

Universidade Federal do Rio Grande do Sul
Instituto de Matemática
Departamento de Estatística



Anais

IV SEMANÍSTICA

IV Semana Acadêmica do Departamento de Estatística

da UFRGS

<http://www.ufrgs.br/semanistica>

Porto Alegre - 20 a 22 de outubro de 2014

O Processo Estocástico k-Factor GARMA, Um Estudo de Simulação e Aplicação em Tempestades Solares

Ian Meneghel Danilevicz^{1 3}

Cleber Bisognin^{2 3}

Resumo: Estudo teórico e aplicado do processo estocástico *k-factor* GARMA($\rho, \mathbf{u}, \lambda, q$). Considerações teóricas, apresentação de resultados de simulações e resultados de aplicação a dados reais. Proposição de diferentes estimadores para os parâmetros $\rho, \mathbf{u}, \lambda, q$ e análise do comportamento desses estimadores para diferentes cenários, nos quais o comprimento de k e a quantidade de *outliers* aditivos foram testados de diversas formas. A série real analisada foi a famosa sunspot.year (disponível no software estatístico R), nela foram aplicados os estimadores propostos, observado a qualidade dos resíduos e proposto um modelo candidato para previsões.

Palavras-chave: *k-factor* GARMA, longa dependência, *outliers*, processo estocástico.

1 Introdução

No estudo de séries temporais tem crescido o interesse por processos de longa dependência (Bisognin), assim como pela identificação de outliers (Tsay). Apresentamos o modelo *k-factor* GARMA, uma generalização dos processos SARFIMA, com a propriedade da longa dependência. Apresentamos algumas considerações teóricas, resultados de simulação de estimadores e uma aplicação em uma série temporal real.

Definição de *k-factor* GARMA

Seja X_t um processo estocástico que satisfaz a equação

$$\varphi(\beta) \prod_{j=1}^k (1 - 2u_j\beta + \beta^2)^{\lambda_j} (X_t - \mu) = \theta(\beta)\varepsilon_t$$

¹ UFRGS - Universidade Federal do Rio Grande do Sul. Email: iandanilevicz@gmail.com

² UFRGS - Universidade Federal do Rio Grande do Sul. Email: cbisognin@ufrgs.br

³ Agradecimentos a bolsa de IC da PIC CAPES.

na qual k é um inteiro finito, $|u_j| \leq 1$ e λ_j é um número fracionário, para $j=1, \dots, k$, μ é a média do processo, $\{\varepsilon_t\}_{t \in \mathbb{Z}}$ é um processo de ruído branco e $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de grau p e q , respectivamente.

Definição de *outlier* Aditivo

Seja Z_t um processo estocástico que satisfaz a equação

$$Z_t = X_t + V_t$$

na qual X_t é o processo k -factor GARMA visto acima e V_t é constituído de variáveis aleatórias independentes e identicamente distribuídas com distribuição dada por

$$H_v = (1 - c)\delta_0 + cG$$

para a qual $0 \leq c \leq 1$, δ_0 é uma distribuição degenerada na origem e G é uma distribuição arbitrária. Além disso, os processos V_t e X_t são independentes. Portanto, Z_t é um processo contaminado por misturas de AO (*Additive Outliers*).

2 Metodologia

Num primeiro momento, realizamos um estudo de simulação de Monte Carlo, no qual submetemos os estimadores a diferentes cenários e comparamos os resultados alcançados por cada um deles. Como estatísticas para avaliar o desempenho de cada um dos estimadores temos a média, o vício, o erro quadrático médio (EQM) e a variância.

Os cenários gerados foram séries temporais contaminadas por *outliers* aditivos (AO), ou seja, séries nas quais há observações com valores discrepantes (erro de aferição por exemplo), mas que não interferem nos resultados seguintes. Também diversificamos o percentual e a grandeza da contaminação na série com o intuito de perceber a robustez dos estimadores, ou seja, até que ponto os estimadores produzem estimativas com baixos valores de vício.

Simulamos séries de tamanho mil com mil e duzentas repetições, p igual a zero, q igual a zero e k igual a dois, pois nosso foco é concentrar esforços no estudo dos estimadores para u e λ . Dessa forma temos processos k -factor GARMA(p, λ, u, q) com $\lambda = \{\lambda_1, \lambda_2\}$ e $u = \{u_1, u_2\}$.

Estimadores Propostos

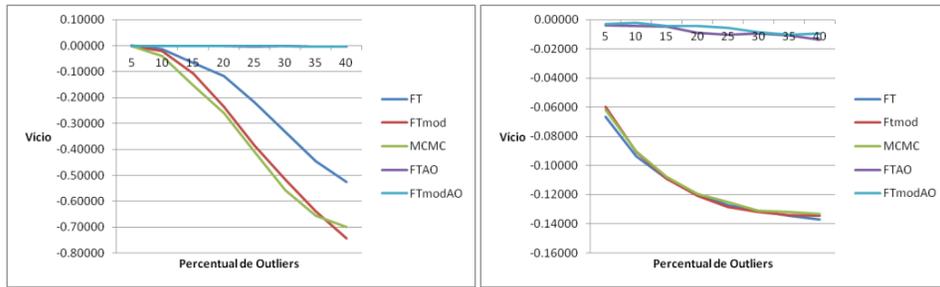
Direcionamos nosso estudo para os estimadores paramétricos, pois os estimadores semi-paramétricos não conseguem estimar todos os parâmetros. Os estimadores analisados são os respectivos:

1. **FT** - extensão do modelo proposto por Fox e Taquq (1995) para os processos ARFIMA, o qual utiliza uma aproximação para a matriz de autocovariância sugerida por Whittle (1954) e a função periodograma;
2. **FTmod** - segue o estimador FT, mas a função a ser minimizada é calculada para mais frequências do que as de Fourier;
3. **MCMC** - utiliza o método de Metropolis-Hasting no estimador FT;
4. **FTAO** - análogo ao estimador FT, porém acrescenta-se uma função de densidade espectral capaz de estimar a magnitude e o percentual de *outliers* no processo;
5. **FTmodAO** - análogo ao estimador FTmod, porém acrescenta-se uma função espectral semelhante a executada no FTAO.

Em um segundo momento, passamos para o desafio de trabalhar com dados reais. Primeiramente analisamos a função periodograma da série temporal, afim de observar quantos picos a função apresenta no intervalo menor do que π . Esse procedimento nos indicará qual o provável melhor valor de k . Depois disso, fazemos um estudo de *outliers*, se há, se não há, qual a magnitude deles. Depois disso, propomos modelos e analisamos os resíduos, principalmente o gráfico de correlação e correlação parcial, para selecionar um modelo competitivo e se necessário acrescentar parâmetros φ e θ . Passado esse modelo estamos mais seguros para propor um modelo de previsão.

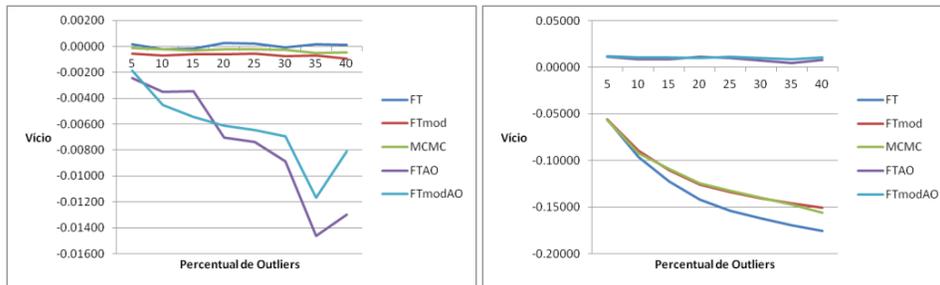
3 Resultados

Simulamos séries com *outliers* com variância da variável de contaminação τ^2 igual a 2, 5 e 10. e probabilidades de Contaminação $c=\{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40\}$. Para economizar espaço, mostramos apenas os resultados para τ^2 igual a cinco, mas o comportamento dos estimadores é semelhante. Os primeiros gráficos são dos vícios dos estimadores e na próxima página estão os respectivos EQMs.



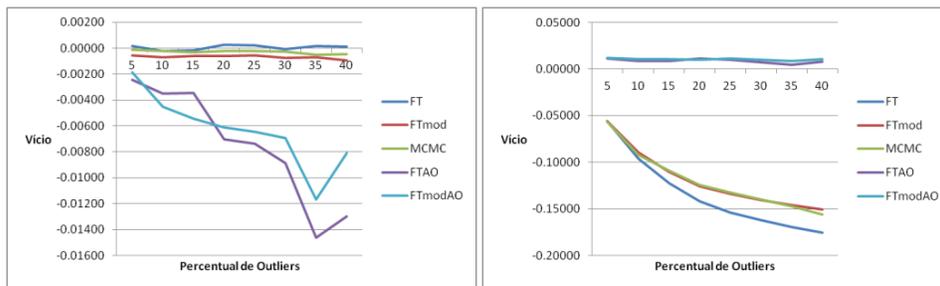
vício u_1

vício u_2



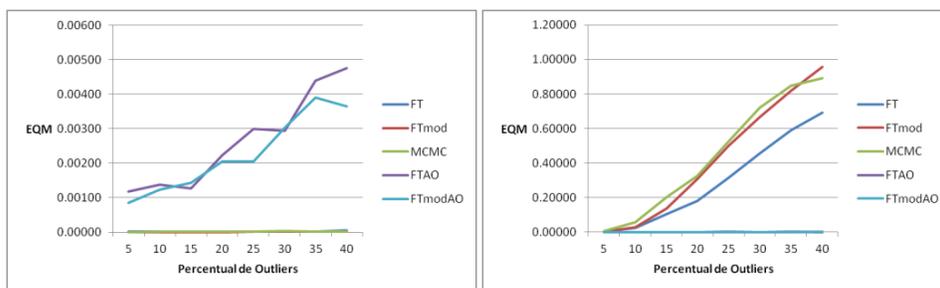
vício λ_1

vício λ_2



EQM u_1

EQM u_2



EQM λ_1

EQM λ_2

Figura 1: Resultados de Simulação

Além dos resultados de simulação (Figura 1) temos os resultados para séries reais. Primeiramente as questões mais genéricas da série temporal *sunspot* (Figura 2), ou seja, uma identificação de *outliers* e um periodograma suavizado.

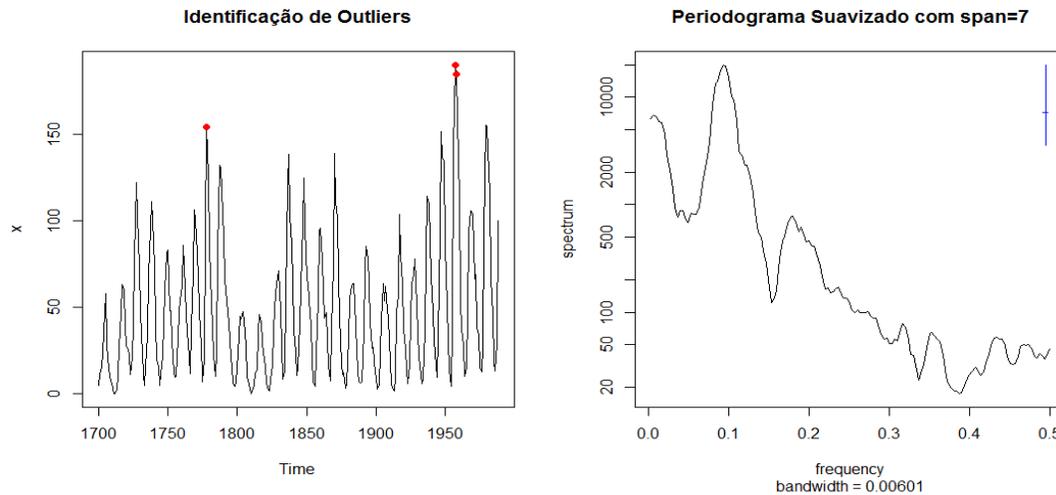


Figura 2: Série Real: *sunspot*.

Testamos os cinco tipos de estimadores para um modelo com $k=3$ e $p=0=q$. Uma vez que os resíduos (ver Figura 3) não estavam muito bons acrescentamos parâmetros p , chegando em $p=3$, o qual produziu resultados minimamente satisfatórios de resíduos. Embora nem todos os *lags* estejam dentro dos intervalos de confiança, o preço a ser pago para ajustá-los seria demasiado caro, ou seja, teríamos que colocar muitos ϕ s e/ou θ s no modelo. O que implica em aumentar de forma expressiva o erro padrão.

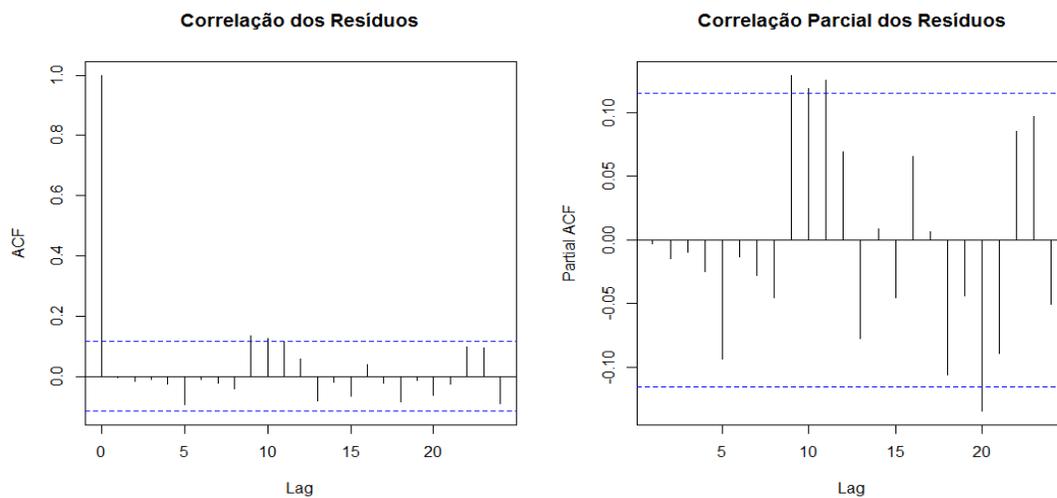


Figura 3: Autocorrelação e Autocorrelação Parcial dos Resíduos.

Por fim temos o modelo *k-factor* GARMA com $k=3$ e $p=3$, $u=\{0.80, 0.84, 0.99\}$, $\lambda=\{0.19, 0.22, 0.23\}$, $\varphi=\{0.01, -0.10, -0.20\}$. estimado pelo método de FTmod. O qual acerta as 25 observações futuras da série, noutras palavras, os valores reais da série estão nos intervalos de confiança, isso foi possível pois buscamos os valores de tempestades na superfície do sol, para anos subsequentes ao disponíveis no R. A Figura 4 a seguir apresenta a previsão da série real para 25 anos a frente.

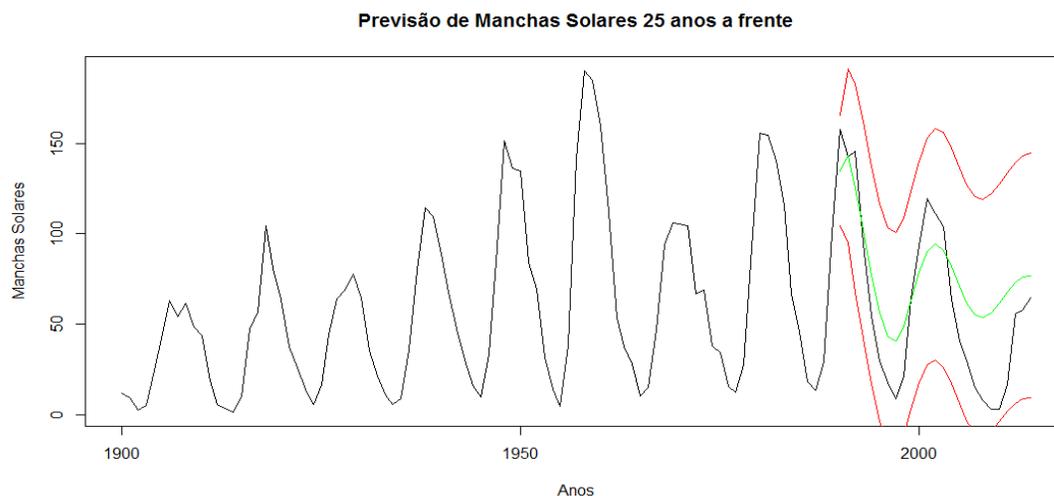


Figura 4: Previsão da Série Real.

4 Conclusões

Claramente temos dois padrões entre os estimadores, um primeiro grupo formado pelos FT, FTmod e MCMC, e um segundo grupo formado pelos FTAO e FTmodAO. Os vícios de FTAO e FTmodAO foram bem menores do que os de FT, FTmod e MCMC com exceção de u_2 , para o qual acontece o inverso. Já no tocante aos EQMs, temos novamente resultados análogos, pois os FTAO e FTmodAO foram melhores do que os estimadores não preparados para os outliers aditivos com exceção do EQM para de λ^2 , no qual o resultado novamente foi o contrário.

Uma primeira aproximação com dados reais foi bastante satisfatória, uma vez que estimamos boas previsões para 25 anos de comportamento do sol. No entanto os ICs podem ter a sua amplitude reduzida uma vez que o desvio padrão dos resíduos diminua. Isso será possível se identificarmos um modelo com ainda melhor aderência do que o atual.

5 Referências

- [1] BISOGNIN; C., LOPES, S.R.C., “Properties of Seasonal Long Memory Processes”, Mathematical and Computer Modelling, 49, 1837-1851, 2009.

[2] Taqqu, M.S.; V. Teverovsky e W. Willinger, “Estimators for Long Range Dependence: an Empirical Study”, *Fractals*, Vol. 3(4), pp. 785-798, 1995.

[3] TSAY, Ruey S., “Outliers, Level Shifts, and Variance Changes in Time Series”, *Journal of Forecasting*, Vol. 7, I-20, 1988.

[4] Whittle, P., “Hypothesis Testing in Time Series Analysis”, *Journal of the American Statistical Association*, Vol. 49, No. 265, pp. 197-200, 1954.