

Boletim



**SOCIEDADE PORTUGUESA
DE ESTATÍSTICA**

Publicação semestral

primavera de 2019



Séries Temporais de Valor Inteiro

Introdução à teoria dos operadores thinning na modelação de séries temporais de valores inteiros

Manuel G. Scotto 12

Métodos de deteção de outliers baseados em wavelets: o caso dos modelos INAR(1) de Poisson

Isabel Silva e Maria Eduarda Silva 22

Modelos de contagem com estrutura periódica

Isabel Pereira, Magda Monteiro e Cláudia Santos 29

Uso de distribuições geométricas autorregressivas na análise de sequências de ADN

Sónia Gouveia 39

Cartas de controlo para o valor esperado de um processo INAR(1) com função ARL sem viés

Manuel Cabral Morais 46

CP-INGARCH: uma classe geral de modelos para séries de contagem

Filipa Alexandra Cardoso da Silva 53

| | |
|---|----|
| Editorial | 2 |
| Mensagem da Presidente | 3 |
| XXIV Congresso - Bolsas de Participação | 4 |
| Notícias | 5 |
| <i>Enigmística</i> | 11 |
| Ciência Estatística | 61 |
| Retrospectiva do Boletim SPE | 65 |
| Edições SPE | 66 |
| Prémios “Estatístico Júnior 2019” | 67 |
| Prémios SPE | 68 |

Informação Editorial

Endereço: Sociedade Portuguesa de Estatística,
Campo Grande. Bloco C6. Piso 4.
1749-016 Lisboa, Portugal.

Telefone: +351.217500120

e-mail: spe@spestatistica.pt

URL: <http://www.spestatistica.pt>

ISSN: 1646-5903

Depósito Legal: 249102/06

Tiragem: 350 exemplares

Execução Gráfica e Impressão: Gráfica Sobreireense

Editor: Fernando Rosado, fernando.rosado@fc.ul.pt

Sociedade Portuguesa de Estatística desde 1980

Métodos de detecção de *outliers* baseados em *wavelets*: o caso dos modelos INAR(1) de Poisson

Isabel Silva, *ims@fe.up.pt*

Faculdade de Engenharia, Universidade do Porto & CIDMA

Maria Eduarda Silva, *mesilva@fep.up.pt*

Faculdade de Economia, Universidade do Porto & CIDMA

1 Introdução

A presença de *outliers* ou observações discordantes nas séries temporais, assim como noutros tipos de dados, pode provocar efeitos adversos na identificação do modelo e na estimação dos parâmetros. Ignorar estas observações díspares pode mascarar a presença de importantes fenómenos subjacentes, impedindo a análise de risco. Várias metodologias para detetar e estimar *outliers* e outros efeitos de intervenção têm sido estabelecidas para modelos ARMA e alguns modelos de séries temporais não lineares, ver por exemplo Chang *et al.* [6], Chen e Liu [7] e Tsay [18]. As abordagens propostas baseiam-se principalmente em métodos iterativos, estimadores robustos e estatísticas baseadas na razão de verosimilhanças.

Segundo Fox [8], no contexto das séries temporais, podem ser considerados dois tipos de *outliers*: os aditivos e os inovacionais. Os *outliers* aditivos são erros externos ou mudanças exógenas que ocorrem num determinado instante e por isso só afetam a observação correspondente ao tempo onde a perturbação acontece. Os *outliers* inovacionais estão associados a mudanças internas ou efeitos endógenos no processo do ruído, afetando as observações subsequentes.

Nos últimos 20 anos, métodos baseados em *wavelets* foram propostos para resolver o problema de detecção da localização de *outliers* em modelos lineares e não lineares (Bilen e Huzurbazar [4], Grané e Veiga [9]). *Wavelets* são funções que combinam propriedades como localização no tempo e em escala, ortonormalidade, diferentes graus de suavidade, suporte compacto e implementação rápida (ver Percival e Walden [12]). A Transformada *Wavelet* Discreta (TWD) é uma poderosa ferramenta para a análise multi-resolução em tempo-escala e consiste na aplicação de filtros de diferentes frequências de corte usados para analisar um sinal em diferentes escalas. Os coeficientes da TWD são capazes de revelar mudanças na variância e de nível, assim como outros tipos de discontinuidades nos dados.

Apesar destes desenvolvimentos, a detecção e estimação de *outliers* no contexto das séries temporais de contagem tem recebido menos atenção. A análise de séries temporais de contagem tem-se tornado uma área de investigação ativa nos últimos anos. Estas séries caracterizam-se por apresentar valores baixos ou nulos, assimetrias e sobre-dispersão e podem ser encontradas nas mais variadas áreas de aplicação. Uma das abordagens mais populares propostas para analisar estes tipos de dados consiste na utilização de modelos baseados numa operação aleatória chamada *thinning*, que conjugada com inovações discretas, permite preservar a natureza discreta das contagens. Assim surge a família de modelos auto-regressivos e de médias móveis de valor inteiro, INARMA, que tem sido amplamente estudada na literatura (ver Scotto *et al.* [13]). Barczy *et al.* [2, 3] propuseram um método dos Mínimos Quadrados Condicionais para a estimação dos parâmetros do modelo INAR(1) contaminado com *outliers* aditivos ou inovacionais, assumindo que os tempos de ocorrência do *outlier* são conhecidos mas os seus tamanhos desconhecidos. Adicionalmente, Silva e Pereira [16] sugeriram uma abordagem Bayesiana para detetar *outliers* aditivos em modelos PoINAR(1), isto é, INAR(1) de Poisson. Recentemente, Bourguignon e Vasconcellos [5]

usaram métodos baseados em características e sinais para obter estimadores robustos dos parâmetros do modelo INAR(1) na presença de *outliers* aditivos.

Neste trabalho são apresentados dois métodos baseados em *wavelets*, propostos por Silva e Silva [14], que permitem a identificação dos tempos de ocorrência dos *outliers*, aditivos ou inovacionais, isolados, múltiplos ou em grupo (*patches*), em modelos PoINAR(1). No primeiro método, adaptado de Grané e Veiga [9], os chamados coeficientes de detalhe obtidos pela aplicação da TWD, usando a *wavelet* de Haar, são comparados com um limiar. No segundo, o método de reamostragem paramétrica de Tsay [19] é usado para obter a distribuição empírica dos tais coeficientes de detalhe.

O resto do trabalho está organizado da seguinte forma: a Transformada *Wavelet* Discreta é brevemente descrita na Secção 2. A Secção 3 apresenta os modelos PoINAR(1) contaminados com *outliers* aditivos e inovacionais. Os procedimentos para detetar o tempo de ocorrência de *outliers* nestes modelos são descritos na Secção 4 e ilustrados num conjunto de dados reais na Secção 5. A Secção 6 contém algumas observações finais.

2 Transformada *wavelet* discreta

Uma *wavelet* é uma função que pode ser vista como uma pequena onda que cresce e decresce num período de tempo limitado (mais detalhes em Percival e Walden [12]). A análise *wavelet* usa versões com diferentes escalas e deslocamentos da chamada *wavelet* mãe, de modo a fornecer a localização no tempo de cada componente espectral de modo semelhante à análise de Fourier, onde funções sinusoidais são usadas para encontrar as componentes em frequência que compõem um sinal.

Seguindo o trabalho de Percival e Walden [12], seja $\mathbf{X} = \{X_t, t = 0, \dots, N-1\}$ uma série temporal, com $N = 2^J$, $J \in \mathbb{N}$. Os coeficientes da TWD $\mathbf{W} = \{W_n, n = 0, \dots, N-1\}$ definem-se por:

$$\mathbf{W} = \mathcal{W} \mathbf{X} \Leftrightarrow [\mathbf{W}_1 \dots \mathbf{W}_J \mathbf{V}_J]^T = [\mathcal{W}_1 \dots \mathcal{W}_J \mathcal{V}_J]^T \mathbf{X},$$

onde \mathcal{W} é uma matriz ortonormal, de dimensão $N \times N$, de versões dilatadas e deslocadas da *wavelet* mãe $\psi(\cdot)$, definidas através de $\frac{1}{\sqrt{d}}\psi\left(\frac{u-t}{d}\right)$ sendo d o parâmetro de dilatação e t o parâmetro de deslocamento, em que $d = 2^j$ e $t = k2^j$, para $j, k \in \mathbb{Z}$.

É possível reconstruir a série temporal à custa da TWD inversa, através de $\mathbf{X} = \mathcal{W}^T \mathbf{W} = \sum_{j=1}^J \mathcal{W}_j^T \mathbf{W}_j + \mathcal{V}_J^T \mathbf{V}_J$.

Na prática, a matriz dos coeficientes da TWD, \mathbf{W} , calcula-se através do chamado algoritmo piramidal proposto por Mallat [10], baseado em filtragens. Especificamente, para uma largura par L , considerem-se o filtro *wavelet* $\{h_l : l = 0, \dots, L-1\}$ (filtro passa-alta) e o filtro de escala $g_l = (-1)^{l+1} h_{L-1-l}$ (filtro passa-baixas). No primeiro passo do algoritmo, dois conjuntos de coeficientes são produzidos pela convolução de \mathbf{X} com $\{g_l\}$ (coeficientes de aproximação do primeiro nível $c\mathbf{A}_1$) e com $\{h_l\}$ (coeficientes de detalhe do primeiro nível $c\mathbf{D}_1$); posteriormente é realizada uma decimação (*downsample*), retendo unicamente as observações filtradas intercaladas. No próximo passo, repete-se o procedimento anterior substituindo \mathbf{X} por $c\mathbf{A}_1$ de modo a obter $c\mathbf{A}_2$ e $c\mathbf{D}_2$. Desta forma, no nível j , a decomposição de \mathbf{X} tem a seguinte estrutura $[c\mathbf{A}_j, c\mathbf{D}_j, c\mathbf{D}_{j-1}, \dots, c\mathbf{D}_1]$.

Os coeficientes de detalhe capturam certas características da série temporal, como mudanças repentinas ou picos, apresentando valores elevados na presença dessas singularidades, podendo então ser usados para detetar *outliers*.

A *wavelet* mãe escolhida neste trabalho foi a *wavelet* de Haar, que pode ser considerada como uma onda quadrada, sendo assim apropriada para séries de contagem. Está definida por:

$$\psi(t) = \begin{cases} -1/\sqrt{2}, & -1 \leq t \leq 0 \\ 1/\sqrt{2}, & 0 < t \leq 1 \\ 0, & \text{caso contrário,} \end{cases}$$

e neste contexto, os filtros passa-baixa correspondem a médias móveis das observações enquanto que os filtros passa-alta correspondem a diferenças móveis das observações.

3 Modelos INAR(1) de Poisson contaminados com outliers

Séries temporais de contagem aparecem nos mais variados domínios de investigação, nomeadamente nas ciências sociais, biologia, economia e finanças, telecomunicações, seguros, entre outros. Uma das abordagens propostas na literatura para analisar este tipo de dados é o modelo auto-regressivo de valor inteiro de primeira ordem, INAR(1), proposto independentemente por Al-Osh e Alzaid [1] e McKenzie [11]. Este modelo usa a operação *thinning* binomial proposta por Steutel e Van Harn [17], de-

finida por $\alpha \circ X = \sum_{k=1}^X Y_k$, onde X é uma variável aleatória de valor inteiro não negativo, $\alpha \in [0, 1]$ e $\{Y_k\}$, $k = 1, \dots, X$, é uma sequência de variáveis aleatórias de Bernoulli independentes e identicamente distribuídas, independente de X , chamada série de contagem. Note-se que $\alpha \circ X | X \sim \text{Bi}(X, \alpha)$ (mais detalhes e propriedades desta operação podem ser encontrados em Silva e Oliveira [15]).

O processo estocástico em tempo discreto de valor inteiro não negativo, $\{X_t\}$, é um processo INAR(1) de Poisson, denotado abreviadamente por PoINAR(1), se satisfaz a seguinte equação:

$$X_t = \alpha \circ X_{t-1} + e_t, \quad (1)$$

onde $e_t \sim \text{Poisson}(\lambda)$, é o processo de chegada, $0 < \alpha < 1$, e para cada t , todas as séries de contagem $\alpha \circ X_{t-1}$ são mutuamente independentes e independentes de $\{e_t\}$. Nestas condições, o processo é estritamente estacionário e $X_t \sim \text{Poisson}(\frac{\lambda}{1-\alpha})$ quando $X_0 \sim \text{Poisson}(\frac{\lambda}{1-\alpha})$.

Diz-se que o processo PoINAR(1) está contaminado com $I \in \mathbb{N}$ outliers aditivos (AO), que ocorrem nos tempos $s_i \in \mathbb{N}$, e têm magnitude $\omega_i \in \mathbb{N}$ para $i = 1, \dots, I$, se

$$Y_t = X_t + \sum_{i=1}^I \delta_{i,s_i} \omega_i,$$

onde X_t é um modelo PoINAR(1), que satisfaz (1), e $\delta_{k,m}$ é uma função indicadora ($\delta_{k,m} = 1$, se $k = m$; $\delta_{k,m} = 0$, se $k \neq m$). Por outro lado, o processo PoINAR(1) diz-se contaminado com $I \in \mathbb{N}$ outliers inovacionais (IO), com magnitude ω_i e ocorrendo nos tempos s_i , $i = 1, \dots, I$, se

$$Y_t = \alpha \circ Y_{t-1} + \eta_t,$$

com $\eta_t = e_t + \sum_{i=1}^I \delta_{i,s_i} \omega_i$, onde $e_t \sim \text{Poisson}(\lambda)$ e $\delta_{k,m}$ está definida como anteriormente.

Note-se que em ambos os casos, o processo sem outliers, X_t , não é observado.

4 Procedimentos para detetar o tempo de ocorrência de outliers

Dois procedimentos, baseados em *wavelets*, para a deteção do tempo de ocorrência de outliers em processos PoINAR(1) podem ser descritos através dos seguintes passos:

Passo 1 Dada uma série temporal de contagem, $\mathbf{Y} = \{Y_t, t = 0, \dots, N\}$, ajusta-se um modelo PoINAR(1)

e estimam-se os resíduos de Pearson¹ $\mathbf{Z} = \{\hat{z}_t, t = 1, \dots, N-1\}$, onde $\hat{z}_t = \frac{Y_t - (\hat{\alpha}Y_{t-1} + \hat{\lambda})}{\sqrt{\hat{\alpha}(1-\hat{\alpha})Y_{t-1} + \hat{\lambda}}}$.

¹ $Z_t = \frac{Y_t - E\{Y_t|Y_{t-1}\}}{\sqrt{\text{Var}\{Y_t|Y_{t-1}\}}}$

Passo 2 Aplica-se a TWD aos resíduos de Pearson de modo a obter os coeficientes de detalhe do primeiro nível, $cD_1 = (d_1, d_2, \dots, d_{N/2})$.

Passo 3a Abordagem pelo limiar:

- (i) Define-se o limiar k_1^a (discutido na Subsecção 4.1).
- (ii) Determina-se o conjunto de índices (ordenados) $S = \{s_1, \dots, s_I\}$, das posições onde os coeficientes de detalhe estão acima do limiar k_1^a . Como sugerido por Grané e Veiga [9], os *outliers* são detetados de forma recursiva. Assim, Z é reconstruído a partir da TWD inversa aplicada aos coeficientes de detalhe modificados de forma a que o maior coeficiente de detalhe (em valor absoluto) seja substituído por zero. O procedimento termina quando não são detetados mais *outliers*.

Passo 3b Abordagem pela reamostragem paramétrica:

- (i) Calcula-se o envelope de aceitação (discutido na Subsecção 4.2).
- (ii) Determina-se o conjunto de índices (ordenados) $S = \{s_1, \dots, s_I\}$, das posições onde os coeficientes de detalhe estão fora do envelope de aceitação.

Passo 4 Para obter a posição exata do *outlier* na série dos resíduos, seja s um elemento genérico de S , calcula-se a média amostral de Z sem as observações $2s$ e $2s - 1$, isto é, $\bar{z}_{N-2} = \frac{1}{N-2} \sum_{i \neq 2s, 2s-1} \hat{z}_i$. Então, o tempo de ocorrência do *outlier* na série dos resíduos é $2s$ se $|\hat{z}_{2s} - \bar{z}_{N-2}| > |\hat{z}_{2s-1} - \bar{z}_{N-2}|$, ou $2s - 1$ caso contrário.

Note-se que no caso de *outliers* em grupo (*patches*), os coeficientes do primeiro nível só permitem detetar o tempo em que começa, mas não a sua duração, sendo necessário recorrer aos coeficientes de detalhe do segundo nível cD_2 (Bilen e Huzurbazar [4] e Grané e Veiga [9]). Assim, no **Passo 3a** há um limiar para cada nível, k_1^{a1} e k_2^{a2} , respetivamente. Analogamente, tem que ser determinado um envelope de aceitação para cada nível no **Passo 3b**.

4.1 Definição do limiar

É sabido que os coeficientes *wavelet* de dados Gaussianos ou de um ruído branco são eles próprios Gaussianos ou ruído branco. Adicionalmente, como referido por Bilen e Huzurbazar [4] e Percival e Walden [12], os coeficientes *wavelet* em W_j são aproximadamente não correlacionados mesmo quando os dados são altamente correlacionados. Contudo, no contexto não Gaussiano das séries de contagem, não há resultados disponíveis para a distribuição dos coeficientes de detalhe da análise *wavelet*. Por isso, são usadas simulações de Monte Carlo para obter a distribuição empírica do máximo dos coeficientes de detalhe (em valor absoluto) para os resíduos de Pearson dos modelos PoINAR(1). O limiar é então definido da seguinte forma. Para cada par de parâmetros (α, λ) no conjunto $\{(\alpha, \lambda) : \alpha = (2k + 1) \times 10^{-1}, k = 0, \dots, 4; \lambda = 2k + 1, k = 0, \dots, 14\}$, 20000 repetições do processo PoINAR(1) correspondente são geradas para cada tamanho de amostra $N = 2^J + 1$, para $J = 7, \dots, 10$. O modelo é ajustado, os resíduos de Pearson, \hat{z}_i , para $i = 1, \dots, N - 1$, são estimados e os máximos dos coeficientes de detalhe do primeiro e segundo nível são obtidos. Os limiares k_1^{a1} e k_2^{a2} são definidos como os $100(1 - a)^\circ$ percentis das distribuições empíricas correspondentes, para $a = a_1$ ou $a = a_2$. Os resultados obtidos indicam que os limiares variam não apenas com o tamanho da amostra N mas também com a combinação específica dos parâmetros α e λ . Portanto, adotando uma estratégia conservadora, para cada tamanho de amostra N os limiares são definidos como o mínimo obtido para todos as combinações de parâmetros em cada nível de decomposição. Os limiares obtidos são apresentados na Tabela 1.

Tabela 1: Limiares correspondentes aos 90° e 95° percentis da distribuição empírica do máximo dos coeficientes de detalhe (de primeiro e de segundo nível), em valor absoluto, para resíduos de Pearson de modelos PoINAR(1).

| N | 128 | 256 | 512 | 1024 |
|--------------|-------|-------|-------|-------|
| $k_1^{0.05}$ | 3.469 | 3.694 | 3.886 | 4.118 |
| $k_1^{0.1}$ | 3.182 | 3.450 | 3.657 | 3.840 |
| $k_2^{0.05}$ | 3.157 | 3.347 | 3.518 | 3.691 |
| $k_2^{0.1}$ | 2.936 | 3.138 | 3.320 | 3.504 |

4.2 Determinação do envelope de aceitação

Tsay [19] propôs um método para obter a distribuição empírica de um funcional escolhido usando amostras *bootstrap* geradas a partir de um modelo ajustado, e então comparar o valor observado para a série com a distribuição empírica obtida. Para este propósito, um envelope de aceitação é calculado a partir dos $100(1 - a/2)^\circ$ e $100a/2^\circ$ percentis desta distribuição empírica. Se o modelo ajustado for adequado, o funcional de interesse dos dados originais deve estar dentro do envelope. Neste trabalho, os funcionais de interesse são os coeficientes de detalhe de primeiro e segundo nível dos resíduos de Pearson do modelo PoINAR(1). Assim, para vários tamanhos de amostra $N = 2^J + 1, J = 7, 8, 9$, e valores dos parâmetros $\{(\alpha, \lambda) : \alpha \in \{0.1, 0.5, 0.9\}; \lambda \in \{1, 5, 9, 13\}\}$, 20000 realizações do processo PoINAR(1) são geradas e os correspondentes resíduos de Pearson são estimados. Para cada série de resíduos de Pearson, a TWD é aplicada para obter os coeficientes de detalhe de primeiro e segundo nível, cD_1 e cD_2 , e os envelopes de aceitação são construídos a partir dos 0.01° e $99,99^\circ$ percentis da distribuição empírica de cD_1 e cD_2 , respectivamente. Mais uma vez, os resultados mostram que os envelopes de aceitação variam não apenas com o tamanho da amostra N , mas também com a combinação dos valores dos parâmetros (α, λ) . Portanto, assumindo uma estratégia conservadora, para cada tamanho de amostra, é escolhido um envelope de aceitação com a amplitude mínima.

5 Ilustração

Nesta seção ilustram-se com dados reais os procedimentos de detecção do tempo de ocorrência de *outliers* descritos anteriormente. Para isso, consideramos uma série temporal com 241 observações relativas ao número de diferentes endereços IP, registrados em períodos de 2 minutos, que acedem ao servidor do Departamento de Estatística da Universidade de Würzburg, entre as 10h e as 18h do dia 29 de novembro de 2005, representada na Figura 1. O valor da média amostral ($\bar{x} = 1.32$), da variância amostral ($\hat{\sigma}^2 = 1.39$) e a análise das funções de autocorrelação e autocorrelação parcial amostrais indicam que pode ser ajustado um modelo PoINAR(1) a este conjunto de dados. A aplicação dos procedimentos descritos anteriormente permitem detetar a ocorrência de um *outlier* na observação $t = 224$ (correspondendo a $S = \{112\}$). Na Figura 2 estão representados o limiar e o envelope de aceitação para este conjunto de dados. A detecção de *outlier* em $t = 224$ está de acordo com os resultados de Weiß[20] e Silva e Pereira [16]. A estimativa CLS proposto por Barczy *et al.* [3] para o tamanho do *outlier* é $\hat{\omega} = 6.79$, o que significa que o valor verdadeiro da observação 224 é aproximadamente igual a 1. Note-se que o trabalho de Weiß[20] indica que o valor verdadeiro é $X_{224} = 1$ enquanto que Silva e Pereira [16] detetam um *outlier* em $t = 224$ com probabilidade 0.99 e estimativas dos parâmetros dadas por $\hat{\alpha} = 0.27, \hat{\lambda} = 0.89$ e $\omega = 7$.

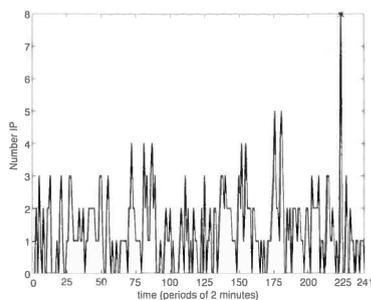


Figura 1: Cronograma da série temporal IP.

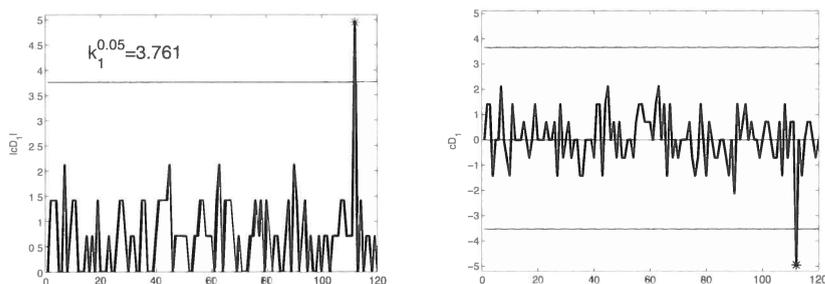


Figura 2: Resultados da detecção do tempo de ocorrência do *outlier* na série temporal IP, com a abordagem do limiar (lado esquerdo) e com a abordagem da reamostragem paramétrica (lado direito).

6 Observações finais

As metodologias apresentadas para a detecção do tempo de ocorrência de *outliers* em modelos PoI-NAR(1) não requerem o conhecimento prévio do número de *outliers* e adequam-se aos casos de um único ou múltiplos *outliers*, do tipo aditivo ou inovacional, assim como em grupo (*patches*). Contudo, a discriminação do tipo de *outlier* é ainda um tema em aberto.

Adicionalmente, os procedimentos propostos podem ser aplicados noutros contextos e também podem ser estendidos para detetar mudanças na estrutura e dinâmica dos processos. Nestes casos, será necessário calibrar os percentis das distribuições empíricas usadas para detetar o tempo de ocorrência de *outliers*, seja na abordagem do limiar ou na abordagem da reamostragem paramétrica. É ainda possível que diferentes aplicações precisem de diferentes níveis de decomposição na TWD.

Agradecimentos Este trabalho foi parcialmente financiado pela FCT- Fundação para a Ciência e a Tecnologia, através do Centro de Investigação e Desenvolvimento em Matemática e Aplicações (CIDMA), projeto UID/MAT/04106/2019.

Referências

- [1] Al-Osh, M. A. e Alzaid, A. A. (1987). First-order integer-valued autoregressive (INAR(1)) process. *J. Time Ser. Anal.* **8**, 261–275.
- [2] Barczy, M., Ispány, M., Pap, G., Scotto, M. e Silva, M. E. (2010). Innovational Outliers in INAR(1) Models. *Commun. Stat. - Theor. M.* **39**, 3343–3362.
- [3] Barczy, M., Ispány, M., Pap, G., Scotto, M. e Silva, M. E. (2011). Additive outliers in INAR(1) models. *Stat. Pap.* **53**, 935–949.

- [4] Bilen, C. e Huzurbazar, S. (2002). Wavelet-Based Detection of Outliers in Time Series. *J. Comp. Graph. Stat.* **11**, 311–327.
- [5] Bourguignon, M. e Vasconcellos, K. L. P. (2018). The effects of additive outliers in INAR(1) process and robust estimation, *Stat. Theory Relat. Fields* **2**, 206–214.
- [6] Chang, I., Tiao, G. C. e Chen, C. (1988). Estimation of time series parameters in the presence of outliers. *Technometrics* **30**, 193–204.
- [7] Chen, C. e Liu, L. M. (1993). Joint estimation of model parameters and outlier effects in time series. *J. Am. Stat. Assoc.* **88**, 284–297.
- [8] Fox, A. J. (1972). Outliers in Time Series. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **34**, 350–363.
- [9] Grané, A. e Veiga, H. (2010). Wavelet-based detection of outliers in financial time series. *Comput. Stat. Data An.* **54**, 2580–2593.
- [10] Mallat, S. G. (1989). A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 674–693.
- [11] McKenzie, E. (1985). Some simple models for discrete variate time series. *Water Resources Bull.* **21**, 645–650.
- [12] Percival, D. e Walden, A. (2006). *Wavelet methods for time series analysis*, Cambridge Series in Statistical and Probabilistic Mathematics, New York, Cambridge University Press.
- [13] Scotto, M. G., Weiß, C. H. e Gouveia, S. (2015). Thinning-based models in the analysis of integer-valued time series: a review. *Stat. Modelling* **15**, 590–618.
- [14] Silva, I. e Silva, M. E. (2018). Wavelet-based detection of outliers in Poisson INAR(1) time series, In Oliveira, T. A., Kitsos, C. P., Oliveira, A., Grilo, L. (Eds), *Contributions to Statistics - Recent Studies on Risk Analysis and Statistical Modeling*, Springer, 183–195.
- [15] Silva, M. E. e Oliveira, V. L. (2004). Difference equations for the higher-order moments and cumulants of the INAR(1) model. *J. Time Ser. Anal.* **25**, 317–333.
- [16] Silva, M. E. e Pereira, I. (2015). Detection of additive outliers in Poisson INAR(1) time series. In Bourguignon, J. P. et al. (Eds.) *CIM Series in Mathematical Sciences - Mathematics of Energy and Climate Change*, Springer, 377–388.
- [17] Steutel, F. W. e Van Harn, K. (1979). Discrete analogues of self-decomposability and stability. *Ann. Probab.* **7**, 893–899.
- [18] Tsay, R. S. (1986). Time series model specification in the presence of outliers. *J. Am. Stat. Assoc.* **81**, 132–141.
- [19] Tsay, R. S. (1992). Model checking via parametric bootstraps in time series analysis. *J. R. Stat. Soc. Ser. C Appl. Stat.* **41**, 1–15.
- [20] Weiß, C. H. (2007). Controlling correlated processes of Poisson counts. *Qual. Reliab. Eng. Int.* **23**, 741–754.



Índice

| | |
|---|----|
| Editorial | 2 |
| Mensagem da Presidente | 3 |
| XXIV Congresso - Bolsas de Participação | 4 |
| Notícias | 5 |
| <i>Enigmística</i> | 11 |

Séries Temporais de Valor Inteiro

| | |
|---|----|
| Introdução à teoria dos operadores thinning na modelação de séries temporais de valores inteiros <i>Manuel G. Scotto</i> | 12 |
| Métodos de deteção de outliers baseados em wavelets: o caso dos modelos INAR(1) de Poisson <i>Isabel Silva e Maria Eduarda Silva</i> | 22 |
| Modelos de contagem com estrutura periódica <i>Isabel Pereira, Magda Monteiro e Cláudia Santos</i> | 29 |
| Uso de distribuições geométricas autorregressivas na análise de sequências de ADN <i>Sónia Gouveia</i> | 39 |
| Cartas de controlo para o valor esperado de um processo INAR(1) com função ARL sem viés <i>Manuel Cabral Morais</i> | 46 |
| CP-INGARCH: uma classe geral de modelos para séries de contagem <i>Filipa Alexandra Cardoso da Silva</i> | 53 |

Ciência Estatística

| | |
|---|----|
| <i>Teses de Doutoramento</i> | 61 |
| <i>Livros e Capítulos de Livros</i> | 64 |
| Retrospectiva do Boletim SPE | 65 |
| Edições SPE | 66 |
| Prémios “Estatístico Júnior 2019” | 67 |
| “Prémio SPE” | 68 |