



Mauro Lawall Evaristo Carlos

**Uma Investigação Estatística de
Modelos para Séries Temporais de
Dados de Contagem: Modelo GARMA e
Modelo Poisson Gama em Espaço
de Estado**

Dissertação de Mestrado

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica da PUC-Rio.

Orientador: Prof. Cristiano Augusto Coelho Fernandes

Rio de Janeiro
Novembro de 2006



Mauro Lawall Evaristo Carlos

**Uma Investigação Estatística de
Modelos para Séries Temporais de
Dados de Contagem: Modelo GARMA e
Modelo Poisson Gama em Espaço
de Estado**

Dissertação apresentada como requisito parcial para obtenção do grau de Mestre pelo Programa de Pós-Graduação em Engenharia Elétrica do Departamento de Engenharia Elétrica do Centro Técnico Científico da PUC-Rio. Aprovada pela Comissão Examinadora abaixo assinada.

Prof. Cristiano Augusto Coelho Fernandes
Orientador

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Álvaro de Lima Veiga Filho

Departamento de Engenharia Elétrica – PUC-Rio

Prof. Hélio Cortes Vieira Lopes

Departamento de Matemática – PUC-Rio

Prof. José Eugenio Leal

Coordenador Setorial do Centro

Técnico Científico

Rio de Janeiro, 24 de novembro de 2006

Todos os direitos reservados. É proibida a reprodução total ou parcial do trabalho sem autorização da universidade, do autor e do orientador.

Mauro Lawall Evaristo Carlos

Mestre pelo Programa de Pós-graduação em Engenharia Elétrica – novembro de 2006 - Engenharia Elétrica da PUC-RIO; Engenheiro Eletricista: ênfase em Telecomunicações e Sistemas de Apoio a Decisão - julho de 2004 -Engenharia Elétrica da PUC-RIO; Emprego atual: Analista de Risco - Telos Fundação Embratel de Seguridade Social

Ficha Catalográfica

Carlos, Mauro Lawall Evaristo

Uma investigação estatística de modelos para séries temporais de dados de contagem : modelo GARMA e modelo Poisson gama em espaço de estado / Mauro Lawall Evaristo Carlos ; orientador: Cristiano Augusto Coelho Fernandes. – 2006.

85 f. ; 30 cm

Dissertação (Mestrado em Engenharia Elétrica) – Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2006.

Inclui bibliografia

1. Engenharia Elétrica – Teses. 2. Poisson. 3. Binomial negativa. 4. Espaço de estado. 5. GARMA I. Fernandes, Cristiano Augusto Coelho. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Engenharia Elétrica. III. Título.

CDD: 658.5

Agradecimentos

Agradeço primeiramente a CAPES pelo fornecimento de uma bolsa que tornou possível a realização deste trabalho e a meu orientador Cristiano Fernandes pela orientação.

Agradeço a meus pais Hudson Queiroz Evaristo Carlos e Regina Lúcia Lawall Evaristo Carlos, pelo apoio incondicional, amor e ensinamentos.

Aos grandes amigos Horácio V. A. Neto, Luiz Ribas, André Fadul, André Galdino, Fernando Pizzaro pelo apoio e incentivo, bagunças... que fizeram o período de Mestrado ficar menos árido.

Ao pessoal da favelinha (L-615) que compartilharam horas de estudo.

Ao pessoal da graduação (TODOS). Especialmente para: Clara Muniz, pelo carinho e apoio incondicional e Luiz Fernando C. R. (famoso LF) pela amizade, sinceridade, financiamentos de almoços e cervejas, cafezinhos, e risadas...muitas risadas.

Ao pessoal do Pires (Rainha do Mar) que sempre serviu o melhor almoço e a cerveja gelada mais gostosa da Gávea. Especialmente para: Gonça e Fernando pela amizade e que infinitas vezes permitiram pendurar a conta no prego. Aos Marcolinos e ao Lú pela amizade e por fazerem a cerveja chegar até a mesa.

Às “meninas da secretaria” Marcia Guerra, Alcina e Ana pelas risadas e incentivos nas horas difíceis.

Ao Fernando Hey (lhe devo muito), Fernando Fontes e Pedro Labouriau pela amizade e por compartilhar a mesma corda em diversas paredes nas alturas.

Aos professores Álvaro Veiga pelos ensinamentos e sinceridade na hora da defesa e ao Hélio Côrtes que também participou da defesa. Gostaria de agradecer também a outros professores (do departamento de Elétrica ou não) que de alguma forma me ajudaram: Marco da Silveira pela credibilidade a mim creditada, ao Marco Aurélio pelo apoio e amizade (especialmente no dia da defesa), Adriana Trix por ter aberto as portas na hora da dúvida (devia ter entrado...), Gilson Brand Baptista por ter acreditado em mim desde cedo, Moisés Szwarcman pela amizade, sinceridade e ensinamentos.

Ao Sergio (Van de Graaff) pelos ensinamentos e amizade

Um agradecimento especial para o Eduardo Fonseca Mendes (Duda) que muito me ensinou, ajudou e incentivou (sem seu apoio esse trabalho não existiria).

Resumo

Carlos, Mauro Lawall Evaristo; Fernandes, Cristiano Augusto Coelho (Orientador). **Uma Investigação Estatística de Modelos para Séries Temporais de Dados de Contagem: Modelo GARMA e Modelo Poisson Gama em Espaço de Estado**. Rio de Janeiro, 2006. 85p. Dissertação de Mestrado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

O presente trabalho tem como objetivo principal investigar por meio de simulação Monte Carlo algumas propriedades estatísticas dos modelos GARMA (*Generalized Autoregressive Moving Average*) para séries temporais de dados de contagem. Os modelos GARMA são uma extensão dos Modelos Lineares Generalizados de McCullagh e Nelder para situações de dados dependentes, caracterizando-se pela adição de um termo extra ao preditor linear, o qual passa a incorporar termos autoregressivos (AR) e de médias móveis (MA). As propriedades estatísticas investigadas foram às condições de estacionariedade dos modelos GARMA e os critérios de identificação da ordem (p,q) dos polinômios AR e MA que definem o modelo. Os resultados encontrados indicam que os critérios AIC BIC e Hannan-Quin utilizados foram razoavelmente eficazes na identificação da ordem dos modelos e que as condições de estacionariedade estabelecidas empiricamente em termo de restrições no espaço paramétrico são bastante complexas exigindo um estudo mais detalhado. Como objetivo secundário testamos os modelo GARMA em séries reais, ajustando os modelos GARMA-Poisson e GARMA-Binomial Negativa ao número de caso de poliomielite nos EUA e ao número de infartos do miocárdio no município do Rio de Janeiro. Os resultados indicam que os modelos foram capazes de explicar, de forma econômica, a variação destas séries.

Palavras-chave

Poisson, Binomial Negativa, Espaço de Estado, GARMA.

Abstract

Carlos, Mauro Lawall Evaristo; Fernandes, Cristiano Augusto Coelho (Advisor). **A Statistical Investigation on Time Series Models for Count Data: GARMA Model and the State Space Poisson Gamma Model.** Rio de Janeiro, 2006. 85p. MSc Dissertation – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

The main objective of this dissertation is to investigate, using Monte Carlo simulations, some statistical properties of GARMA (*Generalized Autoregressive Moving Average*) models for time series of count data. GARMA models are extensions of the Generalized Linear Models to dependent data, in which autoregressive (AR) and/or moving average (MA) terms are incorporated into the linear predictor. The statistical properties targeted in our investigation were the model stationarity conditions and the identification criteria for selection of model orders, the lag structure (p,q) associated with the AR and MA terms. Our results suggest that AIC, BIC and Hann-Quinn criteria worked relatively well in identifying the model order, and that the conditions for stationarity established empirically in terms of parameter space restrictions were not totally conclusive, requiring further investigation. As a secondary objective we tested the model against real data, by fitting both a GARMA-Poisson and a GARMA-Negative Binomial to the series of number of cases of poliomyelitis on the US and the number of heart-attacks in Rio de Janeiro city. The results we found indicate that these models were able to explain, in a parsimonious way, the variation of both series.

Keywords

Poisson, Negative Binomial, State Space, GARMA.

Sumário

| | |
|---|----|
| 1. Introdução | 10 |
| 2. Generalized Autoregressive Moving Average Models (GARMA) | 14 |
| 2.1. Generalized Linear Models (GLM) | 14 |
| 2.2. Definição do modelo GARMA | 16 |
| 2.3. Modelo GARMA-POISSON | 19 |
| 2.4. Modelo GARMA-BINOMIAL NEGATIVA (GARMA-NB) | 20 |
| 2.5. Estimação do vetor de hiperparâmetros | 21 |
| 2.6. Estacionariedade dos modelos GARMA | 22 |
| 2.6.1. Modelos com função de ligação identidade | 22 |
| 2.6.2. Modelos com função de ligação diferente da identidade. | 25 |
| 3. Estacionariedade e Identificação via Simulações Monte Carlo | 28 |
| 3.1. Estacionariedade | 28 |
| 3.2. Identificação pela inspeção visual da função de autocorrelação e da função de autocorrelação parcial | 35 |
| 3.3. Identificação utilizando critérios de informação | 39 |
| 4. Aplicações | 49 |
| 4.1. Número de casos de poliomielite nos Estados Unidos | 49 |
| 4.1.1. Aplicação do modelo GARMA-Poisson | 51 |
| 4.1.2. Aplicação do modelo GARMA-NB | 52 |
| 4.2. Número de enfartos em maiores de 65 anos na cidade do Rio de Janeiro | 59 |
| 4.2.1. Aplicação do modelo GARMA-Poisson | 60 |
| 4.2.2. Aplicação do modelo GARMA-NB | 63 |
| 5. Resultados e conclusões | 71 |
| ➤ Estacionariedade | 71 |
| ➤ Identificação | 73 |
| ➤ Aplicação em séries reais | 74 |
| Referências bibliográficas | 76 |
| Apêndices | 79 |

Lista de figuras

| | |
|--|----|
| Figura 1: Gráficos comparativos relativos às distribuições Normal, Poisson, e Binomial Negativa | 12 |
| Figura 2: Esquema de simulações para investigação da região de estacionariedade dos processo GARMA para dados de contagem | 29 |
| Figura 3: Exemplo de distribuições empíricas incondicionais nos tempos $t=100$, $t=150$ e $t=300$ para o modelo GARMA-Poisson com apenas termos autoregressivos | 30 |
| Figura 4: Exemplo ilustrativo do esquema utilizado para apresentação dos resultados da análise de estacionariedade. Os hiperparâmetros utilizados ϕ_1 , ϕ_2 , θ_1 , θ_2 e k estão apontados ao lado da figura (de baixo para cima) | 31 |
| Figura 5: Análise de estacionariedade para séries geradas pelo modelo GARMA-Poisson. | 32 |
| Figura 6: Análise de estacionariedade para séries geradas pelo modelo GARMA-NB | 34 |
| Figura 7: FAC e FACP para algumas séries geradas. De cima para baixo: GARMA(1,0)-Poisson, GARMA(2,0)-NB, e GARMA(0,1)-Poisson | 36 |
| Figura 8: FAC e FACP médias para séries geradas pelos modelos GARMA(1,0) -NB com parâmetros: $\phi_1=0.6$ e $k=3$ | 37 |
| Figura 9: FAC e FACP médias para séries geradas pelos modelos GARMA(2,0) -Poisson com parâmetros: $\phi_1=0.6$ e $\phi_2=0.2$ | 37 |
| Figura 10: FAC e FACP médias para séries geradas pelo modelo GARMA(0,1) -NB com parâmetros: $\theta_1=0.6$ e $k=3$ | 38 |
| Figura 11: FAC e FACP médias para séries geradas pelos modelos GARMA(2,1) -NB com parâmetros $\phi_1=0.2$, $\phi_2=0.2$, $\theta_1=0.6$ e $k=3$ | 38 |
| Figura 12: FAC e FACP médias para séries geradas pelos modelos GARMA(2,1) -NB com parâmetros $\phi_1=0.2$, $\phi_2=0.6$, $\theta_1=0.2$ e $k=3$ | 38 |
| Figura 13: FAC e FACP médias para séries geradas pelos modelos GARMA(1,1) -Poisson com parâmetros: $\phi_1=0.2$ e $\theta_1=0.6$ | 39 |
| Figura 14: Ilustração do uso da tabela 3 | 43 |
| Figura 15: Série do número de casos de poliomielite e histograma da série | 49 |
| Figura 16: FAC e FACP para a série de poliomielite nos Estados Unidos da América | 50 |
| Figura 17: Série de casos de poliomielite nos EUA e média estimada; resíduos do modelo GARMA(1,2)-Poisson aplicado à série | 51 |
| Figura 18: QQplot e histograma dos resíduos do modelo GARMA(1,2) aplicado à série | 52 |
| Figura 19: FAC e FACP dos resíduos do modelo GARMA(1,2) aplicado à série | 52 |
| Figura 20: Série de casos de poliomielite nos EUA e média estimada; resíduos do modelo GARMA(1,2)-NB aplicado à série | 53 |
| Figura 21: QQplot e histograma dos resíduos do modelo GARMA(1,2)-NB aplicado à série | 54 |
| Figura 22: FAC e FACP dos resíduos do modelo GARMA(1,2) aplicado à série | 54 |

| | |
|---|----|
| Figura 23: Série de casos de poliomielite nos EUA e média estimada; resíduos do modelo ARMA(2,0) aplicado à série | 57 |
| Figura 24: QQplot e histograma dos resíduos do modelo ARMA(2,0) aplicado à série | 57 |
| Figura 25: FAC e FACP dos resíduos do modelo ARMA(2,0) aplicado à série | 57 |
| Figura 26: Série do número de enfartos no município do Rio de Janeiro e histograma da série | 58 |
| Figura 27: FAC e FACP da série do número de enfartos no município do Rio de Janeiro | 59 |
| Figura 28: Série de número de casos de enfarto no município do Rio de Janeiro e média estimada da série; resíduos do modelo GARMA(1,0)-Poisson ajustado à série de enfartos | 62 |
| Figura 29: QQplot e histograma dos resíduos do modelo GARMA(1,0)-Poisson ajustado à série de enfartos | 62 |
| Figura 30: FAC e FACP dos resíduos do modelo GARMA(1,0)-Poisson ajustado à série de enfartos | 62 |
| Figura 31: Histograma de quatro séries, cada uma gerada por um ruído com distribuição Binomial Negativa com $p=0.90$ e respectivamente $k=3$, $k=10$, $k=25$, e $k=50$ | 65 |
| Figura 32: Série de enfartos e média estimada da série; resíduos do modelo GARMA(1,0)-NB ajustado à série | 65 |
| Figura 33: QQplot e histograma dos resíduos do modelo GARMA(1,0)-NB ajustado à série de casos de enfartos no município do Rio de Janeiro | 66 |
| Figura 34: FAC e FACP dos resíduos do modelo GARMA(1,0)-NB ajustado à série de casos de enfartos no município do Rio de Janeiro | 66 |
| Figura 35: Série diária da quantidade de material particulado (10ppm) presente no ar referente ao mesmo período da série de enfartos no município do Rio de Janeiro | 68 |
| Figura 36: Nível diário de monóxido de carbono referente ao mesmo período da série de casos de enfartos no município do Rio de Janeiro | 68 |

Lista de tabelas

| | |
|--|----|
| Tabela 1: Estatísticas descritivas das séries geradas | 12 |
| Tabela 2: Valores dos hiperparâmetros multiplicadores dos termos auto-regressivos (ϕ_1 e ϕ_2) e médias móveis (θ_1 e θ_2) | 41 |
| Tabela 3: Identificação dos dois melhores modelos para séries geradas por GARMA-Poisson | 44 |
| Tabela 4: Identificação dos dois melhores modelos para séries geradas por GARMA-NB com $k=3$ | 45 |
| Tabela 5: Identificação dos dois melhores modelos para séries geradas por GARMA-NB com $k=8$ | 46 |
| Tabela 6: Estatísticas descritivas para a série de poliomielite nos Estados Unidos da América | 50 |
| Tabela 7: Resultados dos critérios BIC, AIC e Hanna-Quinn para a série de casos de poliomielite nos Estados Unidos da América - modelagem GARMA-Poisson | 51 |
| Tabela 8: Resultados dos critérios BIC, AIC e Hanna-Quinn para a série de casos de poliomielite nos Estados Unidos da América - modelagem GARMA-NB | 53 |
| Tabela 9: Estimativa dos hiperparâmetros encontrados ao se ajustar os modelos GARMA(1,2)-Poisson e o modelo GARMA(1,1)-NB à série de casos de poliomielite nos EUA | 55 |
| Tabela 10: Estimativa dos hiperparâmetros encontrados ao se ajustar os modelos GARMA(1,2)-Poisson e o modelo GARMA(1,1)-NB (sem o termo referente à tendência) à série de casos de poliomielite nos EUA | 56 |
| Tabela 11: Análise dos modelos identificados e modelo ARMA para a série de casos de poliomielite nos Estados Unidos | 58 |
| Tabela 12: Estatísticas descritivas da série do número de enfartos no município do Rio de Janeiro | 59 |
| Tabela 33: Resultados dos critérios BIC, AIC e Hanna-Quinn para a série de número de enfartos no município do Rio de Janeiro - modelagem GARMA-Poisson | 60 |
| Tabela 44: Hiperparâmetros estimados para o modelo GARMA(1,0)-Poisson ajustado à série de número de casos de enfarto no município do Rio de Janeiro | 61 |
| Tabela 55: Resultados dos critérios BIC, AIC e Hanna-Quinn para a série de número de enfartos no município do Rio de Janeiro - modelagem GARMA-NB | 63 |
| Tabela 66: Hiperparâmetros estimados para o modelo GARMA(1,0)-NB ajustado à série de número de casos de enfarto no município do Rio de Janeiro | 64 |
| Tabela 77: Estimativa dos hiperparâmetros ao se ajustar os modelos: GARMA(1,0)-Poisson com a variável explicativa PM_{10} ; GARMA(1,0)-Poisson com a variável explicativa CO; GARMA(1,0)-NB com a variável explicativa PM_{10} | 69 |