Morbidade pulmonar e condições climáticas: um estudo na cidade de São Paulo

Thelma Sáfadi [1]; Airlane Pereira Alencar[2] e Pedro Alberto Morettin [3]

Resumo

O modelo fatorial dinâmico cujos fatores seguem um modelo auto-regressivo é utilizado para analisar a associação entre séries climáticas e séries de internação por problemas pulmonares. As séries consideradas foram taxa de internação por tuberculose, taxa de internação por problemas pulmonares, temperatura mínima, índice pluviométrico e umidade relativa mínima na cidade de São Paulo no período de Janeiro de 1998 a Setembro de 2009. Para os dados analisados, foram identificados dois fatores. O primeiro fator associa positivamente as taxas de internação por tuberculose e problemas pulmonares e negativamente com a umidade relativa mínima, o segundo fator associa a taxa de internação por problemas pulmonares com as séries de temperatura mínima, índices pluviométricos e umidade relativa mínima.

Key Words: Análise Bayesiana, modelo fatorial, morbidade por problemas pulmonares, séries climáticas

1 Introdução

Denomina-se doença pulmonar qualquer doença ou distúrbio que cause ou indique uma função pulmonar deficiente. Existem três categorias principais de doença pulmonar: doença pulmonar obstrutiva crônica em que existe uma redução do fluxo de ar, causada pelo estreitamento ou obstrução das vias aéreas, como asma, enfisema e bronquite crônica; doença pulmonar restritiva em que há uma diminuição na quantidade de ar inalada, porque existe uma redução na elasticidade ou quantidade de tecido pulmonar. A terceira categoria inclui as infecções.

A tuberculose é uma infecção causada por um microorganismo chamado Mycobacterium tuberculosis, também conhecido por bacilo de Koch. A doença costuma

[1]Universidade Federal de Lavras, MG
[2]Universidade de São Paulo, SP
[3]Universidade de São Paulo, SP

afetar os pulmões mas pode, também, ocorrer em outros órgãos do corpo, mesmo sem causar dano pulmonar.

A tuberculose é mais comum nas áreas onde há muita pobreza, promiscuidade, desnutrição, má condição de higiene e uma saúde pública deficitária. Os países com maior incidência da doença são a Índia, China, Indonésia, Bangladesh, Nigéria, Paquistão, Filipinas, Congo, Rússia e Brasil.

Geralmente, pega-se a doença pelo ar contaminado eliminado pelo indivíduo com a tuberculose nos pulmões. A pessoa sadia inala gotículas, dispersas no ar, de secreção respiratória do indivíduo doente. Este, ao tossir, espirrar ou falar, espalha no ambiente as gotículas contaminadas, que podem sobreviver, dispersas no ar, por horas, desde que não tenham contato com a luz solar. A pessoa sadia, respirando no ambiente contaminado, acaba inalando esta microbactéria que se implantará num local do pulmão. Em poucas semanas, uma pequena inflamação ocorrerá na zona de implantação. Não é ainda uma doença. É o primeiro contato do germe com o organismo (primoinfecção). Depois disso, esta bactéria pode se espalhar e se alojar em vários locais do corpo.

Neste artigo estamos interessados em explorar a associação entre as séries de taxas de internações hospitalares por tuberculose e por doenças pulmonares e as séries climáticas temperatura mínima, índice pluviométrico e umidade mínima na cidade de São Paulo durante o período de Janeiro de 1998 a Setembro de 2009. O modelo fatorial dinâmico cujos fatores seguem um modelo auto-regressivo é aplicado para examinar esta associação.

2 Modelo Fatorial Dinâmico e Inferência Bayesiana

O modelo fatorial cujos fatores seguem um modelo auto-regressivo é dado pela equação (1)

$$\begin{aligned} y_t &= \beta + Cf_t + e_t, \\ f_t &= \sum_{i=1}^{p} \rho_i f_{t-i} + w_t, \end{aligned} \tag{1}$$

onde y_t é um vetor $q \times 1$ de séries temporais observadas, β é um vetor $q \times 1$ de médias e C é uma matriz $q \times k$ de constantes desconhecidas, denominada matriz de cargas. O erro e_t é um vetor $q \times 1$ suposto independente normal com $e_t \sim N(0, \Gamma)$, onde Γ é uma matriz diagonal, $\Gamma = \mathrm{diag}(\psi_1, \psi_2, ..., \psi_q)$. Os fatores f_t são representados por um vetor $k \times 1$ o qual segue um modelo auto-regressivo multivariado, onde as matrizes AR, ρ_i, são diagonais com $\rho_i = diag(\rho_{i1}, ..., \rho_{ik})$, $i = 1, ..., p$ e $\{\rho_{1j}, \rho_{2j}, ..., \rho_{pj}\}$

satisfazendo a condição de estacionariedade, $j = 1, .., k$ e w_t são k-vetores independentes normais com $w_t \sim N(0, I_k)$, onde I_k é a matriz identidade, e e_t e w_s são independentes para todo t e s.

Para este modelo temos $f_t \sim N(0, U)$ e $U = \sum_{i=1}^{p} \rho_i' U \rho_i + I_k$, para todo t e $y_t \sim N(\beta, \Sigma)$, onde Σ satisfaz $\Sigma = CUC' + \Gamma$.

A matriz de carga C é assumida ser triangular inferior de posto completo para evitar problemas de identificabilidade. Assim,

$$C = \begin{pmatrix} 1 & 0 & 0 & ... & 0 \\ c_{21} & 1 & 0 & ... & 0 \\ c_{31} & c_{32} & 1 & ... & 0 \\ : & : & : & ... & : \\ c_{k1} & c_{k2} & c_{k3} & ... & 1 \\ : & : & : & ... & : \\ c_{q1} & c_{q2} & c_{q3} & ... & c_{qk} \end{pmatrix} .$$

Essa forma da matriz C na inferência Bayesiana é equivalente a dizer que os valores fixos ocorrem com probabilidade um e não serão estimados. O número de fatores será selecionado analizando os autovalores da matriz de covariancia (Peña e Box,1987) e o critério Bayesiano de Schwartz, BIC, utilizado na escolha da ordem do modelo auto-regressivo.

A função de verossimilhança condicionada na p-primeiras observações é dada por

$$\mathcal{L}(Y, F|\beta, C, \Gamma, \rho, Y_1, ..., Y_p, f_1, ..., f_p)$$

$$\propto \prod_{t=p+1}^{n} |\Gamma|^{-1/2} \exp\{-\frac{1}{2}(y_t - \beta - Cf_t)' \Gamma^{-1} (y_t - \beta - Cf_t)\} \times$$

$$\exp\{-\frac{1}{2}(f_t - \sum_{i=1}^{p} \rho_i f_{t-i})'(f_t - \sum_{i=1}^{p} \rho_i f_{t-i})\}, \tag{2}$$

onde $F = (f_{p+1}, ..., f_n)$, $Y = (y_{p+1}, ..., y_n)$ e os parâmetros são $\theta = (\beta, C, \Gamma, \rho)$.

Assumimos prioris independentes dadas por: $P(C)P(\rho)P(\beta) \propto$ constant, $\gamma_i = \psi_i^{-1} \sim Gamma(\alpha_0, \beta_0)$, tal que ψ_i tem distribuição Gama Inversa para cada componente $\Gamma = \text{diag}\{\psi_1, \psi_2, ..., \psi_q\}$.

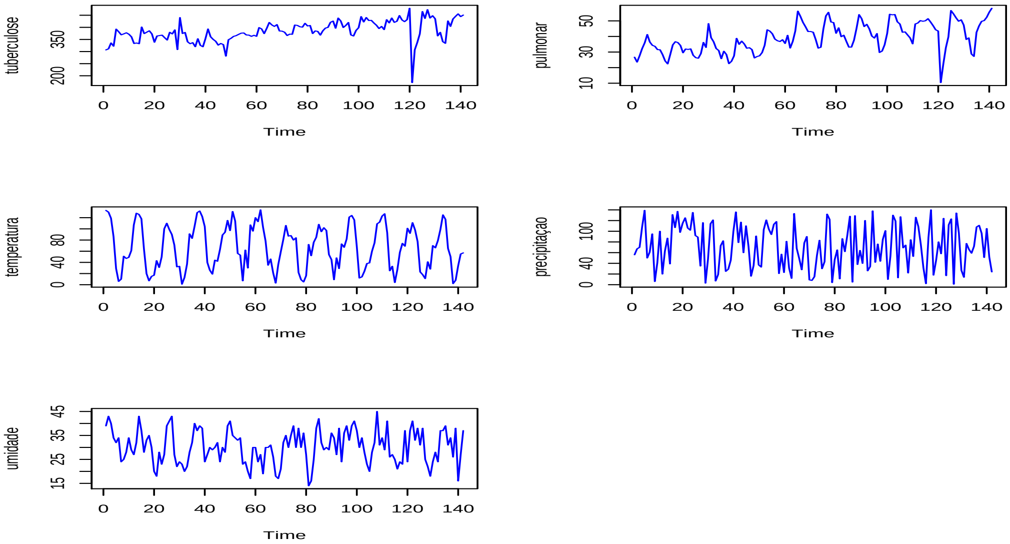

Figura 1: Dados mensais para as taxas de internações por tuberculose e doenças pulmonares, temperatura mínima, índice pluviométrico e umidade mínima.(01/98 a 09/09)

Para a implementação do Amostrador de Gibbs e posterior inferência dos parâmetros são necessárias as distribuições condicionais completas a posteriori as quais podem ser encontradas em Sáfadi e Peña (2008).

3 Morbidade pulmonar e séries climáticas

Neste trabalho consideramos as séries mensais de internações por tuberculose, internações por problemas pulmonares, temperatura mínima, índice pluviométrico e umidade mínima na cidade de São Paulo durante o período de Janeiro de 1998 a Setembro de 2009. Para as séries de internações consideramos a taxa de internações por cada 100.000 habitantes.

Os gráficos das séries são apresentados na Figura 1. Observa-se o caráter sazonal para a maioria das séries, comportamento este observado também nos gráficos das funções de autocorrelação, Figura 2.

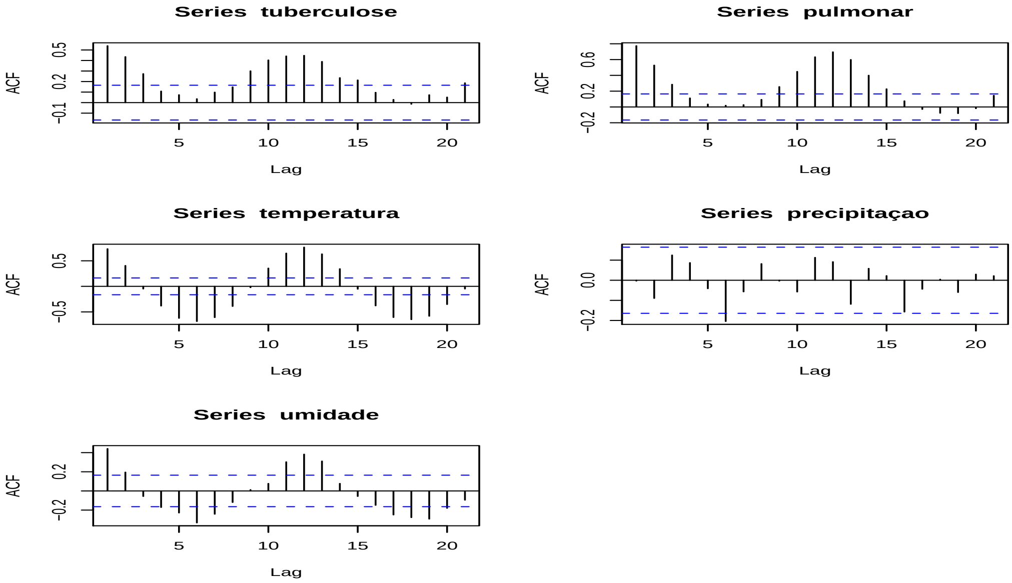

Figura 2: Gráficos das funções de autocorrelação .

Para a análise foram consideradas as séries sazonalmente ajustadas, Figura 3. A escolha do número de fatores é feita a partir do número de autovalores da matriz de covariância maiores que um, Tabela 1. A ordem do modelo autorregressiva para os fatores foi calculada analisando os valores do critério Bayesiano, BIC, para o modelo auto-regressivo vetorial VAR(p) para $p = 1, ..., 4$ (Tabela 2). A partir dos resultados apresentados nas Tabelas 1 e 2, a análise Bayesiana foi conduzida para o modelo fatorial dinâmico com 2 fatores os quais seguem um modelo AR(1).

O Amostrador de Gibbs foi utilizado para gerar duas cadeias, de tamanho 20.000 cada, das posterioris condicionais completas para cada parâmetro. As observações resultantes foram obtidas descartando as 50%-primeiras de cada cadeia e tomando uma a cada 15 observações. O critério de Gelman e Rubin (1992) foi utilizado na verificação da convergência.

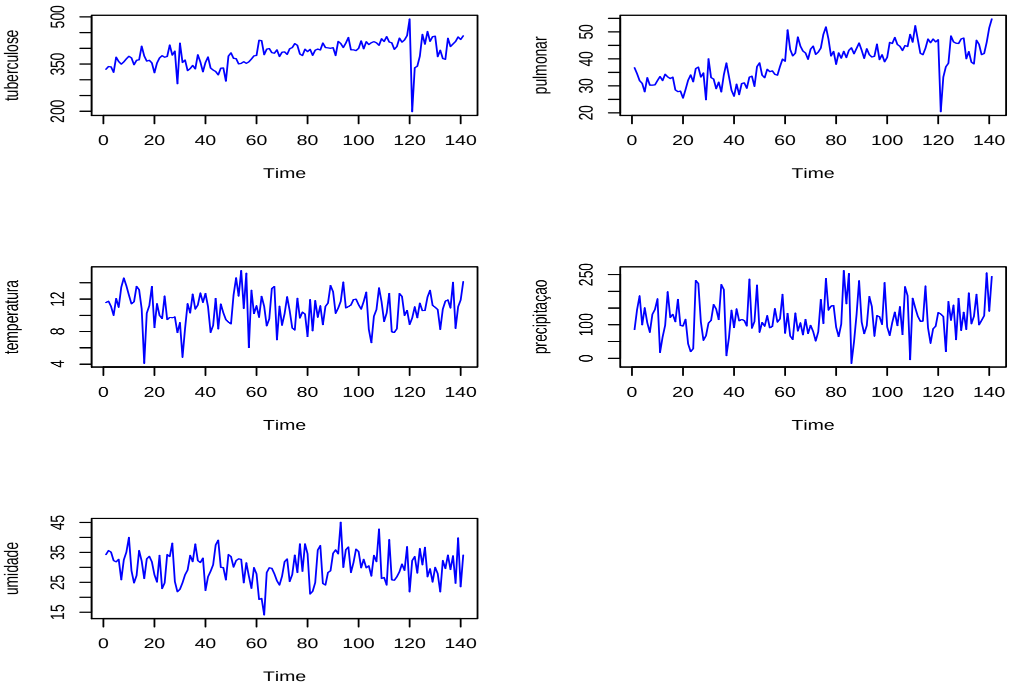

Figura 3: Gráfico das séries sazonalmente ajustadas taxa de internação por problema pulmonar, temperatura mínima, índice pluviométrico e umidade mínima.

Tabela 1: Autovalores da matriz de correlação .

	Autovalores	Proporção	Acumulado
1	1,8642	0,3728	0,3728
2	1,4191	0,2838	0,6567
3	0,9448	0,1890	0,8456
4	0,6190	0,1238	0,9694
5	0,1529	0,0306	1,0000

Tabela 2: BIC para o modelo VAR(p).

p	1	2	3	4
BIC	36,190877	36,854747	37,564536	38,279535

As médias a posteriori da matriz de covariância, Γ , os coeficientes auto-regressivos, a matriz de carga, os respectivos desvios padrão e o critério de Gelman e Rubin R são apresentados na Tabela 3.

Observamos que os parâmetros auto-regressivos ρ_1 e rho_2 são significativos, indicando que estes fatores no instante t são correlacionados com o instante $t-1$. As séries taxa de internação por tuberculose e índice pluviométricos apresentam as maiores variâncias. As séries de internações por problemas pulmonars e umidade possuem as menores variâncias.

A construção da matriz de carga indica que o primeiro fator associa positivamente as taxas de internação por tuberculose e problemas pulmonares e negativamente a umidade relativa mínima, o segundo fator associa a taxa de internação por problemas pulmonares com as séries de temperatura mínima, índices pluviométricos e umidade relativa mínima.

As distribuições marginais a posteriori para cada parâmetro são apresentadas nas Figuras 4, 5 e 6. Observa-se que a grande maioria das distribuições é assimétrica, fato este que poderia ser usado na escolha das prioris.

Tabela 3: Médias a posteriori para os parâmetros do modelos fatorial com 2 fatores- AR(1) e critério R.

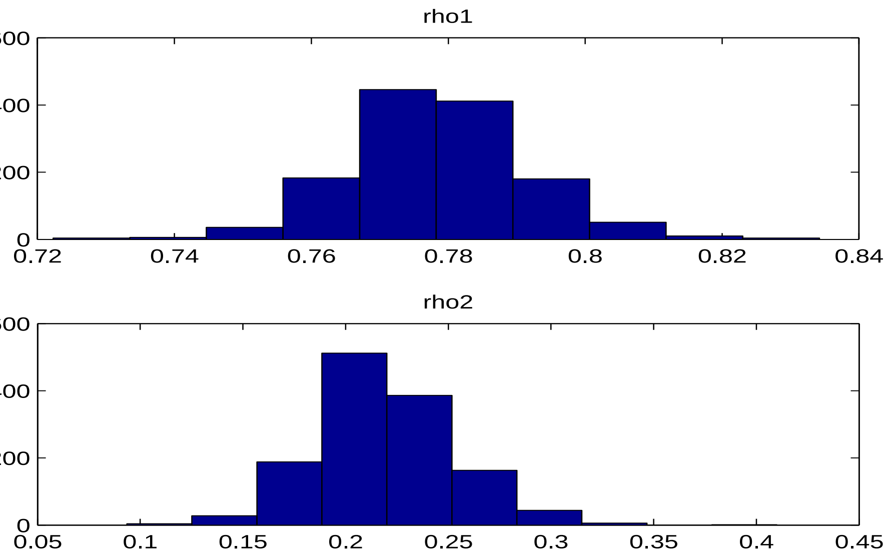

Figura 4: Distribuição marginal a posteriori para os coeficientes auto-regressivos

Parâmetro	média	d.p.	R
ρ_1	**0,7785**	0,0130	1,0021
ρ_2	**0,2185**	0,0338	1,0012
ψ_{11}	**1286,0**	148,43	0,9997
ψ_{22}	0,9000	0,6226	1,0144
ψ_{33}	**3,6000**	0,4257	1,0005
ψ_{44}	**2680,7**	329,46	1,0002
ψ_{55}	1,0000	0,6735	0,9995
c_{21}	**4,3406**	0,2826	1,0037
c_{31}	-0,0259	0,1099	1,0034
c_{32}	**0,4595**	0,1685	1,0043
c_{41}	1,3134	2,9702	1,0029
c_{42}	**19,3130**	4,7094	1,0000
c_{51}	**-0,7360**	0,3146	1,0011
c_{52}	**4,9483**	0,3305	1,0013

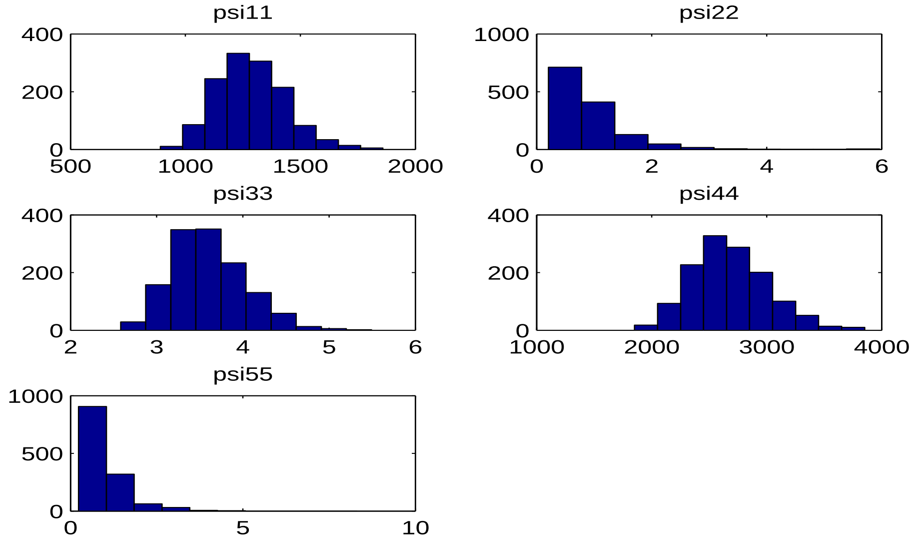

Figura 5: Distribuição marginal a posteriori para os elementos da matriz Γ

4 Conclusão

Concluímos que usando o modelo fatorial é possível ajustar um modelo para as séries de internações e de condições climáticas com um número menor de parâmetros do que seria se ajustassemos modelos vetoriais autorregressivos.

5 Agradecimentos

As autoras agradecem a Fapemig e a Capes projeto Procad 177/2007 o auxilio financeiro.

6 Referências

Gelman, A. ; Rubin, D.B. 1992. Inference from iterative simulation using multiple sequences. *Statistical Science*, **7**, 457-511.

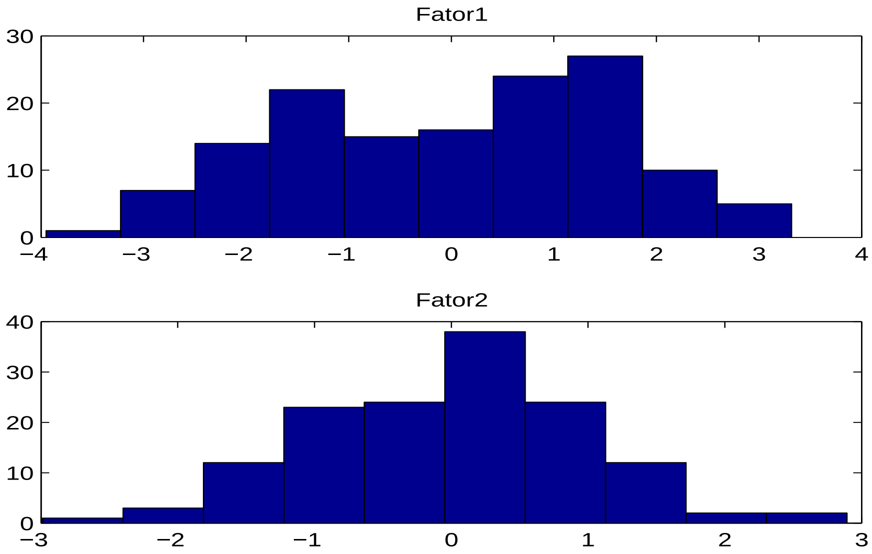

Figura 6: Distribuição marginal a posteriori para os fatores

Peña, D., Box, G.E.P. 1987. Identifying a simplifying structure in time series. *Journal of the American Statistical Association*, 82(399), 836-843.

Sáfadi, T. ; Peña, D. 2008. Bayesian analysis of dynamic factor model: an application to air pollution and mortality in São Paulo, Brazil. *Environmetrics, 19* Wiley InterScience. 582-601.