a covariância entre pares de áreas com mesmo número de vizinhos são diferentes, Besag and Kooperberg (1995). Como vemos em (3.11), a covariância entre um par de áreas não depende apenas dos vizinhos de primeira ordem, mas depende de vizinhos mais longínquos ou de caminhos de ordem k , com $k \rightarrow \infty$. Portanto, a covariância entre pares depende de toda a estrutura do grafo, embora os termos de vizinhança mais longínquos tenham menos importância, pois $\mathbf{W}^k \rightarrow 0$ se $k \rightarrow \infty$.

3.6.3 Convergência e segundo autovalor

A análise da velocidade da convergência de \mathbf{W}^* pode ser feita através da análise da razão entre seus segundo e primeiro autovalores. Para ter uma idéia da convergência e comparar com a convergência de \mathbf{W} , nós calculamos essa razão para \mathbf{W}^* e o segundo autovalor de

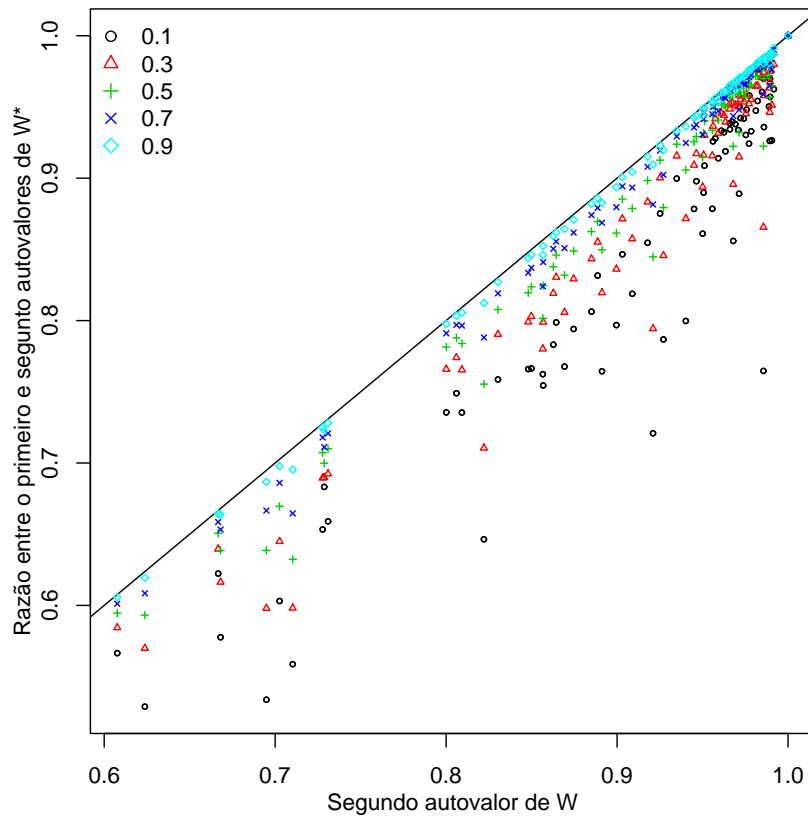


Figura 3.6: Segundo autovalor de \mathbf{W} (vertical) pela a razão entre os dois primeiros autovalores de \mathbf{W}^* (horizontal), considerando diferentes valores de τ_b .

\mathbf{W} , que é a razão entre seus dois primeiros autovalores, pois seu primeiro autovalor é 1. Nós consideramos 76 mapas, incluindo o mapa dos EUA, Brasil, estados dos EUA divididos em condados e estados do Brasil dividido em municípios. Na figura 3.6, temos o gráfico de dispersão entre essas razões.

Na figura 3.6, podemos observar o diagrama de dispersão do segundo autovalor de \mathbf{W} pela razão entre os dois primeiros autovalores de \mathbf{W}^* , considerando diferentes valores de τ_b e τ_y . Nós consideramos $\tau_b = \{0.1, 0.3, 0.5, 0.7, 0.9\}$. e $\tau_y = 1 - \tau_b$. Podemos observar que a razão entre o segundo e primeiro autovalores de \mathbf{W}^* é menor que o segundo autovalor de \mathbf{W} , para todos os 76 mapas considerados e todos os valores de τ_b , exceto aos referentes ao estado de Delaware que tem o segundo autovalor de \mathbf{W} igual a 1 e os dois primeiros autovalores de \mathbf{W}^* iguais.

Esse resultado indica que \mathbf{W}^* converge mais rápido que \mathbf{W} , além disso, $\mathbf{W}^* \rightarrow 0$. Isso implica que a vizinhança mais longinqua influencia menos na estrutura de covariância a posteriori que sua influência na priori, ou no modelo CAR. Notamos também que para valores menores de τ_b a razão é menor e, conseqüentemente a convergência é mais rápida. Esse resultado é bastante intuitivo, pois se a precisão a priori é menor em relação a soma das precisões, é de se esperar que a estrutura de covariância seja menos afetada pela estrutura do mapa.

Capítulo 4

Conclusão

A matriz de covariância dos modelos SAR e CAR e da distribuição a posteriori podem ser decompostas em uma soma infinita de termos. Essa decomposição é válida considerando restrição no parâmetro de correlação de forma a garantir a convergência dos termos da soma. A partir da interpretação desses termos, podemos entender melhor os resultados não intuitivos da estrutura de correlação implicada.

A correlação entre um par de áreas não depende apenas de seus vizinhos, mas sim da estrutura de vizinhança de todo o mapa. A contribuição de um vizinho mais longínquo é dada nos termos de maior grau da soma infinita e esses termos tem uma interpretação associada a caminhos entre as áreas passando por esse vizinho. A influência das áreas mais distantes está associada à velocidade com que os termos da soma infinita convergem. Também vimos que essa convergência está associada ao segundo autovalor da matriz de vizinhança associada e a contribuição de áreas mais distantes é menor em mapas mais densamente conectados, pois vimos que o segundo autovalor de \mathbf{W} diminui a medida que adiciona-se arestas.

Nós observamos também que os resultados não intuitivos são menos frequentes, ou em menor grau, no modelo Bayesiano, a medida que a informação dos dados seja mais relevante que a informação à priori. Isso porque os resultados não intuitivos são causados pela priori. Usando resultados de álgebra, conseguimos compreender a estrutura de covariância da distribuição a posteriori do efeito aleatório espacial.

A seguir colocamos três possíveis linhas de pesquisa para trabalhos futuros:

- A matriz \mathbf{W} que analisamos é a mais utilizada, porém outras definições também podem ser utilizadas. Seria interessante estudar outros tipos de matriz de vizinhança.
- A razão entre o segundo e o primeiro autovalor de \mathbf{W}^* foi apresentado para um conjunto de 76 mapas. Um trabalho futuro poderia ser provar se esse resultado é geral.
- Os resultados deste trabalho são úteis para compreender a estrutura a posteriori de parâmetros com priori CAR. Porém, é necessário verificar se em modelos mais complexos os resultados não intuitivos ocorrem e se podem ser analisados da forma como fizemos neste trabalho.

Referências Bibliográficas

- Anselin, L., Florax, R. J. G. M. and Rey, S. J. (2004). *Advances in Spatial Econometrics: Methodology, Tools and Applications*, Springer.
- Assunção, R. M. (2003). Space varying coefficient models for small area data, *Environmetrics* 14: 453–473.
- Assunção, R. M., Krainski, E. T. and Del Pino, G. (2008). Another close look at the spatial structure of car and sar models, *submetido* .
- Banerjee, S., Carlin, B. P. and Gelfand, A. E. (2004). *Hierarchical Modeling and Analysis for Spatial Data*, CRC Press.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems (with discussion), *Journal of the Royal Statistical Society, Series B* 36: 192–236.
- Besag, J. and Kooperberg, C. (1995). On conditional and intrinsic autoregressions, *Biometrika* 82: 733–746.
- Brémaud, P. (2001). *Markov Chains*, Springer-Verlag, New York.
- Chung, F. (1997). *Spectral Graph Theory*, Vol. 92 of *CBMS Regional Conference Series in Mathematics*, American Mathematical Society: Providence.
- DeGroot, M. H. (1970). *Optimal Statistical Decisions*, McGraw-Hill Book Company, New York.
- Diggle, P. J. and Ribeiro Jr., P. J. (2007). *Model Based Geostatistics*, Springer, New York. 230p.
- Fahrmeir, L. and Lang, S. (2001). Bayesian inference for generalized additive mixed models based on markov random field priors, *Applied Statistics - JRSS C* 50: 201–220.
- Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*, Cambridge University Press.
- Iosifescu, M. (1980). *Finite Markov Processes and Their Applications*, John Wiley and Sons, Chichester.

-
- Lang, B. (2000). BayesX - Software for Bayesian Inference Based on Markov Chain Monte Carlo Simulation Techniques.
- Lunn, D., Thomas, A., Best, N. and Spiegelhalter, D. (2000). WinBUGS – a Bayesian Modelling Framework: Concepts, Structure, and Extensibility, *Statistics and Computing* 10: 325–337.
- R Development Core Team (2007). *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.
URL: <http://www.R-project.org>
- Ribeiro Jr, P. J. and Diggle, P. J. (2001). geoR: a package for geostatistical analysis, *R-NEWS* 1(2): 14–18. ISSN 1609-3631.
URL: <http://CRAN.R-project.org/doc/Rnews/>
- Wall, M. M. (2004). A close look at the spatial structure implied by the car and sar models, *Journal of Statistical Planning and Inference* 121: 311–324.
- Whittle, P. (1954). On stationary process in the plane, *Biometrika* 41: 434–449.