

# O efeito de outliers na estimação de Modelos ARFIMA

Carlos Eduardo Starling  
Ela Mercedes M. de Toscano  
Lupercio França Bessegato  
Universidade Federal de Minas Gerais

## Resumo

No presente trabalho, o objetivo principal é estudar o efeito dos outliers na estimação dos modelos ARFIMA. Para a análise, foi considerada uma apropriada normalização na definição do tamanho do pulo nos outliers, e proporções diferentes no início, meio e final da série. Os resultados das simulações mostram que os métodos de estimação do parâmetro  $d$ , GPH e TAPER GPH, se mostraram ineficientes para essa situação, pois seu vício está aumentando à medida que se acrescentam outliers.

Palavra chave: Outliers, Processos Fracionários.

## 1. Introdução

Os valores de uma série temporal podem, muitas vezes, ser afetados por eventos inesperados tais como: *mudanças de políticas; crises econômicas; ondas inesperadas de frio ou calor; erros de digitação, etc.* A consequência da ocorrência desses tipos de eventos é a criação de observações espúrias que são inconsistentes com o resto da série; tais observações são denominadas de "valores atípicos" ou "outliers". De modo geral, as séries econômicas, não são raras as ocorrências de observações *outliers* cuja influência tem que ser levada em conta. O resultado da análise é a melhor compreensão do fenômeno por trás das mudanças e sua inclusão no modelo, promovendo uma capa cidade de explicação e de previsão melhor.

Este trabalho apresenta um estudo do feitos através de simulações de séries temporais, nos quais pretende-se mostrar a influência de observações atípicas (outliers) na modelagem de processos ARFIMA(0,d,0). O objetivo é estudar a diferença na modelagem das séries temporais com longa dependência de acordo com a quantidade de outliers existentes na série, e da localização desses outliers. As simulações foram feitas no programa S-Plus. Todas as séries simuladas continham 300 observações ( $n=300$ ) e os outliers são do tipo aditivos. Os modelos utilizados no estudo são os ARFIMA(0,d,0), com valores de  $d=0,2$  e  $d=0,45$ . Os estimadores utilizados foram o GPH (definido para processos estacionários) e o TAPER GPH (definido para processos com tendências ou quebras estruturais).

A presença de valores atípicos tem efeito: no comportamento da FAC e FACP amostral, na identificação correta de modelos, na estimação dos parâmetros do modelo ARFIMA.

Nesse estudo, considerou-se como observações atípicas valores ponderados pelo tamanho T da série. Dessa forma foram identificados como outliers os valores que ultrapassassem o intervalo  $[m-k\sigma T^{1/2}; m+k\sigma T^{1/2}]$ , onde k assumiu os valores 0,17, 0,23 e 0,29.

Como as séries simuladas eram de tamanho T=300, com desvio padrão igual a 1, os dados considerados atípicos tinham valores de 3, 4 e 5 de distância da média.

## 2. Modelos ARFIMA(p,d,q)

O modelo ARFIMA (p,d,q) é amplamente usado para descrever séries com características de longa dependência, que denotaremos por I(d). Consideramos a seguinte expressão,

$$(1 - B)^d X_t = a_t \quad -0,5 < d < 0,5$$

onde  $a_t$  é i.i.d.  $(0, \sigma_e^2)$  para  $t = 1, \dots, T$ , B o operador de defasagem. Para qualquer valor real de  $d > -1$ , o operador de diferenças fracionais,  $(1 - B)^d$ , é definido pela sua série de Maclaurin, que é,

$$(1 - B)^d = \sum_{j=0}^{\infty} p_j B^j$$

onde

$$p_j = \frac{\Gamma(j - d)}{\Gamma(j + 1)\Gamma(-d)} = \tilde{O} \frac{k - 1 - d}{k} \quad 0 < k \leq j$$

Um estimador do parâmetro d é método de Geweke e Porter-Hudak (1983) (GPH), indicado para processos estacionários, é baseado na seguinte regressão,

$$\ln I(w_j) = \ln f_U(0) - d \ln[2\text{sen}(w_j/2)]^2 + \ln[I_U(w_j)/f_U(0)]$$

onde  $I(w_j) = \frac{1}{2p} \left| \sum_{t=1}^p y_t \exp(iw_j t) \right|^2$  é o periodograma na frequência  $w_j = 2\pi j/T$ , o qual

depende do tamanho da amostra T. Frequentemente, é usado  $g(n) = T^{1/2}$ . O estimador para o parâmetro fracionário é dado por

$$\hat{d}_p = \frac{- \sum_{i=1}^{g(n)} (x_i - \bar{x}) y_i}{\sum_{i=1}^{g(n)} (x_i - \bar{x})^2} \quad , \text{ com } g(n) \text{ tal que } \frac{g(n)}{n} \rightarrow 0 \text{ , para } n \rightarrow \infty$$

Um outro estimador do parâmetro d, denominado TAPER GPH, é indicado para processos com quebra estrutural ou tendências. Este método é uma variação do GPH, e utiliza o periodograma tapered ao invés do periodograma padrão para estimar a função espectral. Essa modificação fornece mais robustez na estimação de d. O estimador TAPER GPH apresenta melhor comportamento nas séries não estacionárias ( $d > 1$ ), região onde o GPH padrão é fortemente viciado.

O cálculo do GPH taper é igual ao do GPH, porém baseado na série ponderada pelo seguinte fator:  $I_t = \frac{1}{2} \left[ 1 - \cos \left( \frac{2p(t+0,5)}{T} \right) \right]$ .

### 3. Estudo simulado

Com o objetivo de verificar a influência de observações atípicas em modelos ARFIMA (0,d,0), utilizou-se o software S-PLUS para simular 400 réplicas de processos ARFIMA(0,d,0) de tamanho 300, para diversos valores. Às séries simuladas foram adicionados saltos de valor  $k\sigma T^{1/2}$ , onde k assumiu os valores 0,17, 0,23 e 0,29. Estes saltos foram posicionados aleatoriamente no início da série (observações 1 a 100), no meio (101 a 200) e no final da série (201 a 300). A quantidade de outliers adicionada foi de 3%, 5% e 10% em relação ao tamanho da série T=300. Os métodos GPH e TAPER GPH foram utilizados para estimar os processos ARFIMA sem outliers e com outliers. A seguir serão apresentados os alguns resultados obtidos através das simulações.

Os resultados mostram que ao adicionar outliers no meio da série, as estimativas de d sofrem alterações. Ambos os estimadores se mostraram ineficientes para essa situação, pois seu vício está aumentando à medida que se acrescentam outliers. Os dois estimadores apresentaram comportamentos semelhantes, pois as médias das estimativas de ambos caíram com o aumento do número de outliers. A seguir alguns resultados da estimação.

**Tabela 1: Estatísticas descritivas das estimativas de d pelo método GPH dos processos ARFIMA(0,d,0), para d=0,2 e outliers no início da série.**

% de outliers	média	desvio padrão	assimetria	kurtosis	vício	mse
0%	0,198945	0,202199	-0,243427	0,560558	-0,001055	0,040783
3%	0,185363	0,211161	-0,282326	0,253518	-0,014637	0,044692
5%	0,173910	0,223338	-0,155451	0,132232	-0,026090	0,050436
10%	0,164643	0,227970	-0,287020	0,156111	-0,035357	0,053090

**Tabela 2: Estatísticas descritivas das estimativas de d pelo método TAPER GPH dos processos ARFIMA(0,d,0), para d=0,2 e outliers no início da série.**

% de outliers	media	desvio padrao	assimetria	kurtosis	vicio	mse
0%	0,220451	0,238747	-0,316761	0,196420	0,020451	0,057276
3%	0,211661	0,228272	-0,079815	-0,026042	0,011661	0,052114
5%	0,203104	0,225570	0,000007	0,140554	0,000104	0,055400
10%	0,194643	0,222872	0,000007	0,140554	0,000104	0,055400

## 4. Conclusões

Com os resultados obtidos conclui-se que a presença de dados atípicos pode influenciar fortemente as características da série, bem como a estimação do parâmetro de diferenciação fracionária.

Ao contrário do que se esperava, o método de estimação TAPER GPH não se mostrou eficaz nas séries que se aproximam da não estacionaridade.

Para as séries estacionárias, o estimador TAPER GPH se mostrou eficaz quando os outliers estavam presentes no início e no final da série.

Ao negligenciar os dados atípicos ambos métodos não têm bom desempenho. Portanto, faz-se necessário o uso de outros métodos de análise, como por exemplo análise de intervenção, ou a utilização de outros estimadores.

## Referências bibliográficas

- Geweke, J. & Porter-Hudak, S. (1983) The Estimation and Application of Long Memory Time Series Models, *Journal of Time Series Analysis*, 4, 15-39.
- Crato, N. (2001) Aplicações de Modelos de Memória Longa, 9a Escola de Séries Temporais e Econometria.
- Maddala, G. S. and Kim, I. Unit Roots, Cointegration, and Structural Change
- Wei W.W.S. (1980) - TIMES SERIES ANALYSIS - Univariate and multivariate methods
- Sibbertsen, Philipp (2002) – LONG MEMORY IN VOLATILITIES OF GERMAN STOCK RETURNS