

5 Conclusão

Atualmente, a metodologia padrão utilizada nas análises de séries temporais em estudos epidemiológicos sobre os efeitos da poluição do ar na saúde são os modelos aditivos generalizados. Esta classe de modelos, que é uma extensão dos modelos lineares generalizados, relaxa a linearidades das funções no preditor linear, permitindo que outras formas funcionais sejam empregadas. Esta propriedade dos modelos aditivos generalizados permite que os efeitos de covariáveis que constituem fator de confusão nos modelos epidemiológicos sejam facilmente corrigidos.

A característica de corrigir adequadamente os fatores de confusão é devida à flexibilidade das funções suaves que constituem a partição aditiva do modelo. A escolha natural para estas funções suaves são aquelas definidas pelos próprios dados por meio de um operador de suavização. Vários destes operadores de suavização estão disponíveis. Entretanto, as *splines* cúbicas naturais apresentam propriedades matemáticas consideradas atrativas. A mais importante destas propriedades é que estes operadores são lineares, no sentido de que existe uma matriz que transforma os dados observados, produzindo a curva suave estimada. Esta propriedade permite, por exemplo, a utilização de diagnósticos originalmente desenvolvidos para a análise de regressão usual na análise de regressão não-paramétrica.

Não obstante o fato de terem sido admitidos como metodologia padrão nos estudos sobre poluição do ar e efeito na saúde, os modelos aditivos generalizados apresentam limitações em se tratando do ajuste de estruturas intrínsecas nos dados de séries temporais, como tendência, sazonalidade e autocorrelação. Embora o efeito da tendência e da sazonalidade sejam capturados satisfatoriamente pela curva suave estimada a partir da covariável que representa o índice do instante t , o componente de autocorrelação necessita de métodos *ad hoc* para ser devidamente controlado.

Os modelos estruturais em espaço de estados apresentam uma solução apropriada para lidar com essas estruturas das séries temporais no ambiente gaussiano. Para dados de contagem, os modelos Poisson-gama contam com

um componente de nível estocástico que evolui no tempo seguindo um passeio aleatório. Este movimento proporciona um ajuste adequado do componente de autocorrelação. Entretanto, sua formulação estritamente paramétrica os desqualifica como uma opção razoável aos modelos aditivos generalizados.

Neste trabalho, é introduzida uma nova classe de modelos para séries temporais de dados de contagem, os modelos Poisson-gama semi-paramétricos. Nesta classe, os modelos Poisson-gama paramétricos são estendidos para uma formulação híbrida – paramétrica e não-paramétrica –, na qual funções não-lineares são admitidas como parte do preditor. O preditor linear dos modelos Poisson-gama é substituído por uma soma de funções genéricas, das quais as funções lineares constituem um caso particular.

Tal como ocorre nos modelos aditivos generalizados, funções suaves são empregadas para representar a associação das covariáveis com a variável resposta. Devido às suas propriedades matemáticas, as *splines* cúbicas naturais estimadas a partir de cada covariável constituem uma classe adequada de funções para compor a partição não-paramétrica do preditor.

A estimação dos hiperparâmetros dos modelos Poisson-gama semi-paramétricos é obtida por máxima verossimilhança. A partição não-paramétrica com múltiplas covariáveis dos modelos é estimada por meio de uma regressão não-paramétrica pelo algoritmo *backfitting*. Os processos de maximização da verossimilhança e *backfitting* são iterados até que o valor da verossimilhança convirja para um valor finito.

Para os diagnósticos de modelos Poisson-gama semi-paramétricos podem ser empregados extensões daqueles já utilizados nos modelos aditivos generalizados e no Poisson-gama paramétrico. Alguns procedimentos de inferência estatística do modelo são apresentados e estão implementados na biblioteca *pgam*. Entretanto, uma teoria ditribucional adequada dos modelos semi-paramétricos, inclusive os modelos aditivos generalizados, ainda está em desenvolvimento. Os procedimentos apresentados baseiam-se em aproximações e devem ser utilizados basicamente para seleção de modelos.

A fim de comparar os três modelos, uma série epidemiológica, investigando a associação entre material particulado e internações de crianças por doenças do aparelho respiratório no Rio de Janeiro, foi analisada. Foram utilizados um modelo Poisson-gama semi-paramétrico, um modelo Poisson-gama paramétrico e um modelo aditivo generalizado na análise. Os três modelos levaram a resultados semelhantes, sobretudo nos coeficientes das covariáveis de confusão. As curvas das covariáveis da partição aditiva dos

dois modelos semi-paramétricos também apresentaram-se semelhantes.

O coeficiente da covariável que representa a exposição ao poluente foi estimado com valor menor no modelo Poisson-gama semi-paramétrico que no modelo Poisson-gama paramétrico e no modelo aditivo generalizado. Este fato leva a um importante questionamento: se o coeficiente da covariável de exposição é subestimado devido ao método de estimação em si ou se um modelo capaz de ajustar adequadamente as estruturas intrínsecas da série temporal e permitir formas mais flexíveis das covariáveis de confusão for usado, obtém-se uma estimativa do efeito mais próxima do valor verdadeiro desconhecido. Um estudo de simulação dos modelos Poisson-gama semi-paramétricos propostos neste trabalho é o próximo passo natural para a obtenção da resposta a esta questão. Também se faz necessário um estudo mais detalhado das propriedades herdadas e combinadas do Poisson-gama paramétrico, das *spline* cúbicas naturais e do algoritmo *backfitting*.

Não obstante o esforço computacional envolvido na estimação dos modelos Poisson-gama semi-paramétricos, os resultados obtidos são compatíveis com aqueles dos modelos aditivos generalizados. Entretanto, este modelo oferece como vantagem a característica nativa de modelar adequadamente os componentes estruturais intrínsecos de uma série temporal, sem a necessidade de métodos *ad hoc*. Assim, oferece à metodologia de análise de séries temporais epidemiológicas de poluição do ar e efeitos na saúde mais uma poderosa ferramenta para detecção e estimação dos efeitos potencialmente nocivos à saúde humana.