

4

Aplicação e comparação de modelos

4.1

Aplicação em Epidemiologia Ambiental

Estudos epidemiológicos têm fornecido evidências sobre efeitos adversos à saúde devido a poluição do ar. Estes efeitos podem ser observados no aumento da mortalidade e da morbidade da população, principalmente por doenças respiratórias e cardiovasculares. A preocupação com os efeitos nocivos da poluição recebeu maior atenção a partir dos episódios de alta concentração de poluentes ocorridos no Vale Meuse, na Bélgica, em 1930, em Donora, nos Estados Unidos, em 1948 e em Londres, Inglaterra, em 1952. Neste último, o número total de óbitos observado elevou-se em mais de cinco vezes em relação ao esperado para aquela época do ano. Com a adoção de medidas visando controlar a emissão de poluentes na atmosfera, episódios como estes não voltaram a ocorrer. O avanço das técnicas estatísticas tem viabilizado o estudo dos possíveis efeitos da poluição do ar sobre a saúde mesmo quando os níveis ambientais são muito menores do que aqueles tradicionalmente considerados como nocivos [2, 32, 8, 15].

As técnicas de análise de séries temporais têm desempenhado um papel relevante para o desenvolvimento destes estudos. Estes estudos são do tipo ecológico, com dados agregados tanto para a exposição quanto para as informações de saúde. As principais vantagens do uso deste método em estudos epidemiológicos são que um mesmo grupo de indivíduos é avaliado ao longo do tempo e, portanto, a população é seu próprio controle. Fatores como residência, ocupação, dieta, tabagismo e atividades físicas tendem a se manter constantes na população dentro do período investigado e outros fatores de confusão como efeito de calendário, fatores climáticos, epidemias e estruturas da própria série como tendência e sazonalidade podem ser controlados [32, 27].

Nos estudos epidemiológicos que envolvem a análise de séries temporais, o que se pretende usualmente é verificar se variações de curto prazo nos

níveis da variável resposta estão de alguma forma associadas com flutuações de curto prazo na variável de exposição. No caso dos estudos dos efeitos da poluição na saúde, se busca verificar se variações na concentração diária de um ou mais poluentes estão relacionadas com variações na contagem diária de internações hospitalares, óbitos ou outra medida de frequência de eventos de saúde de curto prazo. Também se deseja estimar a magnitude deste efeito, ou seja, qual o acréscimo no número de hospitalizações ou óbitos correspondente a um dado aumento na concentração de poluentes [32, 26].

A análise de correlação entre as variáveis de desfecho e exposição não é adequada para estimar a associação já que podem existir estruturas inerentes à dinâmica das séries temporais, tais como tendência, sazonalidade, ciclos e autocorrelação, que podem enviesar a estimativa. Há também outros efeitos concorrentes relacionados com a variável de desfecho e/ou com a variável de exposição que precisam ser removidos. Estas perturbações confundem as medidas de associação e devem ser adequadamente controladas [19]. Efeitos do calendário – como dias da semana e feriados –, greves, epidemias e condições meteorológicas podem constituir situações de confusão nos estudos de séries temporais de poluição do ar e saúde [32].

A metodologia padrão utilizada atualmente para a análise estatística de séries temporais nos estudos sobre poluição do ar e efeitos na saúde são os modelos aditivos generalizados [27, 23]. Por se tratar de dados de contagem, um dos modelos probabilísticos adequados é o modelo de Poisson. Nestes modelos, com a função de ligação canônica, os coeficientes das covariáveis descrevem a variação percentual no logaritmo da média da variável resposta. A função exponencial avaliada no coeficiente estimado da variável que representa a exposição ao poluente é o risco relativo para o evento de saúde, mortalidade ou morbidade, devido ao aumento em uma unidade do poluente [19].

A fim de ilustrar a aplicação dos modelos Poisson-gama semi-paramétricos, considere-se investigar o efeito de curto prazo da exposição a material particulado com diâmetro aerodinâmico de até $10 \mu\text{m}$ (PM10) sobre o número diário de internações de crianças por doenças do aparelho respiratório (DAR). O desenho epidemiológico empregado é o ecológico de caráter temporal com dados provenientes de fontes secundárias. A população alvo é constituída de todas as crianças residentes no município do Rio de Janeiro com até 5 anos de idade. A série temporal das internações compreende o período de setembro de 2000 a agosto de 2002, totalizando 730 observações.

A série diária de internações hospitalares de caráter não eletivo por

doenças do aparelho respiratório (CID 10 - cap. X) de crianças menores de 5 anos, residentes no município do Rio de Janeiro, foi gerada a partir das internações realizadas no âmbito do Sistema Único de Saúde (SUS) através das Autorizações de Internação Hospitalar (AIH) contidas nos bancos de dados disponibilizados pelo Ministério da Saúde.

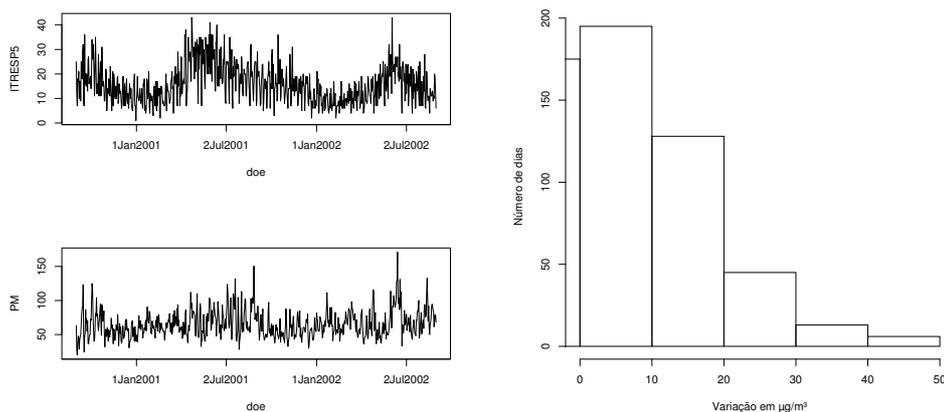
Para obter uma estimativa média da exposição diária da população residente no Rio de Janeiro à poluição, foi construída a série média diária do poluente ambiental PM10, a partir das medidas de quatro estações de monitoramento: Copacabana, Centro, São Cristóvão e Tijuca. Os dados diários do poluente analisado foram fornecidos pela Secretaria Municipal do Meio Ambiente da Cidade do Rio de Janeiro (SMAC).

A exposição média diária da população aos fatores meteorológicos foi obtida através das médias aritméticas das medidas de umidade e de temperaturas mínima, média e máxima referentes aos monitores localizados nos aeroportos de Santa Cruz, Campo dos Afonsos, Galeão e Santos Dumont e no bairro do Maracanã. As medidas foram obtidas dos Departamentos de Meteorologia da Aeronáutica e de Climatologia da Universidade do Estado do Rio de Janeiro, respectivamente. Os dados de intensidade das chuvas foram obtidos na página oficial da Internet da Fundação Instituto de Geotécnica do município do Rio de Janeiro.

Com o objetivo de controlar os fatores de confusão, informações sobre calendário – dias da semana e feriados –, condições meteorológicas e climáticas devem ser introduzidas no modelo. Os eventos de saúde, bem como as concentrações de poluentes, se comportam de forma diferenciada nos dias da semana, logo é necessário introduzir no modelo variáveis indicadoras para dia da semana. Os feriados também modificam o padrão de emissão de poluentes e para estimar adequadamente o efeito da poluição sobre a saúde, o modelo deve contemplar esta fonte de variação. Da mesma forma, a temperatura, a umidade e a precipitação de chuvas modificam o efeito da poluição sobre os eventos de saúde e precisam ser adequadamente controlados. O modelo com apenas os termos para controle de confusão, sem as covariáveis de exposição, é denominado *modelo central* [32, 26].

Os feriados foram agrupados de modo que aqueles com efeito de mesmo sinal e magnitudes semelhantes constituíssem um mesmo fator. As informações sobre as concentrações de poluentes atmosféricos apresentavam uma grande quantidade de dados faltando e foi necessário empregar um método de imputação. Detalhes sobre o método de imputação empregado podem ser obtidos em Junger (2002) [25].

Os gráficos 4.1(a) descrevem as séries de desfecho e de exposição e



4.1(a): Séries de DAR e PM10

4.1(b): Variação média de PM10

Figura 4.1: Valores observados da série de desfecho (DAR de crianças até 5 anos) e da série de exposição (PM10) 4.1(a) e histograma truncado das variações médias diárias positivas de PM10 4.1(b).

se percebe que os dias com maior número de internações coincidem com os dias de maior concentração de material particulado. A série de internações diárias de crianças com até 5 anos de idade devido a doenças do aparelho respiratório tem média 15,75 e desvio-padrão 7,78. Enquanto a série de concentrações médias diárias de material particulado tem média e desvio-padrão iguais a 65,50 e 19,39 respectivamente.

O histograma no gráfico 4.1(b) representa a distribuição truncada das variações que caracterizam um aumento médio diário na concentração de material particulado. Em 193 dias do período de estudo a concentração do poluente sofreu aumentos superiores ou iguais a $10 \mu\text{g}/\text{m}^3$, enquanto que em 387 dias os aumentos ocorridos foram superiores a 0.

Para este problema foi estimado um modelo Poisson-gama semi-paramétrico composto pelos termos lineares listados na tabela 4.1, uma curva *spline* com 7 graus de liberdade equivalentes da temperatura máxima defasada em dois dias e uma curva *spline* com 3 graus de liberdade equivalentes da umidade relativa do ar também defasada em dois dias. Os graus de suavização das curvas de temperatura e umidade foram selecionados de forma empírica e subjetiva, observando-se o diagnóstico do modelo.

Os hiperparâmetros do modelo foram estimados usando o método de otimização BFGS [34]. Os valores estimados estão apresentados na tabela 4.1. O algoritmo de estimação semi-paramétrica convergiu após 6 iterações

¹não estimado numericamente.

Tabela 4.1: Modelo Poisson-gama semi-paramétrico para os dados de internações hospitalares de crianças por doenças respiratórias.

variável	coeficiente	erro-padrão	razão t
fator de desconto	0,86270	0,08518	10,12776
domingo	-0,43353	0,02906	-14,91886
segunda	0,24737	0,02151	11,50122
terça	0,17790	0,02199	8,08801
quarta	0,14524	0,02226	6,52488
quinta	0,15098	0,02248	6,71451
sexta	0,18348	0,02241	8,18548
sábado	-0,47146	-	- ¹
feriados 1	-0,00555	0,16930	-0,03276
feriados 2	-0,32876	0,09515	-3,45503
feriados 3	-0,53535	0,07315	-7,31862
enforcados	-0,41950	0,08831	-4,75039
chuva	-0,00325	0,00167	-1,94134
PM10	0,00104	0,00059	1,77030

e o valor máximo obtido para o logaritmo da verossimilhança é 2.099,538. O valor estimado da função desvio é 826,631. O critério de Akaike aproximado é 1,204.

A análise de diagnóstico indica que o modelo estimado apresenta boa adequação aos dados de internação de crianças por doenças do aparelho respiratório no período de setembro de 2000 a agosto de 2002. O gráfico da figura 4.2, com os valores observados e previstos pelo modelo, mostra que a variabilidade dos valores observados é explorada pelo modelo estimado. Esta percepção é corroborada pela ampliação do gráfico da figura 4.2.

Na figura 4.3, o gráfico de valores observados e previstos é ampliado num fator de sete vezes. É possível perceber que a linha tracejada referente a trajetória das observações é reproduzida com boa precisão pela linha contínua referente às previsões do modelo ajustado.

Na figura 4.4, o gráfico dos resíduos de desvio apresenta-se sem estruturas visíveis e concentrado na região entre -3 e 3. Embora não seja imperativa a normalidade dos resíduos nesta classe de modelos, no gráfico de quantis da figura 4.5, observa-se que os quantis dos resíduos do modelo Poisson-gama semi-paramétrico estimado são aderentes com a reta definida pelos quantis da distribuição normal teórica.

Quanto a estruturas de dependência temporal, nos gráficos da figura 4.8 verifica-se que tanto a função de autocorrelação quanto a função de autocorrelação parcial dos resíduos são estatisticamente iguais a zero.

O gráfico de envelope simulado dos resíduos da figura 4.7 mostra que

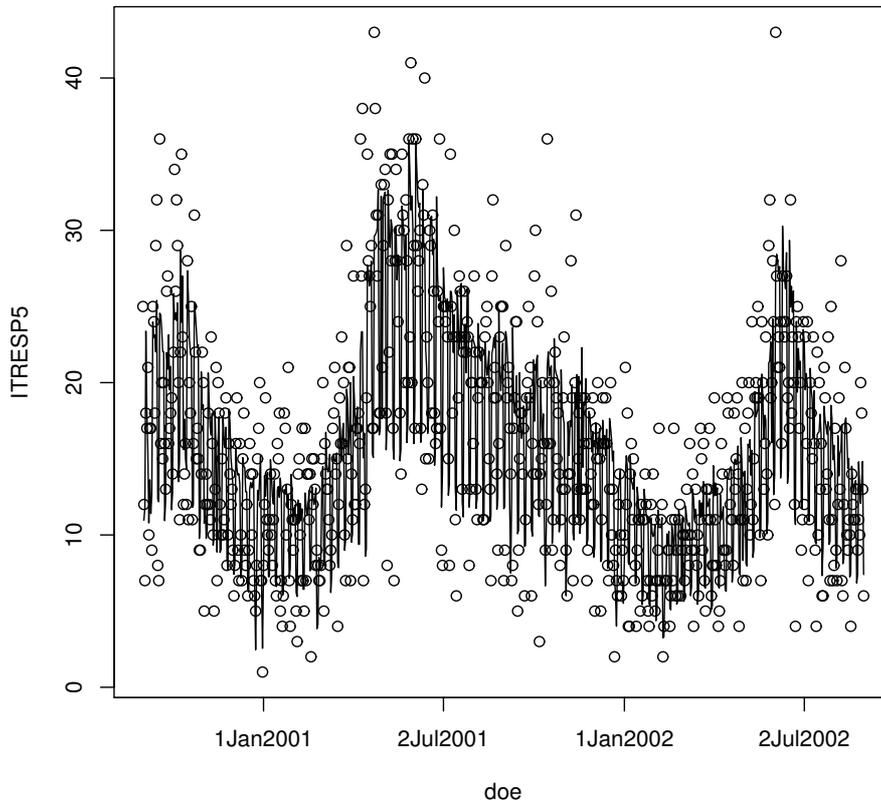


Figura 4.2: Valores observados (pontos) e valores previstos (linha contínua) pelo modelo Poisson-gama semi-paramétrico.

apenas nenhum ponto está fora da banda de confiança. O resultado desta análise indica uma boa adequação do modelo estimado. Este gráfico foi construído com 100 replicações da série de desfecho estimada.

O risco relativo de interações de crianças de até 5 anos de idade por doenças do aparelho respiratório para um aumento de $10 \mu\text{g}/\text{m}^3$ de PM10 pode ser obtido a partir do coeficiente estimado da tabela 4.1. Assim, o risco relativo pode ser calculado de acordo com a expressão

$$RR(10 \mu\text{g}/\text{m}^3) = \exp(10 \cdot 0,00104) = 1,01045 \quad (4-1)$$

o que corresponde a uma variação percentual de 1,05% para uma variação de $10 \mu\text{g}/\text{m}^3$ de PM10. Este efeito estimado não é estatisticamente significativo, pois o intervalo de confiança de 95% dado por $[0,99889; 1,02214]$ contém a unidade.

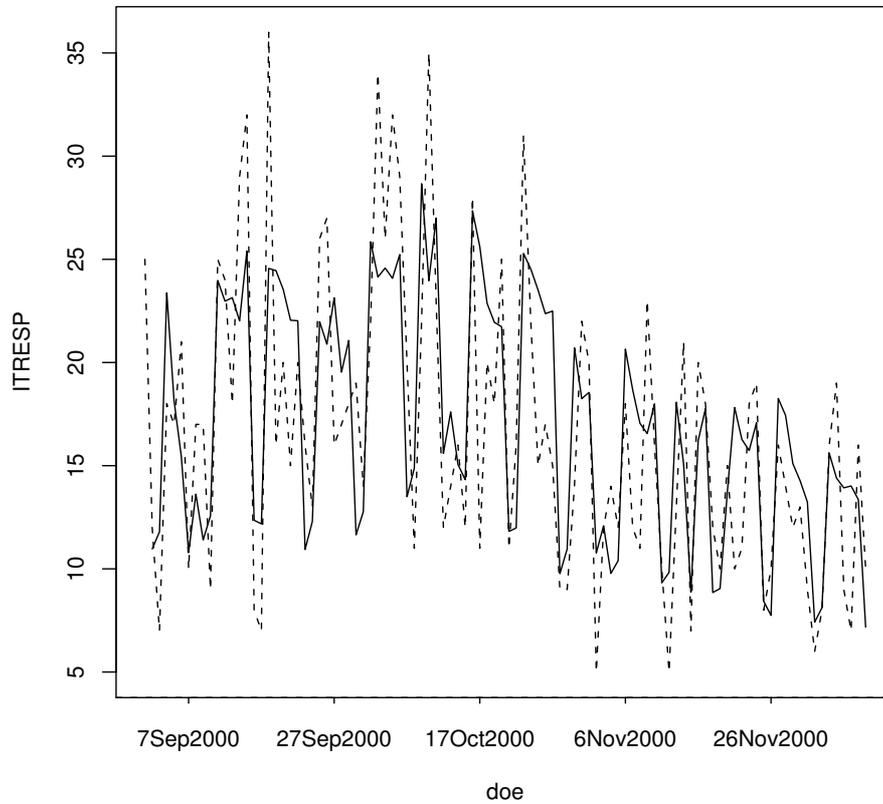


Figura 4.3: Valores observados (linha descontínua) e valores previstos (linha contínua) pelo modelo Poisson-gama semi-paramétrico – ampliação.

4.2

Comparação com outra classe de modelos

Nesta seção, a mesma aplicação é analisada usando outras classes de modelos. O problema de estimar o efeito de curto prazo da concentração de PM10 no ar sobre o número diário de internações de crianças de até 5 anos por doenças do aparelho respiratório é revisto. O modelo Poisson-gama semi-paramétrico é comparado com um modelo Poisson-gama paramétrico, sem suavização de covariáveis, e com um modelo aditivo generalizado (MAG), que hoje é considerado o estado da arte nas análises de séries temporais epidemiológicas em estudos sobre poluição do ar e efeitos na saúde [27, 23].

Na tabela 4.2, estão listadas as estimativas dos hiperparâmetros calculadas seguindo as diferentes abordagens próprias de cada classe de modelos. Na estimação do modelo aditivo generalizado foi considerada como linha de base a segunda-feira. No modelo poisson-gama semi-paramétrico e paramétrico não há linha de base para os fatores sazonais. Para aproximar

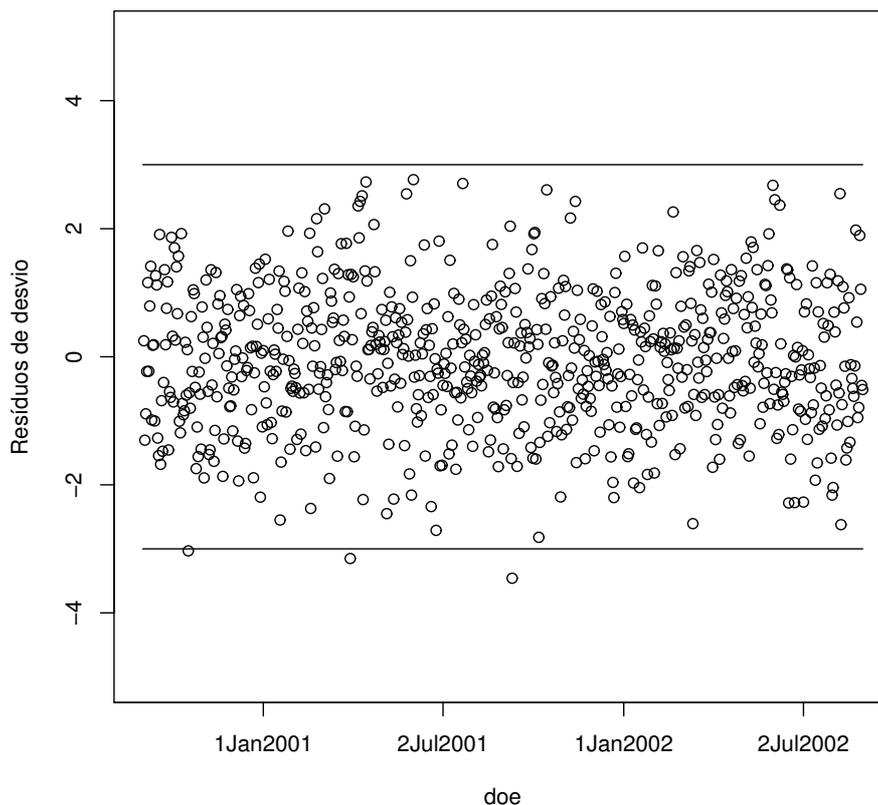


Figura 4.4: Resíduos de desvio padronizados do modelo Poisson-gama semi-paramétrico.

a estimativa do acréscimo em relação a uma linha de base para os dias da semana, é subtraída do coeficiente de um dia específico, por exemplo, a segunda-feira, o valor do coeficiente do dia escolhido como linha de base.

A tabela 4.2 mostra que os coeficientes estimados têm valores próximos nos três modelos, inclusive para os dias da semana se a segunda-feira for definida como linha de base. Diferenças referentes à incerteza das estimativas podem ser atribuídas às características inerentes aos métodos de estimação. As matrizes de covariâncias dos estimadores nos modelos Poisson-gama semi-paramétrico e Poisson-gama paramétrico são, nesta aplicação, estimadas numericamente a partir da matriz hessiana do algoritmo de otimização. Os mesmos critérios para diagnósticos foram empregados nos três modelos.

A estimativa pontual do coeficiente da variável que representa as concentrações de PM10 é menor no modelo Poisson-gama semi-paramétrico que no modelo aditivo generalizado. O risco relativo estimado pelo MAG

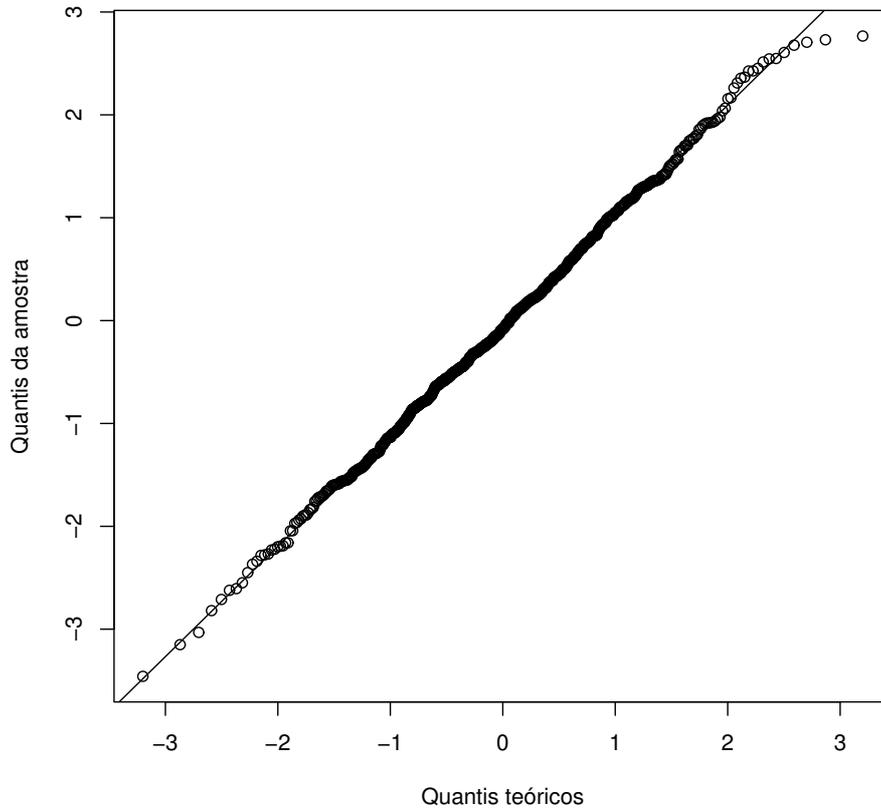
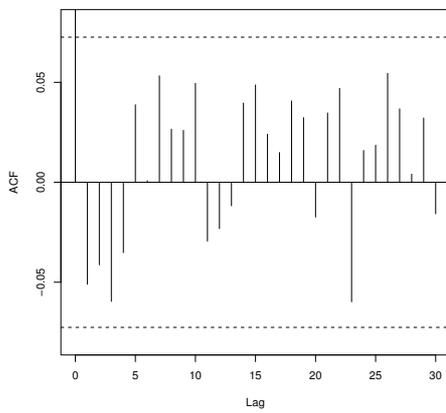
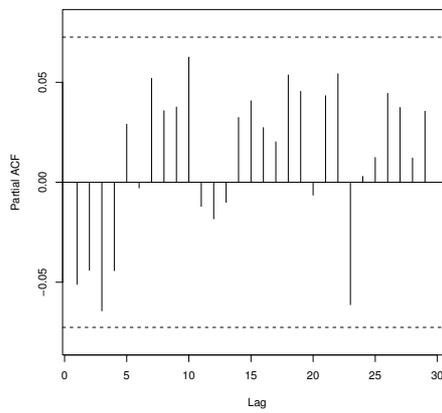


Figura 4.5: Quantis normais dos resíduos de desvio do modelo Poisson-gama semi-paramétrico.



4.6(a): FAC



4.6(b): FACP

Figura 4.6: FAC 4.8(a) e FACP 4.8(b) dos resíduos de desvio do modelo Poisson-gama semi-paramétrico.

Tabela 4.2: Modelo Poisson-gama semi-paramétrico (PGS), Modelo Poisson-gama paramétrico (PGP) e modelo aditivo generalizado (MAG) para os dados de internações hospitalares de crianças por doenças respiratórias³.

variável	modelo	coeficiente	erro-padrão	razão t
fator de desconto	PGS	0,86270	0,08518	10,12776
fator de desconto	PGP	0,86270	0,08518	10,12776
intercepto	MAG	2,78820	0,05155	54,09000
domingo	PGS	-0,68091	0,02906	-14,91886
	PGP	-0,71779	0,02998	-16,32781
	MAG	-0,70825	0,04341	-16,32000
segunda	PGS	0,24737	0,02151	11,50122
	PGP	0,22822	0,02177	10,48299
	MAG	*	*	*4
terça	PGS	-0,06946	0,02199	8,08800
	PGP	0,00331	0,02189	10,57384
	MAG	-0,06534	0,03536	-1,84800
quarta	PGS	-0,10212	0,02226	6,52488
	PGP	-0,07384	0,02229	6,92648
	MAG	-0,13265	0,03618	-3,66600
quinta	PGS	-0,09639	0,02248	6,71451
	PGP	-0,10979	0,02279	5,19592
	MAG	-0,15297	0,03645	-4,19600
sexta	PGS	-0,06389	0,02241	8,18548
	PGP	-0,07194	0,02277	6,87634
	MAG	-0,12590	0,03634	-3,46500
sábado	PGS	-0,71883	-	- ⁵
	PGP	-0,62749	-	-
	MAG	-0,60943	0,04168	-14,62000
feriados 1	PGS	-0,00554	0,16930	-0,03276
	PGP	-0,01346	0,16891	-0,07968
	MAG	0,00427	0,17830	0,02398
feriados 2	PGS	-0,32876	0,09515	-3,45503
	PGP	-0,25659	0,09385	-2,73408
	MAG	-0,19637	0,09795	-2,00500
feriados 3	PGS	-0,53535	0,07314	-7,31862
	PGP	-0,48069	0,07204	-6,67283
	MAG	-0,43078	0,07547	-5,70800
enforcados	PGS	-0,41950	0,08831	-4,75039
	PGP	-0,35286	0,08629	-4,08922
	MAG	-0,32975	0,09199	-3,58400
chuva	PGS	-0,00325	0,00167	-1,94134
	PGP	-0,00251	0,00171	-1,46836
	MAG	-0,00284	0,00178	-1,59900
PM10	PGS	0,00104	0,00059	1,77030
	PGP	0,00181	0,00061	2,98105
	MAG	0,00216	0,00061	3,56400

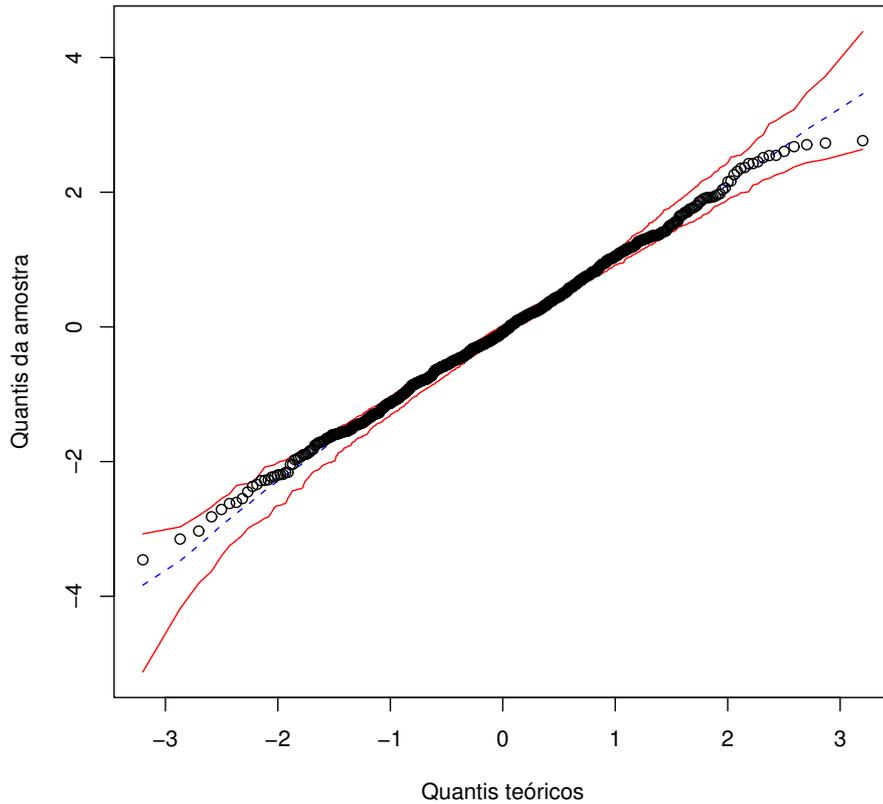


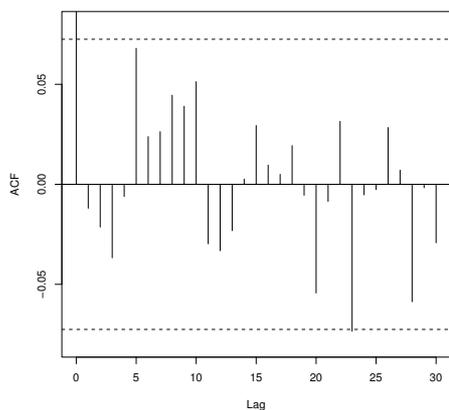
Figura 4.7: Envelope simulado dos resíduos de desvio do modelo Poisson-gama semi-paramétrico.

pode ser calculado de acordo com a expressão

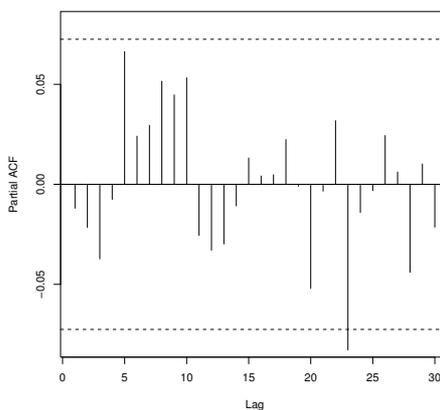
$$RR(10 \mu\text{g}/\text{m}^3) = \exp(10 \cdot 0,00216) = 1,02187 \quad (4-2)$$

e o intervalo de confiança de 95% para o risco relativo é dado por $[1,00978; 1,03410]$. Como o intervalo de confiança não inclui a unidade, o efeito que representa um aumento de 2,19% na média diária de internações para um aumento de $10 \mu\text{g}/\text{m}^3$ de PM10 é estatisticamente significativo.

No modelo Poisson-gama paramétrico, obtém-se um risco relativo igual a 1,01826, calculado da mesma forma que no MAG, que equivale a um aumento de 1,83% na média diária de internações para um aumento de $10 \mu\text{g}/\text{m}^3$ de PM10. O intervalo de confiança de 95% é igual a $[1,00616; 1,03051]$ e, como não inclui a unidade, é estatisticamente significativo o efeito



4.8(a): FAC



4.8(b): FACP

Figura 4.8: FAC 4.8(a) e FACP 4.8(b) dos resíduos de desvio do modelo aditivo generalizado.

estimado para PM10.

Não obstante encontrar coeficientes com valores semelhantes, os três modelos têm fundamentações teóricas e métodos de estimação diferentes. A comparabilidade dos modelos é, então, limitada por estes fatores.

Comparando-se as funções de autocorrelação e autocorrelação parcial dos dois modelos semi-paramétricos, percebe-se que os valores destas funções encontrados no MAG são ligeiramente maior que aqueles encontrados no PGS, apesar de se encontrarem dentro dos limites assintóticos.

Os gráficos 4.9(a) e 4.9(b) da figura 4.9 representam as curvas estimadas para temperatura máxima e umidade relativa sobre os respectivos resíduos parciais na última iteração do algoritmo *backfitting*. Estas curvas \hat{g} correspondem às *splines* cúbicas naturais com 7 e 3 graus de liberdade equivalentes, respectivamente, estimadas a partir dos dados para temperatura máxima e umidade relativa. Como os gráficos da figura 4.9 fazem notar, a curva estimada se aproxima da reta de regressão linear quando os números de graus de liberdade decrescem em direção a valores menores. Os gráficos 4.9(c) e 4.9(d) representam as mesmas curvas estimadas pelo modelo aditivo generalizado. Apenas para referências, os gráficos 4.9(e) e 4.9(f) representam as funções lineares estimadas pelo modelo Poisson-gama paramétrico.

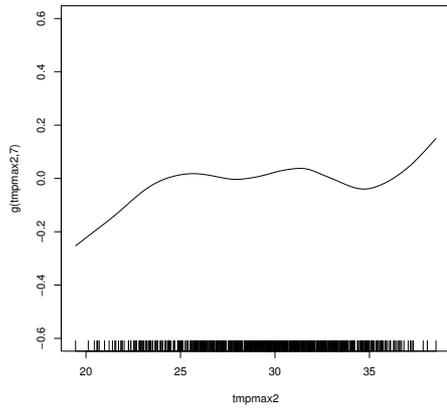
³A fim de facilitar a comparação dos dois modelos, os coeficientes das covariáveis referentes aos dias da semana no modelo Poisson-gama semi-paramétrico, são recalculados em relação a linha de base do modelo aditivo generalizado. Entretanto, os erros-padrão são relativos aos valores originais dos coeficientes.

⁴linha de base.

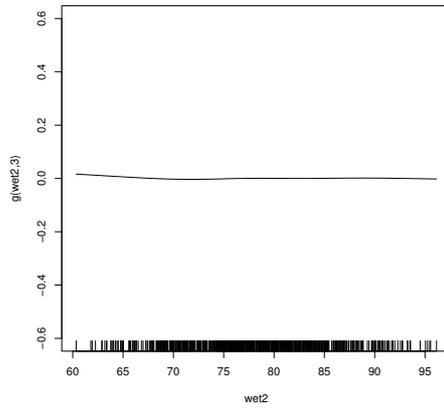
⁵não estimado numericamente.

Os gráficos estão na mesma escala para facilitar a comparação dos três modelos. Ambos procedimentos de estimação semi-paramétrica levam a curvas semelhantes na forma e escala para a covariável de temperatura máxima, porém uma curva mais acentuada para umidade relativa é obtida através do modelo aditivo generalizado.

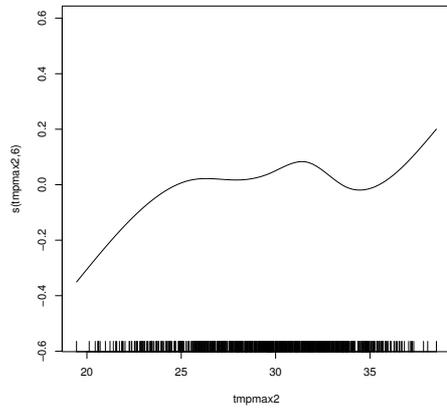
Outro aspecto que merece atenção diz respeito à utilização de recursos computacionais. Nos modelos aditivos generalizados, os estimadores de máxima verossimilhança são obtidos analiticamente a cada iteração do IRLS [28] e do algoritmo *backfitting*. Nos modelos Poisson-gama semi-paramétricos e Poisson-gama paramétricos, as estimativas de máxima verossimilhança dos hiperparâmetros são obtidos via solução de um problema de otimização não linear em um espaço de alta dimensão. Não existem garantias de otimalidade para os hiperparâmetros estimados. Nos modelos Poisson-gama semi-paramétricos é necessário resolver um problema de otimização da função de verossimilhança a cada iteração do algoritmo *backfitting*. Este processo torna a estimação destes modelos dispendiosa em termos de tempo de estimação e recursos computacionais em comparação aos modelos aditivos generalizados.



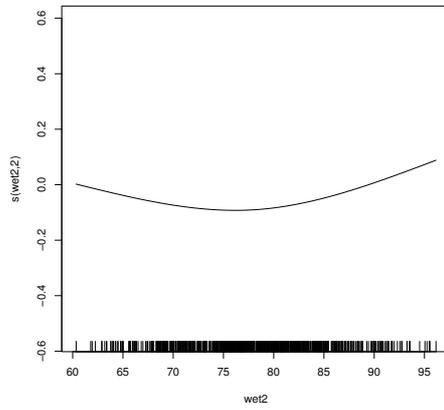
4.9(a): Temperatura máxima - PGS



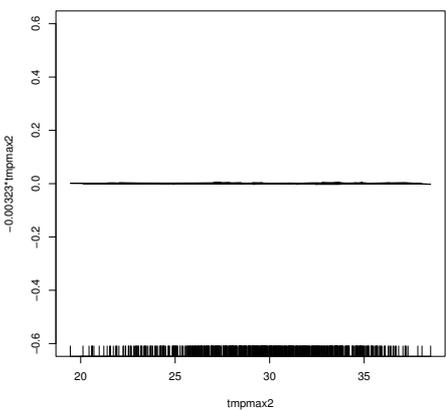
4.9(b): Umidade relativa - PGS



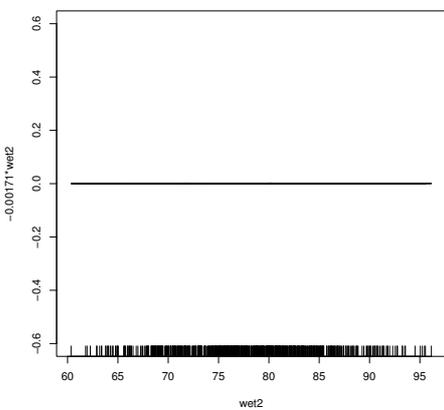
4.9(c): Temperatura máxima - MAG



4.9(d): Umidade relativa - MAG



4.9(e): Temperatura máxima - PGP



4.9(f): Umidade relativa - PGP

Figura 4.9: Suavização das variáveis temperatura máxima e umidade relativa sobre os resíduos parciais do modelo Poisson-gama paramétrico (PGS) 4.9(a) e 4.9(b) e do modelo aditivo generalizado (MAG) 4.9(c) e 4.9(d) e retas estimadas pelo modelo Poisson-gama paramétrico 4.9(e) e 4.9(f).