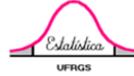




UFRGS
UNIVERSIDADE FEDERAL
DO RIO GRANDE DO SUL

UNIVERSIDADE FEDERAL DO RIO GRANDE DO
SUL
INSTITUTO DE MATEMÁTICA
DEPARTAMENTO DE ESTATÍSTICA



Processos Estocásticos *k-Factor* GARMA, um estudo de
simulação e aplicação à série SOI (*Southern Oscillation
Index*)

Autor: Ian Meneghel Danilevicz
Orientador: Professor Dr. Cleber Bisognin

Porto Alegre, 27 de Novembro de 2015.

Instituto de Matemática
Departamento de Estatística

**Processos Estocásticos *k-Factor* GARMA, um
estudo de simulação e aplicação à série SOI
(*Southern Oscillation Index*)**

Autor: Ian Meneghel Danilevich

Trabalho de Conclusão de Curso
apresentado para obtenção
do grau de Bacharel em Estatística.

Banca Examinadora:
Professor Dr. Cleber Bisognin (orientador)
Professor Dr. Marcio Valk

Porto Alegre, 27 de Novembro de 2015.

Este artigo será submetido a **Journal of Forecasting**.

Agradecimentos

Gostaria de começar agradecendo ao meu orientador Cleber, é um pouco clichê elogiar o orientador, mas o Cleber foi muito parceiro mesmo. Quase três anos de bolsa de IC, muita paciência dele pra fazer um aluno que ainda não tinha cursado a cadeira de Séries Temporais entender o que é uma SDF. Me incentivou a participar de congressos e me instrumentalizou para descobrir o ambiente da pesquisa acadêmica por mim, me deixou quebrando a cabeça com códigos várias tardes, mas também pegou junto para reprogramar funções com erros. Podeira dizer qualquer coisa dele, menos que ele não se dedicou a me fazer crescer, e não se sentiu nem um pouco desprestigiado quando eu disse que não pretendia seguir na UFRGS. Pra mim, um grande professor é aquele que não fabrica cópias de si mesmo.

Agradeço ao Marcio Valk que aceitou ser banca deste TCC, a leitura e crítica por um especialista em Séries Temporais é sempre de extrema valia.

Também gostaria de agradecer ao Luciano, que me deu a dica dessa graduação exótica, e a professora Jandyra que me atendeu no NAE de forma tão acolhedora e eficiente que decidi me tornar como ela. Decidi voltar ao primeiro semestre, mas tive sorte de fazer isso acompanhado de um velho amigo, estudar Cálculo e Álgebra Linear com o Pietro tornou o desafio leve e às vezes até divertido. Decidi voltar a lutar karatê e encontrei um grande parceiro, que me apresentou switch case em C, Ubuntu, Digger e muito heavy metal do leste europeu, este último eu dispensei, mas jamais a tua amizade Psico Mantis, meu brâmane de primeira geração. E assim que me elegerem presidente, tu é o ministro da Fazenda.

Vanessa foi minha professora de introdução a R, Inferência Clássica e Bayesiana, Modelos Mistos e Inferência Bayesiana, certamente uma das culpadas pela minha simpatia por bayesianismo. Lisiane, Fernando e Flávio, também foram professores muito presentes na minha formação. Cada um deles com estilo bastante diferente, mas ótimos. Lisi, nos adotou mesmo, sempre querendo saber o que estávamos planejando e como executaríamos nossos planos de vida. Fernando sempre nos estimulou a ter autonomia, a não esperar, a ler os livros e não apostilas e a seguir estudando sempre. Flávio, uma referência nacional de pesquisa, foi sempre um professor acessível e disposto a nos ajudar com dúvidas e tomadas de decisão.

Aos meus inusitados amigos físicos, Nathan, Eduardo e Marília. Nathan, Nathanielson bróder de mais de 5 mil kms pelo hinterland brasileiro, mestre tapioqueiro e bobão de sempre. Eduardo e Marília meus heróis que me trouxeram um Degroot do além mar por duas libras e muito peso na mochila, amo vcs. E dos meus amigos loucos, poucos são tão loucos e lúcidos como Shambavi, mudei minha vida depois de te conhecer. E estou bem feliz com a troca, também conheci Tandava e Shivanada a partir de ti. Meus sinceros Om Namah Shivaya.

Ao coração bravo e afetuoso que só Dani tem, obrigado pelos vidrinhos de tempero que cê deixou por aqui depois que foi pra Paraíba. Ao Tomas, meu amigo bagunceiro, favelaboy who accept any challenge, e sua exótica forma de cozinhar tudo na mesma panela e "lavar" os pratos. The Wolfi's Team, my friends Julia and James, I sincerelly wish to meet you again, it doesn't matter where. It was a great luck to find you, my plains were just have a nice time, but you turned it into the trip.

Aos amigos de mate Rodrigo Colla, Galo Místico e Gab Dienstman, pois a vida exige momentos prosaicos e contação de causos. Aos novos amigos de Lab, Vitor e Cristiano, não fossem eles apertar uns botões na sala secreta, teria perdido uma semana de simulações. Ao Mathias, com quem pude trocar algumas ideias para estas simulações quando ele estava fazendo o seu TCC sobre os processos ARFIMA. Ao Lucas Gabriel e a Carol pelos

café da Física com Tau ou Roller Derby. À Elisa, parceira de pedal, preciosista com a espessura da tapioca e descolada com a ABNT da vida. Depois de tanto conversar sobre estatística multivariada, vou querer escrever um artigo contigo, pode até ser sobre as cactáceas do Paraguay.

À família certinha e convencional que não tenho, adoro vcs tudo. Aos couchsurfers, warmshowers, woofers e bewelcomes que me hospedaram e/ou pararam aqui em casa e, principalmente, a minha irmã que me apresentou essas ferramentas entre muitas outras, minha queridíssima irmã caçula, minha mestre.

Não fosse essa galera, que mais parece um crowdfund do que um agradecimento de TCC, fazer estatística seria uma chatice. No entanto, meus dois agradecimentos especiais são para Krishna e Karina. É até difícil saber o que dizer de vocês, as caronas de moto, os rolês de bici, os planos de ir pra China, estudar Processos Estocásticos em Janeiro de 40°, as tequilas, as mudanças, o "decide a minha vida pq tá foda", os livros, as séries, os planos e contra-planos, a cumplicidade. Se abrir o jornal destroi o otimismo, vocês são minha resistência à depressão, pelas bobagens que dizem e pelas ideias de que ainda podemos fazer algo.

Ao Tom Zé, seus acordos e deboches foram meu café para escrever e reescrever projetos de simulação sem fim, aquele combustível pro Jimmy aqui não se render. Ao Cesaro e o Herbert que passaram madrugadas invertendo matrizes.

Processos Estocásticos *k-Factor* GARMA, um estudo de simulação e aplicação à série SOI (*Southern Oscillation Index*)

Ian Meneghel Danilevicz*, Cleber Bisognin**¹

** Departamento de Estatística da UFRGS ** Departamento de Estatística da UFRGS

Contato: iandanilevicz at gmail.com, cbisognin at ufrgs.br

Homepage: <https://sites.google.com/site/iandanilevicz/>

Resumo Neste artigo estudam-se os processos estocásticos *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminados ou não por *outliers*. Sendo que essas interferências podem ser de dois tipos, *outliers* aditivos e *outliers* inovadores. Propõe-se três estimadores paramétricos inspirados nos modelos apresentados por Fox e Taqqu, sendo o primeiro para séries não contaminadas e os dois últimos cada um para um dos referidos tipos de contaminação. No entanto, como avaliar se uma série está contaminada e ainda por cima por qual tipo de *outlier*? Para traçar um critério objetivo para essa questão, adaptou-se o algoritmo SODA para os processos *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). No entanto, o SODA é útil não só para identificar *outliers*, como também para retirá-los da série. Tem-se assim várias formas de enfrentar o problema de contaminação, como tratam-se de adaptações para um tipo de processo estocástico ainda não testado, convém proceder com simulações antes de aplicar em séries reais. Finalmente, depois de conferir a eficiência de cada um dos procedimentos propostos por simulação, aplica-se a referida metodologia em uma série real, qual seja, a SOI, *Southern Oscillation Index*, uma variável chave no entendimento do efeito ENSO, *El Niño Southern Oscillation*, ou mais simplesmente, *El Niño*.

Palavras-Chave Longa dependência, processos estocásticos *k-Factor* GARMA, *outliers*, *Southern Oscillation Index*

Abstract In this paper were discuss stochastic process as *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), which can be contaminated or not for outliers. The interference caused by outliers might be from two kind, the additive outliers or the innovative outliers. Here are propose three parametric estimation methods, all them inspired on the Fox and Taqqu work, so the first method is design to estimate the parameters of some stochastic process without contamination and the others each one to one class of outlier. However, in a practical sense, how a research can know if there is or not outliers in a serie? To have some objective criteria to categorize each observation as an outlier or not, the SODA algorithm was adapted to the stochastic process *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). However, SODA is even more useful that just classify outliers, it is an algorithm to clean the time series. So there is many ways to solve the outliers problem, but all them are old methods to a new stochastic process, so it is earnest procedure checking by simulation first, i. e., after test the related methods by simulation, there is more trustful on apply the discussed tecnics on some real time series as the SOI, Southern Oscillation Index, a key variable to forecast the ENSO phenomenon, *El Niño Southern Oscillation*, which affects great part of the Pacific Ocean and coast cuntries.

Keywords long memory, stochastic processes *k-Factor* GARMA, outliers, Southern Oscillation Index

1 Introdução

Em séries temporais, talvez não exista um modelo tão simples e difundido como o modelo Autoregressivo ou AR(p), que juntamente com os modelos MA(q) e ARMA(p, q) são modelos simples e úteis para séries estacionárias e com curta dependência. Para séries sem estacionariedade foram propostos os processos ARIMA(p, d, q). No entanto, em caso de média ou longa dependência, ver Palma (2007) e Guégan (2005), é mais apropriado eleger modelos como o ARFIMA(p, d, q) em que d não precisa ser igual a 1, mas sim qualquer valor fracionário. Acrescentando-se sazonalidade às séries, tem-se os respectivos modelos SARMA(p, q) \times (P, Q) $_s$, SARIMA(p, d, q) \times (P, D, Q) $_s$ e SARFIMA(p, d, q) \times (P, D, Q) $_s$, que são análogos aos modelos ARMA(p, q), ARIMA(p, d, q) e ARFIMA(p, d, q).

Também é possível traçar estratégias de modelagem de processos estocásticos a partir de um olhar no domínio da frequência ao invés de sempre observar as séries no domínio do tempo. Entende-se que se a função densidade espectral de uma série é ilimitada na frequência zero, um bom candidato para modela-la é o processo

¹Agradecimentos ao CNPq pelo financiamento durante a pesquisa e ao CPC-NOAA por disponibilizar a série SOI gratuitamente à comunidade científica internacional

ARFIMA(p, d, q). No entanto, se a função densidade espectral for ilimitada em alguma frequência G diferente de zero no intervalo $[0, \pi]$ é mais apropriado utilizar um modelo Gegenbauer(u, λ) ou GARMA(p, u, λ, q), este último uma sugestão de Gray et al. (1989), que propõe a junção do modelo ARMA(p, q) e Gegenbauer(u, λ).

Finalmente, Giraitis e Leipus (1995) e, depois, Woodward et al. (1998) estendem os modelos Gegenbauer(u, λ) e GARMA(p, u, λ, q), respectivamente, aos modelos k -Factor Gegenbauer($\mathbf{u}, \boldsymbol{\lambda}$) e k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) para os quais a função densidade espectral é ilimitada para um número finito k de frequências, chamadas de frequências de Gegenbauer ou frequências G , no intervalo $[0, \pi]$.

Apresentado o processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), este será mais formalmente definido na seção seguinte, seção 2, assim como suas principais propriedades. Também será discutido algumas formas de contaminações do processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), visto que identificação e tratamento de *outliers* é, atualmente, uma das questões mais quentes em séries temporais. Aqui serão discutidos os dois principais tipos de *outliers*, já apresentados por Fox (1972), quais sejam *outliers* aditivos, AO, e *outliers* inovadores, IO. Os primeiros consistem em observações anômalas que aparecem pontualmente em um processo estocástico, enquanto os segundos são alterações estruturais na série que influenciam uma dita sub-série.

Visto que o processo está definido, o próximo passo é ter uma forma de estimar os parâmetros do modelo proposto. Na seção 3 apresentam-se três estimadores paramétricos para os modelos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). São eles o estimador FT, FTAO e FTIO, todos eles são inspirados nas técnicas apresentadas por Fox e Taquq (1986) de minimizar uma razão entre a função densidade espectral teórica e sua estimação pelo periodograma para processos ARFIMA(p, d, q) e, por isso, ganharam ao nome o prefixo FT. O primeiro deles, FT, foi desenhado para estimar os parâmetros de um processo teoricamente livre de *outliers*. O segundo, FTAO, é para quando se espera a presença de *outliers* do tipo I, ou *outliers* aditivos. Analogamente, o último, FTIO, é mais indicado para séries com contaminação do tipo II, ou seja, *outliers* inovadores.

Na seção 4, apresenta-se o algoritmo SODA, *Spectrum-based outlier detection algorithm*, um método de identificação e correção de *outliers* em séries temporais. Tem-se a oportunidade assim de retomar a questão de como enfrentar contaminações nas séries, além da escolha de um método de estimação entre os três que foram propostos. Choy (2001) apresenta a técnica SODA para modelos ARMA(p, q), no entanto o SODA é estendido aqui para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) alterando-se a função densidade espectral de um modelo pelo outro. Sendo, no entanto, necessário testar por simulação se os resultados propostos por Choy podem ser estendidos. Tal empreendimento faz parte da seção 5, a qual é dedicada a mostrar os resultados por simulação para os métodos de estimação de parâmetros assim como para avaliar a eficiência do algoritmo SODA para os dois tipos de *outliers*.

A sexta seção é dedicada a mostrar uma aplicação de parte das técnicas discutidas neste artigo em uma série temporal real. *El Niño* ou ENSO, *El Niño Southern Oscillation*, é um evento climático popularmente conhecido por ser de difícil previsibilidade e, portanto, o evento escolhido como desafio ao modelo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Diversos Institutos se dedicam a estudar esse fenômeno e propor previsões que envolvem interpolações de muitas variáveis tais como temperatura, circulação dos ventos, correntes marítimas, pressão atmosférica, entre outras. Uma variável considerada chave nesse estudo é a SOI, *Southern Oscillation Index*, que é, basicamente, a diferença de pressão atmosférica na superfície do mar de Darwin (Austrália) e Tahiti (Polinésia Francesa).

2 Processo k -Factor GARMA

Nesta seção definem-se os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Estes processos são tratados como uma generalização dos processos ARFIMA(p, d, q) e SARFIMA(p, d, q)x(P, D, Q)s e foram inicialmente apresentados por Giraitis e Leipus (1995) e por Woodward et al. (1998).

Definição 2.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo estocástico que satisfaz a equação

$$\phi(\mathcal{B}) \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{\lambda_j} (X_t - \mu) = \theta(\mathcal{B}) \varepsilon_t, \quad (1)$$

onde 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 ruído branco, $\phi(\cdot)$ e $\theta(\cdot)$ são os polinômios de grau p e q , respectivamente definidos por $\phi(z) = \sum_{n=0}^p (-\phi_n) z^n$ e $\theta(z) = \sum_{m=0}^q (-\theta_m) z^m$ então, $\{X_t\}_{t \in \mathbb{Z}}$ é um processo auto-regressivo de média móvel k -Factor Gegenbauer de ordem $(p, \mathbf{u}, \boldsymbol{\lambda}, q)$, onde $\mathbf{u} = (u_1, \dots, u_k)'$ e $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_k)'$, denotado por k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$).

A proposição a seguir apresenta alguns resultados sobre o processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), cuja demonstração pode ser encontrada em Giraitis e Leipus (1995) e Woodward et al. (1998).

Proposição 2.1. *Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) conforme a Definição 2.1. Então, as seguintes afirmações são verdadeiras:*

- I) *O processo $\{X_t\}_{t \in \mathbb{Z}}$ é estacionário se u_j são distintos e $\lambda_j < 0.5$, ou quando $|u_j| < 1$ e $0 < \lambda_j < 0.25$, ou quando $|u_j| = 1$, para $j = 1, \dots, k$, e além disso, todas as raízes do polinômio $\phi(z)$ estão fora do círculo unitário;*
- II) *O processo estacionário $\{X_t\}_{t \in \mathbb{Z}}$ possui longa dependência se satisfaz as condições do item i) desta proposição e, além disso, $\lambda_j > 0$, para $j = 1, \dots, k$;*
- III) *O processo estacionário $\{X_t\}_{t \in \mathbb{Z}}$ é causal se e somente se $\phi(z) \neq 0$, para $|z| \leq 1$;*
- IV) *O processo estacionário $\{X_t\}_{t \in \mathbb{Z}}$ é inversível se e somente se $\theta(z) \neq 0$, para $|z| \leq 1$;*
- V) *A função densidade espectral do processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), ver equação (1), é dada por*

$$f_x(w) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(e^{-iw})|^2}{|\phi(e^{-iw})|^2} \prod_{j=1}^k |2(\cos(w) - u_j)|^{-2\lambda_j}, \quad (2)$$

onde $0 \leq w \leq \pi$ e $G_j = \cos^{-1}(u_j)$ são as chamadas frequências de Gegenbauer.

A definição de densidade espectral apresentada acima desconsidera a presença de interferências na série, tais como *outliers*. No entanto, estes são frequentemente encontrados em séries reais. Bisognin (2007) esclarece que estudar *outliers* requer refletir sobre duas questões. A primeira sobre a classificação dos mesmos, que pode ser a tipologia proposta por Fox (1972), entre *outliers* aditivos ou inovadores. Enquanto a segunda questão se refere ao tipo de modelo gerador de *outliers*. Denby e Martin (1979) e Bustos e Yohai (1986), entre outros, consideram o modelo de contaminação por mistura, isto é, os *outliers* AO e IO são gerados por uma dada distribuição de probabilidade. Hotta e Neves (1992) e Palma (1998) descrevem métodos paramétricos de contaminação, que

consistem em alocar os *outliers* de tipo I ou II em posições pré-determinadas de uma série temporal. Logo abaixo, apresenta-se a definição de um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminado por um método paramétrico.

Definição 2.2. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). O processo $\{Z_t\}_{t \in \mathbb{Z}}$ segue o modelo paramétrico de contaminação se

$$Z_t = X_t + \omega f(\mathcal{B})I_t(T), \quad (3)$$

onde $\omega \in \mathbb{R}$ representa a magnitude do *outlier*, $f(\mathcal{B})$ é uma função de intervenção e $I_t(T)$ é uma variável indicadora do tempo T em que o *outlier* ocorre na série.

Pelo modelo gerador de *outliers* visto acima é possível obter *outliers* do tipo I, assim como do tipo II.

Definição 2.3. Se a função de intervenção dada pela equação (3) for $f(\mathcal{B}) = 1$, então pode-se afirmar que $\{Z_t\}_{t \in \mathbb{Z}}$ é um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com contaminação paramétrica do tipo I, pois a contaminação dependerá apenas da variável *dummy* $I_t(T)$.

Agora, para se ter um caso de contaminação inovadora, a função intervenção precisa ser um pouco mais complexa como exposto abaixo.

Definição 2.4. Se a função de intervenção, $f(\mathcal{B})$, dada pela equação (3) for o seguinte filtro

$$f(\mathcal{B}) = \frac{\phi(\mathcal{B})}{\theta(\mathcal{B})} \prod_{j=1}^k (1 - 2u_j \mathcal{B} + \mathcal{B}^2)^{\lambda_j}, \quad (4)$$

então tem-se que $\{Z_t\}_{t \in \mathbb{Z}}$ é um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com contaminação paramétrica do tipo II. Neste caso a variável *dummy* $I_t(T)$ indica quando a interferência começa e o filtro $f(\mathcal{B})$ indica qual o decaimento que essa interferência terá até se tornar desprezível.

A seguir, é definido o método de contaminação por mistura. Proposto, inicialmente, por Beran (1994) para *outlier* aditivos.

Definição 2.5. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico satisfazendo a equação

$$Z_t = X_t + I_t V_t, \quad (5)$$

onde $\{X_t\}_{t \in \mathbb{Z}}$ é um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) definido na equação (1). $\{I_t\}_{t \in \mathbb{Z}}$ é um processo, no qual as variáveis aleatórias seguem uma distribuição Bernoulli com probabilidade de sucesso igual a c . O processo $\{V_t\}_{t \in \mathbb{Z}}$ é constituído de variáveis aleatórias que seguem uma distribuição normal $\mathcal{N}(0, \tau_V^2)$. Portanto Z_t é um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com contaminação por mistura do tipo AO.

Diferentemente do caso AO, a contaminação por mistura de tipo II exige que a manipulação ocorra nos erros do processo estocástico.

Definição 2.6. Seja F_ξ uma distribuição normal contaminada, tal que

$$F_\xi = (1 - c)F_\varepsilon + cG_\varepsilon, \quad (6)$$

onde $0 \leq c \leq 1$ é a probabilidade de mistura, F_ε é uma distribuição normal não contaminada $\mathcal{N}(0, \sigma_\varepsilon^2)$ e G_ε uma distribuição arbitrária com uma dispersão superior a σ_ε^2 , então $\{\xi_t\}_{t \in \mathbb{Z}}$ são os erros de um processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com contaminação por mistura do tipo IO.

Assim estão definidos os dois tipos de *outliers* e as duas formas de contaminação. Para os processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminados por mistura, a função densidade espectral também é afetada. A seguir apresentam-se a função densidade espectral apropriada para estes processos quando sofrem contaminação AO, assim como IO.

Primeiramente, vê-se o caso AO para o qual Bisognin(2007) demonstra que se os processos $\{X_t\}_{t \in \mathbb{Z}}$ e $\{V_t\}_{t \in \mathbb{Z}}$ são independentes, vistos na equação (5), então a função densidade espectral de $\{Z_t\}_{t \in \mathbb{Z}}$, pode ser formulada como $f_Z(w) = f_X(w) + f_V(w)$, para todo $w \in [0, \pi]$.

Seja $f_X(w)$ a função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$, dada pela equação (2), e $f_V(w) = \frac{c\tau_V^2}{2\pi}$ a função densidade espectral do processo $\{V_t\}_{t \in \mathbb{Z}}$, então

$$f_Z(\omega) = \frac{\sigma_\varepsilon^2}{2\pi} \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \prod_{j=1}^k |2(\cos(w) - u_j)|^{-2\lambda_j} + \frac{c\tau_V^2}{2\pi}, \quad (7)$$

onde $f_Z(\omega)$ é a função densidade espectral do processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminado por *outliers* aditivos.

Prosseguindo com o processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminado por *outliers* inovadores, também neste caso a função densidade espectral requer algumas modificações. A prova pode ser conferida em Bisognin (2007).

Seja $f_X(\cdot)$ a função densidade espectral do processo $\{X_t\}_{t \in \mathbb{Z}}$ e $f_\xi(\omega) = \frac{(1-c)\sigma_\varepsilon^2 + c\tau_V^2}{2\pi}$ a função densidade espectral do processo $\{\xi_t\}_{t \in \mathbb{Z}}$. Então

$$f_Z(\omega) = \left[\frac{(1-c)\sigma_\varepsilon^2 + c\tau_V^2}{2\pi} \right] \frac{|\theta(e^{-i\omega})|^2}{|\phi(e^{-i\omega})|^2} \prod_{j=1}^k |2(\cos(w) - u_j)|^{-2\lambda_j}, \quad (8)$$

onde $f_Z(\omega)$ é a função densidade espectral do processo $\{Z_t\}_{t \in \mathbb{Z}}$ com ruído $\{\xi_t\}_{t \in \mathbb{Z}}$, ou seja, k -Factor GARMA com *outliers* inovadores.

A seguir apresenta-se a definição de função periodograma, que é um estimador da função densidade espectral.

Definição 2.7. Seja $\{X_t\}_{t=1}^n$ é uma série temporal qualquer. Então a função periodograma da série $\{X_t\}_{t=1}^n$ é dada por

$$I(w) = \frac{1}{2\pi n} \left| \sum_{t=1}^n X_t e^{-itw} \right|^2, \forall w \in [0, \pi] \quad (9)$$

3 Métodos de Estimação dos Parâmetros

Nesta seção apresentam-se os métodos propostos para a estimação dos parâmetros de processos k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Sabe-se *a priori* que conforme uma série é contaminada por *outliers*, que as estimativas dos parâmetros são inevitavelmente afetadas, como enfatizam Tsay (1986), Martin e Yohai (1986) e Peña (1990). No entanto, esse efeito pode ser em alguns casos minimizado. Neste trabalho são propostos três métodos de

estimação, um para quando a série está isenta de interferências e mais dois, um para cada tipo de contaminação já discutida.

3.1 Método *FT*

Apresenta-se aqui a extensão do estimador de máxima verossimilhança aproximado proposto por Fox e Taquq (1986) e Whittle (1951). Neste método, todos os parâmetros do modelo são estimados conjuntamente. Fox e Taquq (1986) apresentam as condições para as quais esse estimador seja consistente e tenha uma distribuição normal assintótica, que consiste fundamentalmente na estacionariedade da série estudada.

Definição 3.1. Seja $\{X_t\}_{t \in \mathbb{Z}}$ um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) sem nenhum tipo de contaminação, ver definição 2.1, com função densidade espectral dada pela expressão (2). Fazendo-se uso de $I_Z(w)$, o periodograma dado na equação (9), pode-se estimar $\boldsymbol{\eta}$. Seja

$$Q(\boldsymbol{\eta}) = \int_{-\pi}^{\pi} \frac{I(w)}{f_X(w; \boldsymbol{\eta})} dw, \quad (10)$$

onde $\boldsymbol{\eta}$ é o vetor de parâmetros que se deseja estimar dado por $\boldsymbol{\eta} \equiv \{\phi_1, \dots, \phi_p, u_1, \dots, u_k, \lambda_1, \dots, \lambda_k, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2\}$, então, o estimador de máxima verossimilhança aproximado, denotado por *FT*, é o valor de $\boldsymbol{\eta}$ que minimiza a função $Q(\cdot)$.

3.2 Método *FTAO*

Como visto anteriormente, séries contaminadas por *outliers* aditivos são melhor compreendidas por um processo estocástico que considere-os, sendo necessário para tanto utilizar a função densidade espectral própria desse processo. Dessa forma, o método *FTAO* consiste na mesma minimização da razão apresentada na expressão (10), porém substituindo-se a função densidade espectral e o vetor de parâmetros $\boldsymbol{\eta}$ pelos parâmetros adequados.

Portanto, se $\{Z_t\}_{t \in \mathbb{Z}}$ é um processo estocástico *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminado por *outliers* aditivos, recomenda-se o método *FTAO*, que minimiza $Q(\boldsymbol{\eta})$, tal qual na equação (10), com função densidade espectral dada por (7) e o vetor $\boldsymbol{\eta}$, tal que $\boldsymbol{\eta} \equiv \{\phi_1, \dots, \phi_p, u_1, \dots, u_k, \lambda_1, \dots, \lambda_k, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2, c\tau_V^2\}$.

3.3 Método *FTIO*

Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo estocástico *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminado por *outliers* inovadores. Para a estimação dos parâmetros deste processo utiliza-se a função $Q(\boldsymbol{\eta})$, dada pela expressão (10), onde $f_Z(w)$ é a função densidade espectral dada pela equação (8). Portanto o estimador de máxima verossimilhança aproximado, denotado por *FTIO*, de $\boldsymbol{\eta} \equiv \{\phi_1, \dots, \phi_p, u_1, \dots, u_k, \lambda_1, \dots, \lambda_k, \theta_1, \dots, \theta_q, (1-c)\sigma_\varepsilon^2 + c\tau_V^2\}$ minimiza a função $Q(\boldsymbol{\eta})$, onde a $f_Z(w)$ é dada pela equação (8).

4 Identificação e tratamento de *outliers* pelo algoritmo *SODA*

O algoritmo *SODA*, *Spectrum-based outlier detection algorithm*, foi idealizado como um método de identificação e tratamento de *outliers* em séries temporais. Choy (2001) avalia o desempenho desse algoritmo para modelos AR (p) e ARMA (p, q). Neste trabalho generalizou-se o algoritmo *SODA* aos processos *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Resumidamente, o algoritmo deve ser capaz de estimar um valor ω para cada observação

de uma dada série temporal, que indica o quão distante essa observação está do seu valor esperado, e um valor γ que indica o quão provável é que aquela observação seja de fato um ponto anômalo. γ é sempre positiva e quanto maior mais provável que a observação correspondente seja um *outlier*. Sendo que para obter ambas as estatísticas, o algoritmo requer basicamente a função densidade espectral e o periodograma do processo estocástico em questão.

Definição 4.1. Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), $I_Z(w)$ a função periodograma e $f_Z(w)$ a função densidade espectral, dadas pelas equações (9) e (2), respectivamente. Considere a estatística

$$D(I_Z, f_Z) = \int_{-\pi}^{\pi} (\log f_Z(w) + \frac{I_Z(w)}{f_Z(w)}) dw. \quad (11)$$

Seja T a marcação temporal, tal que $1 \leq T \leq n$. Então os estimadores de ω são dados por

$$\hat{\omega}(n, T) = \{\omega | \min\{D(I_{Z-\omega\delta}, f)\}\} \quad (12)$$

e para o caso AO e IO, respectivamente,

$$\hat{\omega}_{AO}(n, T) = \frac{\int_{-\pi}^{\pi} \sum_{t=1}^n Z_t e^{iw(t-T)} / f_Z(w) dw}{\int_{-\pi}^{\pi} 1 / f_Z(w) dw} \quad (13)$$

e

$$\hat{\omega}_{IO}(n, T) = \frac{\int_{-\pi}^{\pi} (\sum_{t=1}^n Z_t e^{iw(t-T)}) (\sum_{s=0}^{n-T} \psi_s e^{-iws}) / f_Z(w) dw}{\int_{-\pi}^{\pi} |\sum_{t=0}^{n-T} \psi_t e^{iws}|^2 / f_Z(w) dw}, \quad (14)$$

onde $\{\psi_k\}_{k \geq 0}$ são os coeficientes da representação média móvel infinita do processo $\{Z_t\}_{t \in \mathbb{Z}}$.

Deve ficar claro da definição acima, que a função densidade espectral utilizada, sempre é a definida na equação (2), ou seja, a função que não leva em conta contaminações de qualquer tipo. Pode-se, ainda, definir a variância assintótica para cada um dos casos.

Teorema 4.1. *Seja $\{Z_t\}_{t \in \mathbb{Z}}$ um processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) contaminado por outliers, então os estimadores baseados na aproximação de Whittle(1953) $\hat{\omega}_{AO}(n, T)$ e $\hat{\omega}_{IO}(n, T)$, definidos respectivamente pelas equações (13) e (14) tem distribuição:*

1. $\hat{\omega}_{AO}(n, T) \rightarrow N(\omega, \sigma_{AO}^2)$, quando $n \rightarrow \infty$, onde

$$\sigma_{AO}^2 = \frac{(2\pi)^2}{\int_{-\pi}^{\pi} 1/f(w) dw} \quad (15)$$

2. $\hat{\omega}_{IO}(n, T) \rightarrow N(\omega, \sigma_{IO}^2)$, quando $n \rightarrow \infty$ e $(n - T) \rightarrow \infty$, onde

$$\sigma_{IO}^2(m) = \frac{(2\pi)^2}{\int_{-\pi}^{\pi} |\sum_{t=0}^m \psi_t e^{iwt}|^2 / f(w) dw} \quad (16)$$

Resta agora definir as respectivas estatísticas Γ para cada um dos tipos de contaminação:

$$\Gamma_{AO, T} = \frac{\hat{\omega}_{AO}(n, T)}{\sigma_{AO}} = \frac{\int_{-\pi}^{\pi} \sum_{t=1}^n Z_t e^{iw(t-T)} / f(w) dw}{2\pi \sqrt{\int_{-\pi}^{\pi} 1/f(w) dw}} \quad (17)$$

e

$$\Gamma_{IO,T} = \frac{\hat{\omega}_{IO}(n,T)}{\sigma_{IO}(n-T)} = \frac{\int_{-\pi}^{\pi} (\sum_{t=1}^n Z_t e^{iwt}) (\sum_{s=0}^{n-T} \psi_s e^{-iw(s+T)}) / f(w) dw}{2\pi \sqrt{\int_{-\pi}^{\pi} |\sum_{t=0}^{n-T} \psi_t e^{iwt}|^2 / f(w) dw}} \quad (18)$$

Como ambas $\Gamma_{AO,T}$ e $\Gamma_{IO,T}$ tem distribuição assintótica $N(0,1)$, pode-se considerar essas estatísticas como critério para detecção de *outliers*. Para isso, é necessário estabelecer um ponto crítico C . Conforme sugerido por Chang et al. (1988) pode-se considerar $C = 4$ para baixa sensibilidade, $C = 3.5$ para média sensibilidade e $C = 3$ para alta sensibilidade em processos ARMA (p, q) . Esses valores, no entanto, precisam ser validados por simulação para os processos *k-Factor* GARMA $(p, \mathbf{u}, \boldsymbol{\lambda}, q)$, exatamente o que será executado na próxima seção.

5 Simulações de Monte Carlo

Nesta seção apresentam-se os resultados de simulação de Monte Carlo para as estimativas dos parâmetros do processo estocástico pelos diferentes estimadores propostos e o sucesso na identificação e tratamento de *outliers* pelo método SODA. Todas as simulações tomam como base o modelo *k-Factor* GARMA $(p, \mathbf{u}, \boldsymbol{\lambda}, q)$, sendo $p = 1 = q$, $\phi = 0.6$, $\mathbf{u} = \{-0.7, 0.5, 0.8\}$, $\boldsymbol{\lambda} = \{0.1, 0.2, 0.3\}$ e $\theta = -0.3$. O tamanho das séries foi sempre o mesmo, qual seja $n=700$, pois o tamanho da série real analisada corresponde a 739 meses. As contaminações das séries seguem o seguinte padrão, na subseção 5.1 quando se comparam os métodos FT, FTAO e FTIO, as contaminações são do tipo por mistura com as seguintes probabilidades de contaminação $c \in \{0, 0.01, 0.025, 0.05\}$. Já na subseção 5.2 quando se analisa o desempenho do algoritmo SODA são efetuadas contaminações paramétricas sempre na posição $T = 400$. A grandeza dos *outliers* foi mantida constante em $\omega = 5$. Finalmente, todos os casos foram considerados para o caso AO e IO.

5.1 Comparação entre os métodos FT, FTAO e FTIO

Nesta subseção observam-se os resultados de simulação para os estimadores paramétricos FT, FTAO e FTIO. Os gráficos destas simulações estão no Apêndice A e as tabelas no Apêndice B. As tabelas apresentam o valor médio das estimativas e o erro padrão, mas como esses dados provém de mil repetições cada, $re = 1000$, dessa forma os erros padrão ficaram muito pequenos. Por esse motivo, nos gráficos não constam os erros padrão de cada estimativa, mas linhas pontilhadas 10% acima e 10% abaixo do valor da amplitude aceitável para o parâmetro. Sendo que as amplitudes aceitáveis correspondem aos limites que os parâmetros podem assumir mantendo a estacionariedade do processo, ou seja, 10% corresponde a 0.1 em \mathbf{u} , ϕ_1 e θ_1 que podem variar entre -1 e 1, no entanto corresponde a 0.025 em $\boldsymbol{\lambda}$ que deve estar restrito entre 0 e 0.5. Tem-se assim um auxílio para avaliar o viés dos estimadores, dado que a escala dos gráficos não é constante.

Na Figura 6 apresentam-se as estimativas de cada um dos oito parâmetros do processo *k-Factor* GARMA $(p, \mathbf{u}, \boldsymbol{\lambda}, q)$, $p = 1 = q$, pelo método FT, assim como por FTAO, para quatro diferentes probabilidades de contaminação do processo, quais sejam $c \in 0, 0.01, 0.025, 0.05$. Fica claro que a partir desse nível de rigor, ambos os métodos não tem bons resultados para nenhuma das posições do vetor $\boldsymbol{\lambda}$, com relação a u_3 ambos apresentam boas estimativas médias, para u_2 novamente ambos tem bons resultados com exceção de FTAO para o caso de contaminação $c = 0$, ou seja, nos casos em que as séries temporais não estão contaminadas, o que é compreensível uma vez que o FTAO não foi desenhado para casos sem *outliers*. Finalmente, u_1 apresenta

problemas de estimação em ambos os métodos. A diferença entre FT e FTAO mostra-se mais nitidamente, quando se comparam os resultados de ϕ_1 e θ_1 , no qual ambos são melhor estimados por FTAO, em que este método resiste a contaminação enquanto FT mostra-se com crescente vício. No entanto, observando-se o método FT para o processo para o qual ele foi idealmente proposto, ou seja, quando $c = 0$, ele é um bom estimador de todos os parâmetros com exceção de u_1 .

Sobre a variabilidade das estimações, convém analisar pontualmente cada erro padrão das Tabelas 4 e 5 no Apêndice B. Na Tabela 4, referente ao estimador FT, tem-se que os maiores erros padrão se referem a u_1 , valores entre 0.0182 e 0.0212, valores aparentemente pequenos, no entanto correspondem a um desvio padrão da ordem de 0.5755 a 0.6704, valores altos dado que u_1 deveria estar no intervalo entre -1 e 1. O próximo parâmetro com alta variabilidade é o ϕ_1 que apresentou erro padrão entre 0.0115 e 0.0155. O parâmetro θ_1 também apresenta valores altos de erro padrão entre 0.0127 e 0.0148. As demais estimativas transcorreram de forma mais precisa, apresentando erros padrão bem menores. Convém comparar esses resultados com os do método FTAO, cujos resultados estão na Tabela 5. A estrutura de resultados do FTAO é semelhante, ou seja, a maior imprecisão está nos mesmos parâmetros, no entanto u_1 tem erros padrão entre 0.0171 a 0.0201, ϕ_1 entre 0.0092 e 0.0128 e θ_1 entre 0.0095 e 0.0118, ou seja, todos os parâmetros com variabilidade crítica tem melhores resultados em FTAO do que em FT.

A comparação entre FT e FTIO trás algumas complicações adicionais. A primeira delas é que uma contaminação $c = 0.01$ não equivale a 1% das observações da série contaminada, mas muito mais, pois cada contaminação IO instaura um novo regime em uma subsérie, ou seja, na contaminação IO, as observações seguintes da observação T também são afetadas. Assim $c \in \{0.025, 0.05\}$ mostraram-se impraticáveis na estimação tanto de ϕ_1 quanto de θ_1 . Portanto, é interessante comparar FT e FTIO somente quando $c = 0.01$, para a qual tem-se que FT tem melhor resultado em λ_1 , enquanto FTIO em u_1 , sendo que nos outros quatro parâmetros ambos se saem minimamente bem, como vê-se pela Figura 7. Novamente, lembra-se que o objetivo inicial de FT era estimar séries não contaminadas, o que ocorre com bastante sucesso, sendo a única exceção o u_1 .

No tocante a variabilidade das estimativas de FT e FTIO, percebe-se que os mesmos parâmetros apontados no caso AO novamente apresentam as maiores variabilidades em erros padrão. Pela Tabela 6, tem-se os erros padrão do FT para u_1 entre 0.0182 e 0.0185, ϕ_1 entre 0.0114 e 0.0119, θ_1 entre 0.0122 e 0.0128. Esses mesmos parâmetros tendem a ter menores erros padrão quando comparados com os resultados de FTIO presentes na Tabela 6, quais sejam u_1 entre 0.0092 e 0.0126, ϕ_1 entre 0.0069 e 0.0149, θ_1 entre 0.0096 e 0.0156. Portanto, fica claro que o u_1 do FTIO é mais preciso do que o FT, mas os resultados de ϕ_1 e θ_1 já são mais complexos e não é possível declarar que há um melhor método independentemente do nível de contaminação.

5.2 Desempenho do algoritmo SODA

O algoritmo SODA, tal como proposto por Choy (2001), é um processo iterativo de remoção de *outliers*. Isto é, dada uma série temporal qualquer estimam-se valores de Γ para cada uma de suas observações, deve-se então analisar se o maior valor de Γ é maior ou igual ao valor da constante C previamente definida, se for maior, subtraí-se da referida observação o valor de ω , caso contrário se assume que a série não possui contaminações. A nova série, da qual foi retirado o efeito do *maior outlier*, tem todas as suas posições reavaliadas pelo SODA, novamente, o maior valor de Γ será comparado a constante C . Esse procedimento segue até o maior Γ ser menor do que C e, portanto, não haver justificativa para seguir a iteração.

Começando pelo caso de contaminações aditivas, neste caso o algoritmo SODA classificou a posição 400 como a apresentando o maior valor Γ_{AO} em 99.7% dos casos, ou seja, o algoritmo funciona muito bem para identificar *outliers* aditivos. Na Figura 1 mostra-se o histograma do módulo de ω_{AO} e de Γ_{AO} , ambas as estatísticas tem comportamento que lembra uma distribuição normal mesmo que não passem em um teste de Shapiro-Wilk com um nível de significância $\alpha = 0.05$, visto o expressivo tamanho da amostra, 1000. Sendo a média do módulo de ω_{AO} igual a 5.2572 e o desvio padrão 0.7033, já Γ_{AO} tem média igual a 3.9185 e desvio padrão igual a 0.5127. Lembrando que o verdadeiro valor de ω_{AO} era 5, ou seja, o SODA fez uma estimação da grandeza dos *outliers* aditivos. Quanto aos valores de Γ_{AO} , pode-se justificar que estão de acordo com os valores propostos por Choy (2001), em que 4 é um bom valor de C para baixa sensibilidade e 3 para alta sensibilidade.

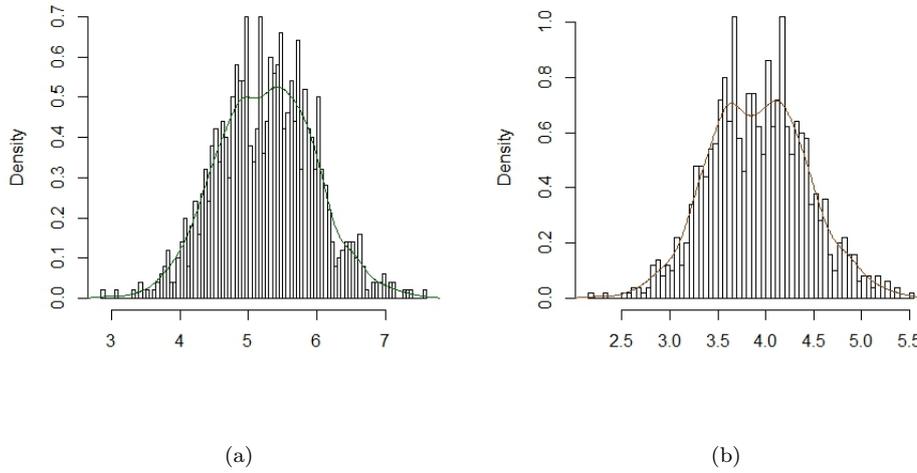


Figura 1: Estatísticas do SODA para 1000 séries contaminadas por AO: (a) ω (b) Γ

Na Tabela 1 observam-se os valores estimados para cada um dos parâmetros do modelo pelo método FT. Tanto as séries originais, que possuem contaminação, quanto as séries tratadas, que conforme dito anteriormente em 99.7% dos casos foram corretamente corrigidas, ambas foram estimadas pelo mesmo método FT. No geral, parece não haver grande diferença entre uma estimação e outra, com exceção dos parâmetros u_1 e θ_1 , em que a estimação das séries tratadas apresentaram um vício levemente maior. De qualquer forma, o objetivo principal deste procedimento é avaliar a qualidade da identificação e estimação de *outliers*, a reestimação da série fica em segundo plano, pois um único *outlier* não influencia tanto uma série que possui 700 observações. No entanto, influência no tocante aos resíduos, pois 10.1% das séries originais apresentavam comportamento normal pelo teste de Shapiro-Wilk com um nível de significância $\alpha = 0.05$, esse valor cresce para 93.1% se as séries forem tratadas no mesmo teste de Shapiro-Wilk. Não tão gritante é a diferença quanto à autocorrelação, em que 95.5% das séries originais passavam com o mesmo α no teste de Box-Ljung, e novamente há um incremento para 97.2% das séries tratadas não apresentam autocorrelação no mesmo teste.

No caso de contaminações por *outliers* inovadores, seguiu-se o mesmo modelo e a mesma posição de contaminação, qual seja, a posição 400. No entanto, o algoritmo foi bem sucedido na percepção da posição contaminada em apenas 74.0% dos casos, ou seja, em 260 séries a posição com o maior Γ_{IO} não era foi a posição 400, deriva deste fato que todo o procedimento de correção da série não ocorre de forma eficaz e, além disso, cria uma nova interferência, pois um filtro aplicado no local errado se comporta como um novo *outlier* inovador.

Tabela 1: Estimativa dos Parâmetros de 1000 séries contaminadas por AO, $\omega = 5$, antes e depois da aplicação do SODA

Séries Contaminadas por AO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.4860	0.4789	0.8155	0.0712	0.1605	0.2907	0.4940	-0.2570
EP	0.0175	0.0117	0.0012	0.0011	0.0021	0.0021	0.0115	0.0127
Séries Tratadas								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.4442	0.4845	0.8179	0.0750	0.1683	0.2888	0.4801	-0.2213
EP	0.0182	0.0121	0.0012	0.0011	0.0019	0.0021	0.0114	0.0126

As estimativa de ω_{IO} não foram tão precisas quanto as de ω_{AO} , mas ainda assim bastante razoáveis, a média do módulo de ω_{IO} foi 3.633 e o desvio padrão a 0.287, conforme se observa na Figura 2. Já no tocante a Γ_{IO} , tem-se uma média igual a 1.685 e desvio padrão de 0.287, ou seja, valores bem abaixo de Γ_{AO} para uma contaminação que foi da mesma ordem de grandeza.

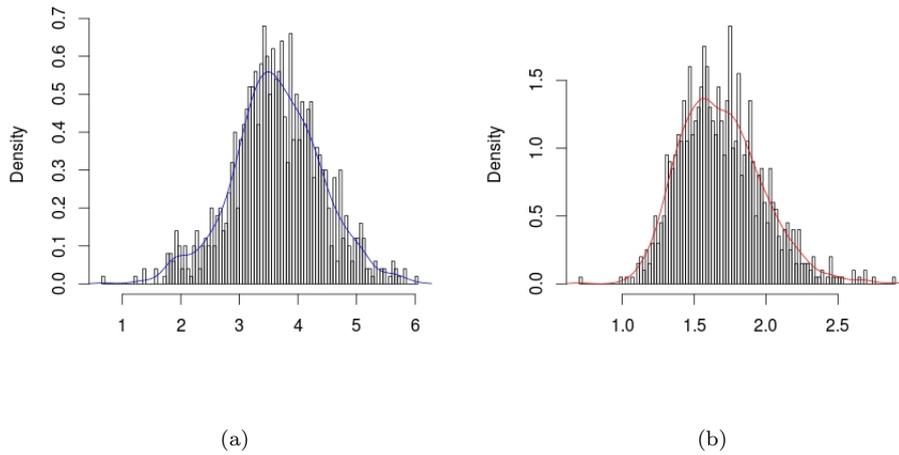


Figura 2: Estatísticas do SODA para 1000 séries contaminadas por IO: (a) ω (b) Γ

O tratamento de um *outlier* inovador é bastante delicado, pois exige que as alterações ocorram nos resíduos. Isto é, os resíduos da série contaminada são tratados e uma nova série é construída a partir de um filtro que usa como parâmetros as estimativas obtidas pelo método FT para a série com resíduos. Esta nova série deveria seguir os mesmos parâmetros e estar livre de *outliers*. No entanto, a menos que se saiba quais as posições infectadas e quais os verdadeiros parâmetros da série, o que acontece é uma sucessão de incertezas. Lembrando, que nestas simulações o algoritmo acertou apenas 74.0% das verdadeiras posições de contaminação, por isso não causa surpresa que na reestimação de parâmetros exibida na Tabela 2, as séries tratadas tenham tido um desempenho pior que as originais contaminadas. Destacam-se os parâmetros u_1 , u_2 , ϕ_1 e θ_1 com nítida piora nas séries tratadas em relação às séries originais.

Na análise de resíduos comparando as séries originais com as tratadas pelo algoritmo SODA, tem-se que exatamente 74.0% das séries temporais originais apresenta normalidade a um α de 0.05 e tem-se exatamente o mesmo percentual de séries normais após o tratamento. Fica evidente que o problema consiste no erro na suposição do lugar do *outlier* inovador. Ainda sobre os resíduos, há apenas uma minoria das séries cujos

resíduos não são autocorrelacionados, são 5.3% e 5.2% entre as séries originais e tratadas, respectivamente. Essa informação é suficiente para elucidar que estimar uma série contaminada por um único *outlier* inovador pelo método FT implica em problemas de autocorrelação, sendo o tratamento difícil.

Tabela 2: Estimativa dos Parâmetros de 1000 séries contaminadas por IO, $\omega = 5$, antes e depois da aplicação do SODA

Series Contaminadas por IO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.4530	0.5088	0.8155	0.0749	0.1736	0.2938	0.5000	-0.2530
EP	0.0179	0.0105	0.0012	0.0011	0.0020	0.0022	0.0115	0.0126
Series Tratadas								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.2786	0.5196	0.8179	0.0801	0.1747	0.2989	0.2103	-0.4513
EP	0.0174	0.0088	0.0016	0.0014	0.0021	0.0022	0.0151	0.0135

No geral, percebe-se que as estimativas de Γ_{IO} ficaram abaixo das estimativas de Γ_{AO} para os mesmos valores de ω , mesmos comprimentos de série e mesmo modelo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$). Embora Γ_{IO} e Γ_{AO} devessem idealmente apresentar a mesma distribuição, o fato do caso IO subestimar os valores de Γ_{IO} acarreta em um fator de segurança, pois uma falsa correção de *outliers* inovadores trás consequências muito mais graves do que uma falsa correção de *outliers* aditivos.

6 Aplicação

A série SOI, *Southern Oscillation Index*, consiste na diferença de pressão atmosférica entre Darwin e Tahiti, uma série meteorológica vital para o clima do planeta, pois está fortemente conectada ao ENSO, *El Niño Southern Oscillation*, conforme sustentam Webster e Yang (1992). Valores negativos da SOI coincidem com episódios de *La Niña* e valores positivos com *El Niño*. Esse fenômeno climático tem amplos efeitos influenciando o regime de monções na Índia, conforme demonstraram Shukla e Paolino (1983), a abundância de chuvas na Austrália conforme destacam Rimmington e Nicholls (1993), assim como o clima do oeste dos Estados Unidos da América, conforme Hansen et al (1997) e inclusive Antártida conforme o artigo de Turner (2004). Além disso, boa parte da fauna marinha do Pacífico sofre influência das oscilações térmicas do ENSO, podendo encontrar mais ou menos fontes de alimento e tornando-se mais ou menos populosas em decorrência disso, conforme descrevem Trathan et al (2007). Das muitas séries possíveis para avaliar o ENSO, a SOI é fortemente recomendada por Salisbury e Wimbush (2002), devido ao menor número de interferências próprias que não tem qualquer relação com *El Niño* ou *La Niña*.

A SOI é observada e divulgada pelo CPC-NOAA, Climate Prediction Center - National Oceanic and Atmospheric Administration² desde 1951 até Julho de 2015, ou seja, 775 meses. Aqui, utilizam-se os dados de Janeiro de 1951 até Julho de 2012 para todas as análises e modelagens, 739 observações, sendo que as 36 observações posteriores referentes ao intervalo entre Agosto de 2012 até Julho 2015 foram reservadas para comparar a eficiência das previsões. Conforme mencionado, esta série consiste na diferença de pressão atmosférica entre Darwin e Tahiti, sendo que ambas medições são previamente padronizadas pelas suas médias e desvios padrões

²Departamento do Governo Estadunidense de Monitoramento Climatológico, ver http://www.cpc.ncep.noaa.gov/information/who_were/index.shtml

conforme o período de 1951 até 1980³. Como ela está definida por uma padronização, seria de se esperar que ela apresentasse dados com distribuição normal, o que não chega a acontecer de forma substancial (p-valor igual a 0.0011 pelo teste Shapiro-Wilk). No entanto, essa não normalidade está condicionada a algumas observações, como vê-se no gráfico 3(b), que formam uma cauda inferior um pouco mais pesada do que o esperado.

Antes de uma conclusão mais robusta sobre a presença ou não de contaminação na série, analisa-se o sinal das estimativas de ω na Tabela 3, que em todos os casos tanto para AO quanto para IO, são negativos. Isto é, se há *outliers* na série eles devem ser os responsáveis por essa cauda inferior mais pesada.

O gráfico 3(a) ainda sugere que a média é constante, ou seja, que o processo é estacionário, o que os testes de Phillips-Perron (p-valor igual a 0.01) e Dickey-Fuller (p-valor igual a 0.01) confirmam. Seguindo num nível descritivo, observamos que o gráfico 3(c) mostra um nível de decaimento moderado, nem rápido, característico de modelos com curta dependência, nem lento, próprio de processos com longa dependência. Finalmente o gráfico 3(d) mostra vários picos em diferentes frequências do periodograma entre zero e 0.5, ou seja, um modelo ARFIMA(p, d, q) ou um GARMA(p, u, λ, q) dificilmente serão suficientes para uma modelagem apropriada, pois como enfatizam Collet et al. (2003) os picos do periodograma são uma informação chave na escolha do modelo $k - Factor$ Gegenbauer($\mathbf{u}, \boldsymbol{\lambda}$).

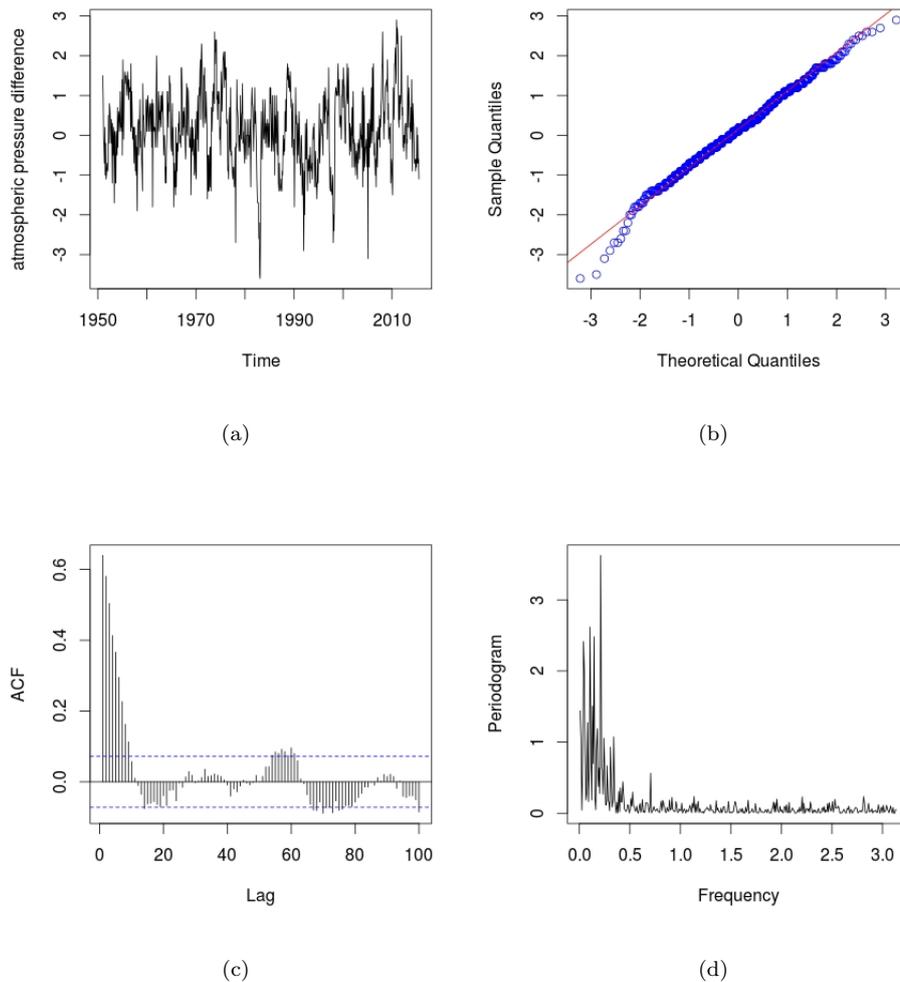


Figura 3: SOI Diagnóstico: (a) SOI (b) Gáfico QQ, (c) ACF e (d) Periodograma

³ver <https://www.ncdc.noaa.gov/teleconnections/enso/indicators/soi/>

A literatura estabelece que a ocorrência de *El Niño* ou *La Niña* flutua entre 2 a 7 anos⁴. Portanto, fica bastante difícil propor um modelo SARIMA com sazonalidade fixa. Essa flutuação tão esparsa se explica em parte pelo fenômeno ter décadas de maior atividade e outras de relativa inexistência. Esse tipo de padrão poderia ser classificado como o comportamento esperado de *outliers* inovadores, ou seja, *outliers* que causam alterações estruturais na série.

Tabela 3: Resultados dos 4 maiores Γ_{AO} e Γ_{IO} pelo SODA para a série SOI, sendo o modelo $k=3$, $p=0=q$

Posição	Data	ω_{AO}	Γ_{AO}	Posição	Data	ω_{IO}	Γ_{IO}
650	Fev 2005	-3.051	1.388	650	Fev 2005	-3.108	1.239
326	Fev 1978	-2.512	1.143	326	Fev 1978	-2.512	0.999
123	Mar 1961	-2.416	1.100	123	Mar 1961	-2.173	0.864
637	Jan 2004	-1.910	0.869	493	Jan 1992	-1.990	0.792

Apesar das considerações sobre a possível contaminação da série SOI, os valores tanto de Γ observados na Tabela 3 são demasiado baixos para efetuar um procedimento de reestimação de alguma observação da série pelo algoritmo SODA. Pode-se argumentar, que o maior Γ_{IO} , 1.239, não é tão baixo, pois o algoritmo mostrou uma tendência a subestimar Γ_{IO} , no entanto, lembra-se mais uma vez que corrigir contaminações do tipo II é um procedimento de risco que deve ser evitado, a menos que as evidências de contaminação sejam muito altas. Por esse motivo, não se procederá com nenhuma correção pelo SODA, mas nada impede a tentativa de estimar os parâmetros da série usando outro estimador que não o FT. Inclusive, os de Γ_{AO} foram ligeiramente superiores aos de Γ_{IO} , portanto não há porque não testar o FTAO, além do FTIO pela proposição teórica de haver um padrão já comentado na literatura da série SOI. É curioso, que somente tenham aparecido valores negativos de ω na Tabela 3, ou seja, devem ser todos momentos de *El Niño* e nenhum de *La Niña*, além de serem todos em meses próximos a Fevereiro, apesar de em anos separados por décadas.

Dadas as características da SOI acima discutidas, a estratégia será testar um apanhado de modelos, quais sejam, todas as combinações de *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) com k entre 1 e 4, assim como combinações de p e q iguais a zero ou um. Tem-se assim 16 modelos candidatos que foram estimados por três métodos, quais sejam FT, FTAO e FTIO. Definiu-se como critério mínimo para um modelo competitivo, apresentar resíduos não correlacionados⁵ e normalidade dos resíduos⁶. Entre os modelos competitivos, a escolha do modelo final fica a cargo de um equilíbrio entre medidas de erro, MAE, *Mean Absolute Error*, e ajuste, AIC, *Akaike Information Criterion*. Além desse modelo final, foi estimado um modelo pelo método automático presente no R, o popular *auto.arima*, qual seja o ARMA(p, q), $p = 2 = q$, onde $\phi = \{1.7640, -0.7985\}$ e $\theta = \{-1.3424, 0.4590\}$.

Pelos critérios eliminatórios, sobram 8 modelos estimados por FT, 13 modelos por FTIO e nenhum por FTAO. Pelo critério de AIC, a comparação entre modelos análogos por FT e FTIO, obtiveram-se sempre valores menores para FT, ou seja, a inclusão de um parâmetro para os *outliers* não se justificou perante um critério de parcimônia. Desses 13 modelos com melhores AIC, alguns estimados por FT e outros por FTIO, selecionou-se o com menor MAE, qual seja o *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), $p = 1 = q$, onde $\phi_1 = 0.8498$, $\mathbf{u} = \{-0.8257, -0.6437, 0.9774\}$, $\boldsymbol{\lambda} = \{0.0328, 0.0330, 0.1231\}$, $\theta_1 = -0.5871$, que apresentou MAE igual a 0.510 contra 0.572 do melhor modelo ARIMA.

Na Figura 4 fica claro que o processo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) é superior ao ARMA(p, q) para está previsão, pois o segundo rapidamente converge para a média enquanto o primeiro, por ter longa dependência, pode

⁴ver: <https://www.ncdc.noaa.gov/teleconnections/enso/>

⁵*p-value* maior ou igual a 0.05 no teste Box-Ljung para lag igual 24

⁶*p-value* maior ou igual a 0.05 no teste Shapiro-Wilk

estimar 36 passos a frente mantendo as flutuações da série. Pois, como argumentam Ferrara e Guégan(2001), se não há problemas de *misspecification* deve ser plenamente viável fazer previsões a longo prazo para processos que possuem longa dependência. Além disso, os intervalos de confiança se mantêm mais estáveis com o acréscimo de tempo do que modelos ARMA(p, q), no gráfico No gráfico 4 é apresentado apenas o intervalo de confiança do processo k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), que a 95% de confiança contempla 36 dos 36 verdadeiros pontos observados na SOI, já o intervalo com 95% de confiança do processo ARMA(p, q) cresce tão rápido que não agrega informação útil.

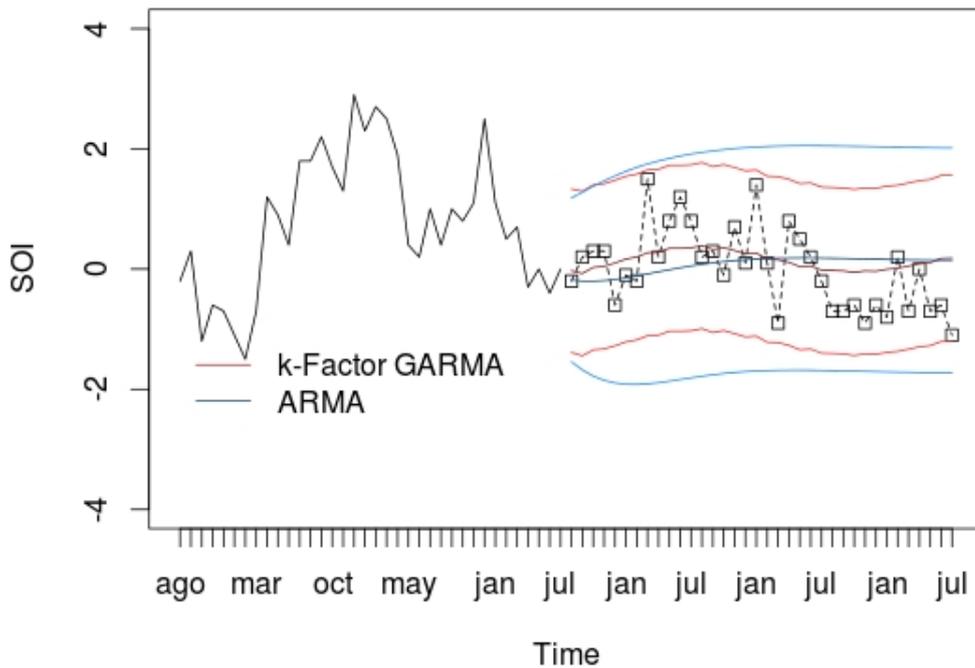


Figura 4: Previsão Mensal da SOI pelos Modelos k -Factor GARMA e ARMA, Agosto de 2013 a Julho de 2015

Pensando-se num sentido prático e simples, estimar corretamente o sinal da SOI já é um resultado útil ao pesquisador preocupado com o ENSO. Pois como discutido, valores negativos da SOI coincidem com as condições propícias ao acontecimento do *El Niño*, assim como os valores positivos com *La Niña*. Por esse critério, o k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) estimou o sinal da série corretamente 26 vezes, ou seja, apresenta 72.2% de acerto, enquanto o modelo ARMA(p, q) acerta 17 vezes, isto é, apenas 47.2% dos casos.

Finalmente, apresenta-se a Figura 5, na qual realiza-se uma previsão para Agosto de 2015 até Julho de 2017. Percebe-se que Agosto de 2015 registra o valor negativo de -0.565, o menor valor no intervalo de previsão. Lembrando que valores negativos são propícios ao *El Niño*. Durante todo o ano de 2016 os valores da previsão ficam muito próximos de zero, no entanto, em Maio de 2017 a previsão da SOI é de 0.273 e em Julho de 2017 alcança 0.276, ou seja, valores positivos e boas chances de ocorrer o evento *La Niña*, resfriamento do Oceano Pacífico.

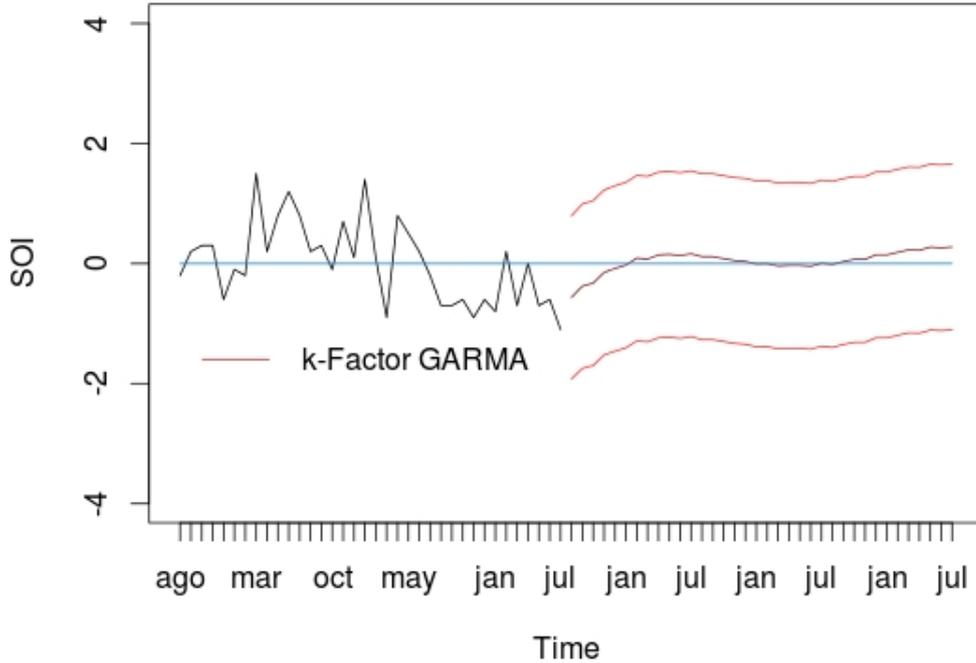


Figura 5: Previsão Mensal da SOI pelo Modelo k -Factor GARMA, Agosto de 2015 a Julho de 2017

7 Conclusão

Neste trabalho discutiu-se o processo estocástico k -Factor GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$), enfatizaram-se os casos de séries longas com ao menos 700 observações e com $k = 3$, visto que essas são características da série SOI, e foi uma preocupação manter o diálogo entre a parte aplicada e teórica deste artigo. Este foi padrão para as simulações, nas quais testaram-se contaminações paramétricas e por mistura com *outliers* aditivos e inovadores. Uma magnitude de ω igual a 5 foi considerada como um bom valor de teste, pois *outliers* menores pouco alteram a amplitude da série e, conseqüentemente, sua estimação, enquanto *outliers* maiores são mais raros e, obviamente, mais fáceis de identificar.

No tocante aos estimadores derivados do proposto por Fox e Taqqu, tem-se que os três estimadores FT, FTAO e FTIO alcançaram bons resultados quando aplicados a casos compatíveis. O estimador FT se saiu suficientemente bem em séries não contaminadas e, inclusive, mostrou-se suficientemente robusto para casos de contaminação AO e IO, esta última quando de ordem moderada, ou seja, quando $c = 0.01$. No entanto, diante de contaminações AO, ainda se recomenda o uso do estimador FTAO, visto que este tem menor variabilidade. Sobre as contaminações IO, percebe-se que se diante de contaminações iguais ou maiores a $c = 0.025$ não há como estimar corretamente, no entanto para casos menos extremos, tanto o estimador FT quanto o FTIO apresentam resultados satisfatórios. Sendo que se há contaminação do tipo IO, o FTIO ainda é um pouco mais preciso e apresenta um pouco menos de variabilidade.

A adaptação do algoritmo SODA foi bem sucedida para as contaminações do tipo AO, para as quais o SODA identificou a posição de contaminação de forma correta em quase todos os casos, além disso o algoritmo foi capaz

de corrigir o problema tratando as séries de forma adequada. Não pode-se dizer que esse procedimento melhorou a estimação dos parâmetros, visto que uma única contaminação por um ω não tão grande em uma série longa não afeta muito a série temporal. No entanto, a retirada de um único *outlier* dessas séries foi suficiente para que os resíduos que em maioria eram não normais se tornassem em absoluta maioria normais. Finalmente, pode-se confirmar os resultados obtidos por Choy (2001) para o caso ARMA e propor os mesmos valores de C para classificar a estatística Γ_{AO} . Já o caso de contaminação IO foi mais dificultoso para o SODA e não houve o mesmo sucesso. Embora o algoritmo tenha acertado a posição da maioria das contaminações, o percentual de erro foi bastante grande. Disso decorre, que as correções das séries tornaram-se em grande parte adição de novos erros e, conseqüentemente, a reestimação das séries após o tratamento foi pior do que a estimação das séries originais. Os valores de Γ_{IO} foram menores do que Γ_{AO} , mas não se chegou a sugerir valores de C , apenas alertou-se do perigo associado a aceitar falsos-positivos, visto que uma correção na posição equivocada pode destruir as características da série temporal em questão.

A aplicação em uma série temporal real foi bastante satisfatória, pois alcançou-se um modelo com baixo erro de previsão e parcimônia de parâmetros, além de resíduos normais e não correlacionados. O intervalo de confiança do modelo engloba todos os 36 verdadeiros pontos a frente sem inflar exponencialmente como um modelo de curta dependência faria. Portanto, reúnem-se aqui todos os motivos para a adoção de um modelo *k-Factor* GARMA($p, \mathbf{u}, \boldsymbol{\lambda}, q$) no lugar de um modelo mais simples na previsão da série SOI, mesmo que esse último tenha menores valores de AIC, somente essa vantagem não se justifica uma vez boas previsões da SOI repercutem em ganho de informação sobre o ENSO e o clima do Oceano Pacífico e entornos.

Referências

- [1] Beran J. 1994. *Statistics for Long-Memory Processes*. New York: Chapman & Hall.
- [2] Bisognin C. 2007. *Estimação e previsão em processos SARFIMA $(p, d, q)x(P, D, Q)s$ na presença de outliers*. Tese (Doutorado em Matemática) – Instituto de Matemática. Universidade Federal do Rio Grande do Sul, Porto Alegre.
- [3] Bustos OH, Yohai VJ 1986. Robust Estimates for ARMA Models *Journal of the American Statistical Association* **81**: 155-168.
- [4] Chang IH, Tiao GC, Cheng C 1988. Estimation of Time Series Parameters in the Presence of Outliers *Technometrics* **30**: 193-204.
- [5] Choy K 2001. Outlier detection for stationary time series *Journal of Statistical Planning and Inference* **99**(2): 111-127.
- [6] Collet J, Guégan D, Valdes-Sosa P 2003. How Shall We Determine the Number and the Location of the Gegenbauer Frequencies? An Empirical Approach, Note de Recherche IDHE-MORA n 2003-09.
- [7] Denby L, Martin RD 1979. Robust Estimation of the First-Order Autoregressive Parameter *Journal of the American Statistical Association* **74**: 140-146.
- [8] Ferrara L, Guégan D 2001. Forecasting with *k-factor* Gegenbauer Processes: Theory and Applications *Journal of Forecasting* **20**: 581-601.

- [9] Fox AJ 1972. Outliers in Time Series *Journal of the Royal Statistical Society* **B-43**: 350-363.
- [10] Fox R, Taqqu MS 1986. Large-sample Properties of Parameter Estimates for Strongly Dependent Stationary Gaussian Time Series *The Annals of Statistics* **14**: 517-532.
- [11] Giraitis L, Leipus R 1995. A Generalized Fractionally Differencing Approach in Long Memory Modelling *Lithuanian Mathematical Journal* **35**(1): 53-65.
- [12] Gray HL, Zhang NF, Woodward WA 1989. On Generalized Fractional Processes *Journal of Time Series Analysis* **10**(3): 233-257.
- [13] Guégan D 2005. How can we define the concept of long memory? An econometric survey *Econometric Reviews* **24** (2):113 - 149.
- [14] Hansen JW, Hodges AW, Jones JW 1997. ENSO Influences on Agriculture in the Southeastern United States *Journal of Climate* **11**: 404-411.
- [15] Hotta LK, Neves MMC 1992. A Brief Review on Test for Detection of Outliers of Time Series Models *Revista Colombiana de Estadística* **44**: 103-148.
- [16] Martin RD, Yohai VJ 1986. Influence Functionals for Time Series *The Annals of Statistics* **14**(3): 781-818.
- [17] Palma JA, 1998. *Outliers em Séries Temporais - Uma Abordagem no Domínio dos Modelos ARMA*. Dissertação de Mestrado, apresentada na Universidade de Lisboa - FCUL. Lisboa.
- [18] Palma W, 2007. *Long-Memory Time Series - Theory and Methods*. New Jersey: John Wiley.
- [19] Peña D 1990. Influential observation in time series *Journal of Business and Economic Statistics* **8**:235-242.
- [20] Rimmington GM, Nicholls N 1993. Forecasting wheat yields in Australia with the Southern Oscillation Index *Australian Journal of Agricultural Research* **44**(4): 625 - 632.
- [21] Salisbury JI, Wimbush M 2002. Using modern time series analysis techniques to predict ENSO events from the SOI time series *Nonlinear Processes in Geophysics, European Geosciences Union* **9** (3/4): 341-345. [jhal-00302126j](#)
- [22] Shukla J, Paolino DA 1983. The Southern Oscillation and Long-Range Forecasting of the Summer Monsoon Rainfall over India *Monthly Weather Review* **111**: 1830-1837.
- [23] Tsay RS, 1986. Time Series Model Specification in the Presence of Outliers *Journal of the American Statistical Association* **81**: 132-141.
- [24] Turner J 2004. The El Niño - Southern Oscillation and Antarctica *International Journal of Climatology* **24**: 1-31.
- [25] Trathan PN, Forcada J, Murphy EJ 2007. Environmental forcing and Southern Ocean marine predator populations: effects of climate change and variability *Phil. Trans. R. Soc. B* **362**: 2351-2365.
- [26] Webster PJ, Yang S 1992. Monsoon and Enso: Selectively Interactive Systems *Quarterly Journal of the Royal Meteorological Society* **118**: 877-926.

- [27] Whittle P 1951. *Hypothesis Testing in Time Series Analysis*. New York: Hafner.
- [28] Whittle P 1953. Estimation and information in stationary time series. *Ark Math* **2**: 423-434.
- [29] Woodward WA, Cheng QC, Gray HL 1998. A k -Factor GARMA Long-Memory Model *Journal of Time Series Analysis* **19**(4): 485-504.
- [30] Yohai VJ 1987. High breakdown point and high efficiency robust estimates for regression *Annals of Statistics* **15**: 642-656.

8 Apêndice A

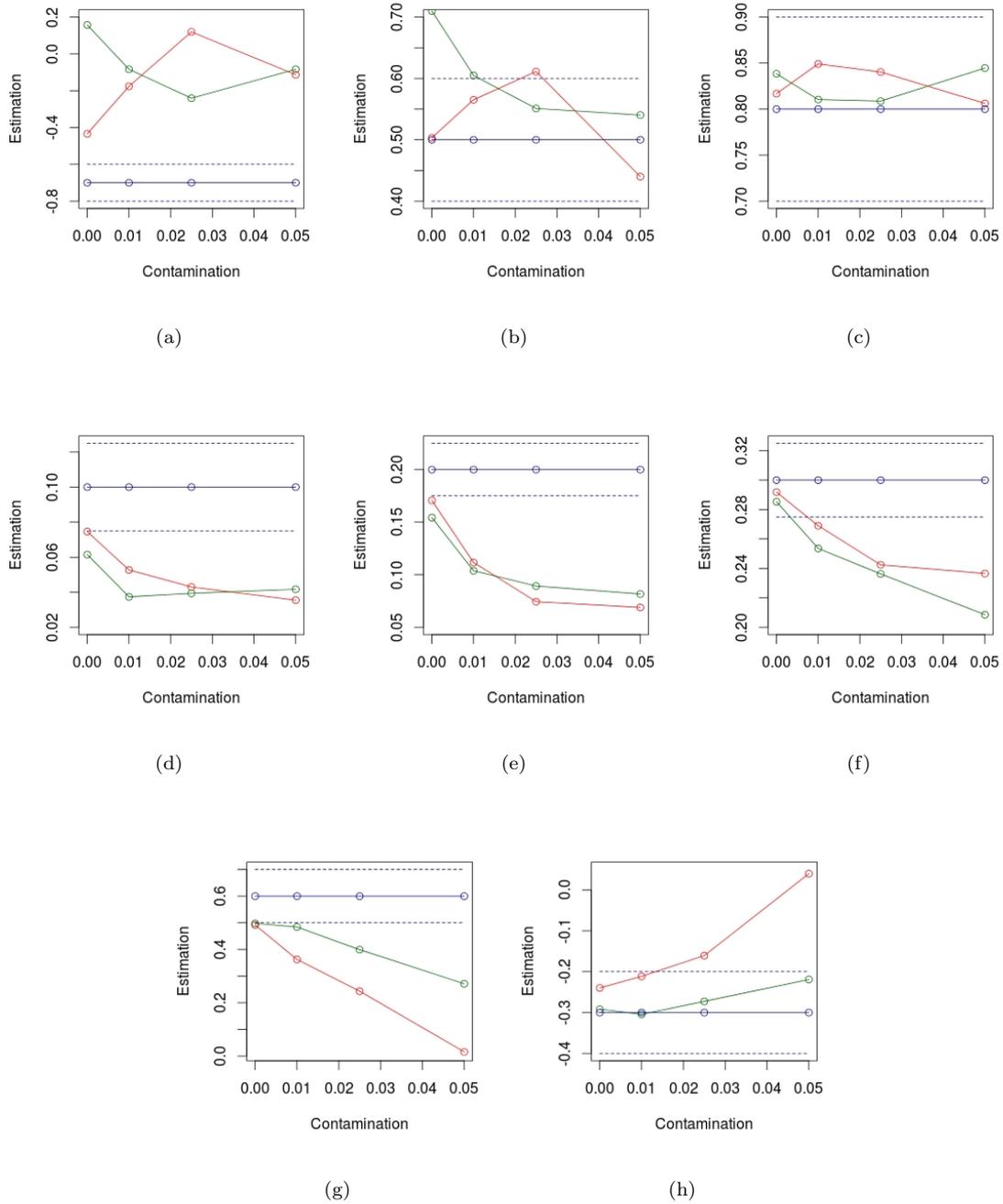


Figura 6: Estimativas por FT, vermelho, e FTAO, verde, dos parâmetros de 1000 repetições de processos k -Factor GARMA $(1, \mathbf{u}, \boldsymbol{\lambda}, 1)$ com $\mathbf{u} = -0.7, 0.5, 0.8$, $\boldsymbol{\lambda} = 0.1, 0.2, 0.3$, $\phi = 0.6$ e $\theta = -0.3$; $n=700$ e contaminação aditiva de magnitude $\omega=5$: (a) u_1 , (b) u_2 , (c) u_3 , (d) λ_1 , (e) λ_2 , (f) λ_3 , (g) ϕ , (h) θ , além disso a linha azul representa o valor do parâmetro e o pontilhado 10% de variação em relação ao parâmetro

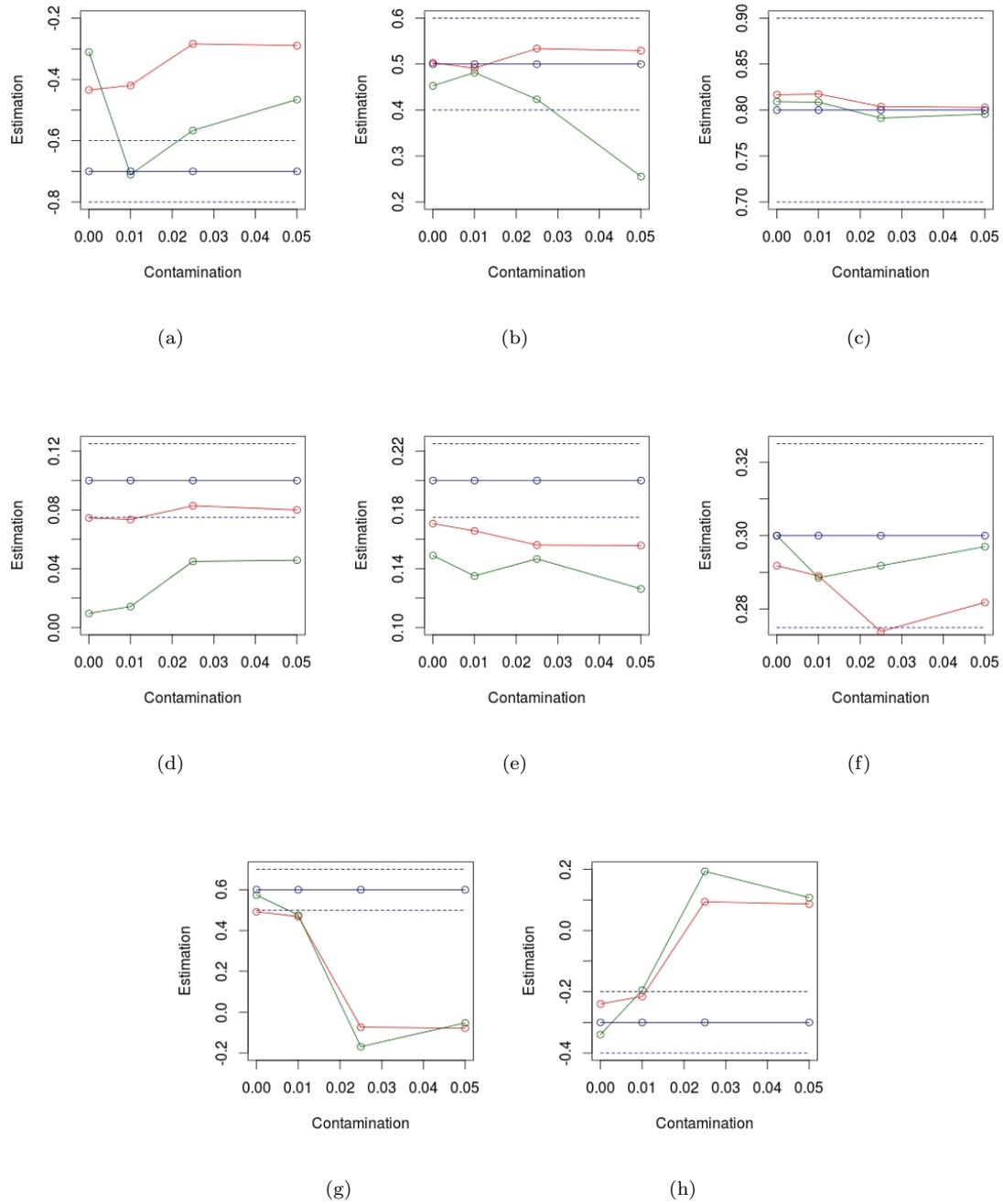


Figura 7: Estimativas por FT, vermelho, e FTIO, verde, dos parâmetros de 1000 repetições de processos k -Factor GARMA $(1, \mathbf{u}, \boldsymbol{\lambda}, 1)$ com $\mathbf{u} = -0.7, 0.5, 0.8$, $\boldsymbol{\lambda} = 0.1, 0.2, 0.3$, $\phi = 0.6$ e $\theta = -0.3$; $n=700$ e contaminação inovadora de magnitude $\omega=5$: (a) u_1 , (b) u_2 , (c) u_3 , (d) λ_1 , (e) λ_2 , (f) λ_3 , (g) ϕ , (h) θ , além disso a linha azul representa o valor do parâmetro e o pontilhado 10% de variação em relação ao parâmetro

9 Apêndice B

Tabela 4: Estimativa dos Parâmetros por FT de 1000 séries com contaminação gradativa por AO, $\omega = 5$ e $c \in \{0, 0.01, 0.025, 0.05\}$.

Séries com $c = 0$ de contaminação AO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.4342	0.5032	0.8168	0.0746	0.1707	0.2918	0.4918	-0.2398
EP	0.0182	0.0109	0.0012	0.0011	0.0020	0.0021	0.0115	0.0127
Séries com $c = 0.01$ de contaminação AO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.1763	0.5652	0.8491	0.0527	0.1116	0.2690	0.3625	-0.2114
EP	0.0212	0.0101	0.0022	0.0010	0.0020	0.0019	0.0146	0.0148
Séries com $c = 0.025$ de contaminação AO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	0.1198	0.6112	0.8403	0.0430	0.0744	0.2424	0.2434	-0.1608
EP	0.0193	0.0068	0.0022	0.0009	0.0012	0.0017	0.0150	0.0141
Séries com $c = 0.05$ de contaminação AO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.1136	0.4400	0.8062	0.0355	0.0688	0.2365	0.0159	0.0396
EP	0.0187	0.0090	0.0012	0.0011	0.0012	0.0017	0.0155	0.0133

Tabela 5: Estimativa dos Parâmetros por FTAO de 1000 séries com contaminação gradativa por AO, $\omega = 5$ e $c \in \{0, 0.01, 0.025, 0.05\}$.

Séries com $c = 0$ de contaminação AO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	127.4570	0.1573	0.7101	0.8384	0.0615	0.1544	0.2853	0.4974	-0.2918
EP	1.6928	0.0171	0.0054	0.0015	0.0016	0.0014	0.0020	0.0108	0.0109
Séries com $c = 0.01$ de contaminação AO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	143.7090	-0.0836	0.6048	0.8104	0.0374	0.1037	0.2536	0.4842	-0.3045
EP	1.1214	0.0200	0.0046	0.0010	0.0010	0.0012	0.0019	0.0092	0.0095
Séries com $c = 0.025$ de contaminação AO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	144.9905	-0.2407	0.5510	0.8087	0.0394	0.0892	0.2362	0.3991	-0.2728
EP	1.0230	0.0201	0.0056	0.0010	0.0009	0.0010	0.0019	0.0108	0.0099
Séries com $c = 0.05$ de contaminação AO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	134.6728	-0.0838	0.5403	0.8446	0.0417	0.0817	0.2085	0.2715	-0.2191
EP	1.1765	0.0200	0.0098	0.0023	0.0009	0.0010	0.0018	0.0128	0.0118

Tabela 6: Estimativa dos Parâmetros por FT de 1000 séries com contaminação gradativa por IO, $\omega = 5$ e $c \in \{0, 0.01, 0.025, 0.05\}$.

Séries com $c = 0$ de contaminação IO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.4342	0.5032	0.8168	0.0746	0.1707	0.2918	0.4918	-0.2398
EP	0.0182	0.0109	0.0012	0.0011	0.0020	0.0021	0.0115	0.0127
Séries com $c = 0.01$ de contaminação IO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.4197	0.4912	0.8175	0.0735	0.1657	0.2890	0.4685	-0.2157
EP	0.0185	0.0120	0.0130	0.0011	0.0020	0.0021	0.0119	0.0128
Séries com $c = 0.025$ de contaminação IO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.2834	0.5340	0.8037	0.0828	0.1561	0.2739	-0.0735	0.0940
EP	0.0184	0.0036	0.0008	0.0013	0.0017	0.0019	0.0114	0.0122
Séries com $c = 0.05$ de contaminação IO								
Parâmetro	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	-0.2892	0.5295	0.8028	0.0800	0.1558	0.2818	-0.0782	0.0866
EP	0.0182	0.0035	0.0007	0.0012	0.0018	0.0018	0.0115	0.0123

Tabela 7: Estimativa dos Parâmetros por FTIO de 1000 séries com contaminação gradativa por IO, $\omega = 5$ e $c \in \{0, 0.01, 0.025, 0.05\}$.

Séries com $c = 0$ de contaminação IO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	1.0216	-0.3104	0.4529	0.8092	0.0095	0.1490	0.3000	0.5736	-0.3400
EP	0.0018	0.0095	0.0065	0.0008	0.0006	0.0014	0.0015	0.0069	0.0096
Séries com $c = 0.01$ de contaminação IO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	1.3563	-0.7105	0.4816	0.8086	0.0142	0.135	0.2885	0.4758	-0.1947
EP	0.0045	0.0108	0.0048	0.0013	0.0006	0.0017	0.0018	0.0094	0.0115
Séries com $c = 0.025$ de contaminação IO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	1.8491	-0.5665	0.4235	0.7913	0.0450	0.1467	0.2918	-0.1695	0.1933
EP	0.0069	0.0092	0.0051	0.0014	0.0013	0.0018	0.0014	0.0149	0.0156
Séries com $c = 0.05$ de contaminação IO									
Parâmetro	$c\tau^2$	u_1	u_2	u_3	λ_1	λ_2	λ_3	ϕ	θ
Média	2.7233	-0.4657	0.2555	0.7956	0.0458	0.1263	0.2970	-0.0527	0.1076
EP	0.0092	0.0126	0.0125	0.0009	0.0015	0.0024	0.0016	0.0131	0.0146