Yule-Walker estimation of periodic INAR signals

Márton Ispány

Faculty of Informatics, University of Debrecen
Hungary

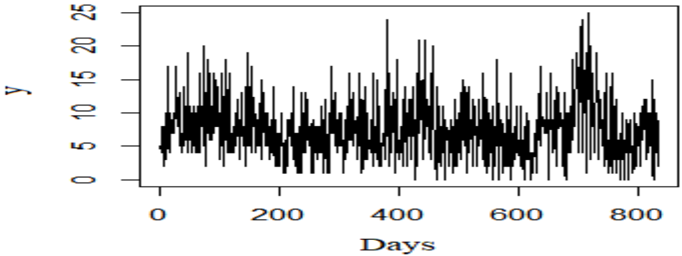
The 2024 IEEE 3rd Conference on Information Technology and Data Science
CITDS 2024
Debrecen, Hungary
26–28 Aug, 2024

Outline

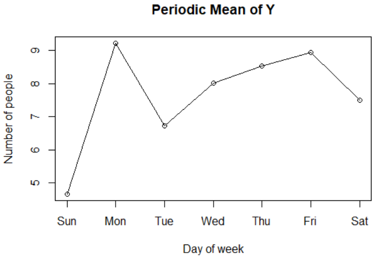
1. Data1: daily number of people who got antibiotics for the treatment of respiratory problems
2. Data2: daily number of parcels picked up from one pickup point (PUP)
3. Periodic integer-valued autoregressive (PINAR) process
4. Yule-Walker estimation
5. Simulation study
6. Real data applications
7. Conclusions

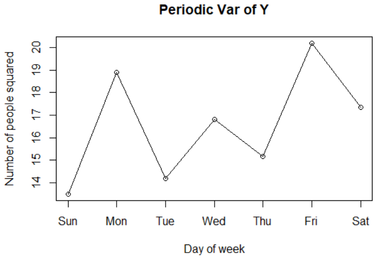
Data1

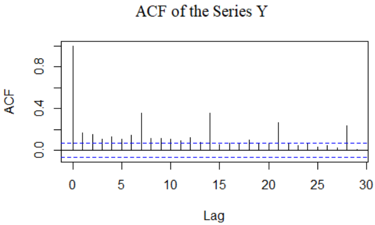
- Time series: The daily number of people who got antibiotics for treating respiratory problems from the public health care system in the emergency service.
- Duration: May 26, 2013, to September 05, 2015, resulting in $T = 833$ daily ($n = 119$ weeks) observations.
- Source:This real data set was obtained from the network records system welfare of the municipality Vitória-ES, Brazil. (Filho et al., 2021)

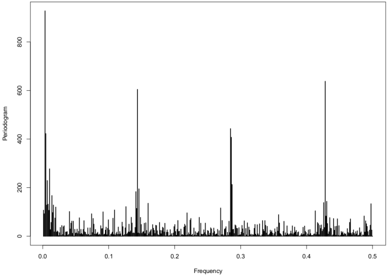


Basic plots of Data1









Periodic ACF and PACF of Data1

Autocovariance function of periodic time series with period $S \in \mathbb{N}$: $\gamma_s(h) = \mathrm{Cov}(Y_t, Y_{t-h})$, where $s = 1, \ldots, S$, $h \in \mathbb{N}_0$, such that $t \equiv s \mod S$.

Periodic ACF

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$
Sunday	0.01	**0.26**	**0.18**	**0.24**	**0.28**	0.11	0.15	0.02	0.07	**0.29**
Monday	**0.38**	-0.12	0.14	**0.23**	**0.18**	**0.19**	**0.29**	0.13	-0.11	0.04
Tuesday	**0.33**	**0.34**	-0.02	0.10	**0.23**	**0.39**	**0.42**	**0.18**	**0.37**	0.03
Wednesday	**0.27**	0.05	**0.17**	0.10	0.16	**0.33**	**0.29**	**0.23**	0.14	0.14
Thursday	**0.18**	**0.36**	**0.23**	**0.31**	0.01	**0.18**	**0.29**	**0.22**	**0.25**	0.11
Friday	**0.25**	0.16	**0.20**	0.14	0.16	**0.17**	**0.18**	**0.30**	**0.23**	0.13
Saturday	**0.20**	0.10	-0.03	-0.05	**−0.18**	0.03	**0.30**	0.10	-0.07	0.16

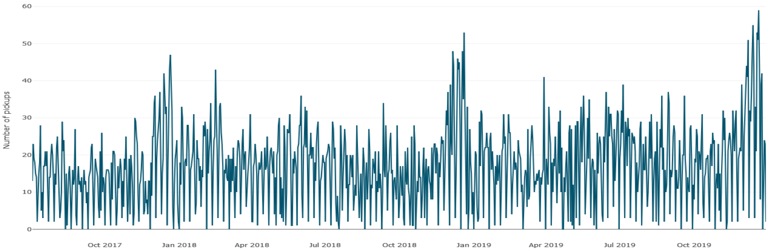
Periodic PACF

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$
Sunday	0.01	**0.26**	0.12	**0.20**	**0.18**	0.00	0.03	-0.02	-0.01	0.16
Monday	**0.38**	-0.14	0.08	**0.18**	0.07	0.01	**0.22**	-0.06	-0.08	-0.02
Tuesday	**0.33**	**0.24**	0.00	-0.00	0.15	**0.32**	**0.29**	0.01	**0.26**	0.04
Wednesday	**0.27**	-0.04	0.10	0.10	0.11	**0.27**	**0.18**	0.03	0.09	-0.07
Thursday	**0.18**	**0.33**	0.13	**0.18**	0.01	0.10	**0.17**	0.04	0.02	-0.02
Friday	**0.25**	0.12	0.10	0.06	0.03	**0.18**	0.08	**0.18**	0.13	-0.05
Saturday	**0.20**	0.05	-0.07	-0.11	**−0.21**	0.08	**0.26**	0.03	-0.13	**0.21**

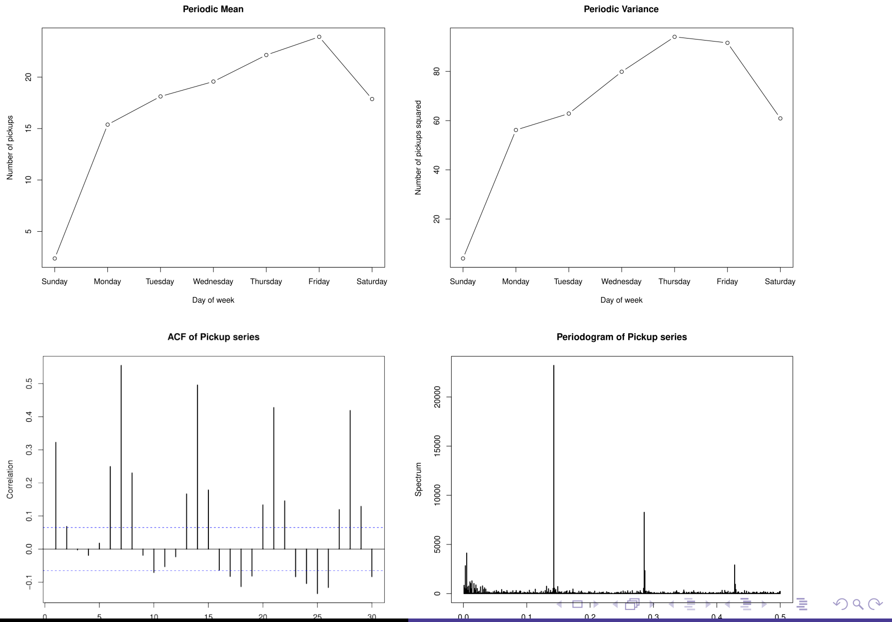
bold: significant correlation

Data2

- Time series: The daily number of parcels picked up from one pickup point (PUP) at a PUP management company.
- Duration: July 3, 2017 to December 29, 2019, resulting in $T = 910$ daily ($n = 130$ weeks) observations.
- Source: https://github.com/cabani/ForecastingParcels (Nguyen, et al., 2023)



Basic plots of Data2



Periodic ACF

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$
Sunday	0.072	0.008	0.118	-0.021	0.036	0.120	-0.042	0.000	-0.058	-0.012
Monday	**0.261**	**0.215**	**0.370**	**0.287**	**0.321**	0.075	0.169	0.084	0.184	0.186
Tuesday	**0.328**	**0.438**	**0.241**	**0.208**	0.081	**0.281**	0.060	0.168	**0.205**	0.115
Wednesday	**0.548**	**0.479**	**0.373**	**0.215**	**0.342**	0.171	**0.222**	**0.238**	**0.238**	**0.232**
Thursday	**0.486**	**0.450**	**0.196**	**0.278**	**0.196**	**0.222**	**0.308**	**0.406**	**0.245**	0.096
Friday	**0.521**	0.149	**0.381**	**0.351**	**0.314**	**0.398**	**0.368**	**0.363**	0.097	**0.337**
Saturday	**0.244**	**0.332**	**0.238**	**0.341**	**0.312**	**0.443**	**0.406**	**0.260**	**0.234**	0.135

Periodic PACF

	$h=1$	$h=2$	$h=3$	$h=4$	$h=5$	$h=6$	$h=7$	$h=8$	$h=9$	$h=10$
Sunday	0.072	-0.011	**0.174**	**0.237**	0.009	-0.037	-0.068	-0.045	-0.088	-0.021
Monday	**0.261**	0.142	0.114	**0.240**	0.091	0.015	0.033	0.081	0.067	-0.022
Tuesday	**0.328**	**0.326**	**0.260**	-0.116	0.085	0.142	0.118	-0.045	0.027	0.005
Wednesday	**0.548**	**0.290**	0.002	0.134	-0.004	**0.297**	0.065	**0.204**	-0.086	-0.156
Thursday	**0.486**	**0.264**	0.064	-0.008	0.152	0.117	0.152	0.063	-0.104	0.017
Friday	**0.521**	0.027	0.096	0.109	-0.021	-0.025	-0.081	**0.202**	-0.080	0.010
Saturday	**0.244**	**0.327**	**0.258**	0.074	0.164	0.157	0.003	-0.052	0.020	0.158

bold: significant correlation

Observations:

- periodicity (varying environment)
- strong seasonality

Overviews on count time series

- Davis RA, Fokianos K, Holan SH, Joe H, Livsey J, Lund R, Pipiras V, Ravishanker N. (2021) Count time series: a methodological review. Journal of the American Statistical Association 116:1533–1547.
- Davis RA, Holan SH, Lund R, Ravishanker N. (2016) Handbook of Discrete-Valued Time Series CRC Press, New York, NY.
- Fokianos, K. (2012) Count time series models. Handbook of Statistics., C. R. Rao, C. Rao, and V. Govindaraju (eds); 30, 315–347, Elsevier, Amsterdam.
- Weiss CH. (2018) An Introduction to Discrete-Valued Time Series. John Wiley & Sons, Boca Raton, FL.

Branching process with immigration (BPI)

$$X_k = \sum_{j=1}^{X_{k-1}} \xi_{k,j} + \varepsilon_k, \quad k \in \mathbb{N}, \quad X_0 = 0$$

$\boldsymbol{\xi}_k := \{\xi_{k,j} \mid j \in \mathbb{N}\}$ and $\{\varepsilon_k \mid k \in \mathbb{N}\}$ i.i.d. sequences of r.v.'s
To avoid degeneracy: $\mathsf{P}(\varepsilon = 0) < 1$
Reformulation by (generalized) thinning operator:

$$X_k = \boldsymbol{\xi}_k \circ X_{k-1} + \varepsilon_k, \quad k \in \mathbb{N}, \quad X_0 = 0$$

Parameters: $m := \mathsf{E}[\xi]$, $\sigma^2 := \mathsf{Var}[\xi]$, $\lambda := \mathsf{E}[\varepsilon]$, $b^2 := \mathsf{Var}[\varepsilon]$
Existence: always since $k \in \mathbb{N}_0$

Classification:

$m < 1$	$m = 1$	$m > 1$
subcritical	**critical**	**supercritical**

Integer valued autoregression (INAR)

INAR(1) model, Al-Osh and Alzaid (1987), McKenzie (1985)

$$Y_k = \alpha_k \circ Y_{k-1} + \varepsilon_k, \quad k \in \mathbb{Z}$$

$\{\alpha_k \circ \mid k \in \mathbb{Z}\}$, are i.i.d. binomial thinning operators with parameter $\alpha \in [0, 1]$, $\{\varepsilon_k \mid k \in \mathbb{Z}\}$ are i.i.d. $\mathbb{N}_0$-valued r.v.'s
To avoid degeneracy: $\mathrm{P}(\varepsilon = 0) < 1$
Reformulation in BPI form:

$$Y_k = \sum_{j=1}^{Y_{k-1}} \xi_{k,j} + \varepsilon_k, \quad k \in \mathbb{Z}$$

$\{\xi_{k,j} \mid k \in \mathbb{Z}, j \in \mathbb{N}\}$ i.i.d. Bernoulli r.v.'s with mean α

Parameters: $\alpha := \mathrm{E}[\xi]$, $\lambda := \mathrm{E}[\varepsilon]$, $b^2 := \mathrm{Var}[\varepsilon]$

Existence: $\alpha < 1$ (subcritical)

In the critical case $\alpha = 1$, there is no solution!

Periodic INAR signals

$\{Y_t\}$ is called periodic INAR (PINAR) process with period $S \in \{2, 3, \ldots\}$ and autoregressive orders $\boldsymbol{p} := \{p_s\}$ if

$$Y_{kS+s} = \sum_{i=1}^{p_s} \alpha_{s,i}^k \circ Y_{kS+s-i} + \varepsilon_{kS+s},$$

$k \in \mathbb{Z}$, where $\{\alpha_{s,i}^k \circ\}$ are i.i.d. binomial thinning operators with mean $\alpha_{s,i} \geq 0$ for all $i = 1, \ldots, p_s, s = 1, \ldots, S$

Y_{kS+s}: the value of series during the sth season of period k

Autoregressive parameters: $\{\alpha_{s,i}, \mid i = 1, \ldots, p_s, s = 1, \ldots, S\}$

Immigration process $\{\varepsilon_t\}$: periodic sequence of $\mathbb{N}_0$-valued r.v.'s, i.e., for each s, $\{\varepsilon_{kS+s} \mid k \in \mathbb{Z}\}$ are i.i.d.r.v.'s

Immigration parameters: $\lambda_s = \mathrm{E}(\varepsilon_{kS+s}) \in \mathbb{R}_+$, and $\sigma_s^2 = \mathrm{Var}(\varepsilon_{kS+s}) > 0$, $k \in \mathbb{Z}$.

Periodic AR (PAR) model:

$$X_{kS+s} = \sum_{i=1}^{p_s} \alpha_{s,i} X_{kS+s-i} + \varepsilon_{kS+s}$$

Periodic time series models

Examples:

- PINAR(1) model, Monteiro et al. (2010)
- $\text{INAR}(1)_S$ model, Bourguignon, Vasconcellos, Reisen, I (2014), Buteikis and Leipus (2020)
- PINAR(2) model with periodic immigration, Morina et al. (2011)
- $\text{PINAR}(1, S)$ model, Filho et al. and I (2021, 2024)
- PINARMA(p,q) model Bentarzi and Aries (2020)

Periodic models (PARMA): Gladyshev (1961), Gardner et al. (2006), and Hurd and Miamee (2007). Inference: Lund and Basawa (2000) and Basawa and Lund (2001).
Seasonal models (SARMA): Chatfield and Prothero (1973)
SPARIMA models: Basawa, Lund, and Shao (2004), Koopman et al. (2006), Hindrayanto et al. (2010)

BP(I) in varying environment

General inhomogeneous branching processes:

- Domain of peer-to-peer file sharing networks, Adar and Huberman (2000), Zhao et al. (2005)
- Modeling biodiversity or macroevolution, Aldous and Popovic (2005), Haccou and Iwasa (1996)
- Epidemic–type Aftershock Sequence (ETAS) in seismology, Farrington et al. (2003)

Heterogeneous INAR models (Bernoulli offspring):

- Understanding and predicting consumers' buying behavior, Böckenholt (1999)
- Modeling the premium in the bonus-malus scheme of car insurance, Gourieroux and Jasiak (2004)

The supercritical case was studied by Goettge (1976), Cohn and Hering (1983), Jagers and Nerman (1985), D'Souza and Biggins (1992, 1993), D'Souza (1994).

State-space (MBPI) representation

Main tool: stationary vector representation á la Gladyshev

Define the matricial thinning operator $\Xi\circ = (\xi_{i,j}\circ)$, Latour (1997), as

$$(\Xi \circ \boldsymbol{Y})_i = \sum_{j=1}^{S} \xi_{i,j} \circ Y_j, \quad i = 1, \ldots, S$$

Let $\boldsymbol{Y}_k = (Y_{kS+1}, \ldots, Y_{kS+S})^\top$ and $\boldsymbol{\varepsilon}_k = (\varepsilon_{kS+1}, \ldots, \varepsilon_{kS+S})^\top$

Implicit state-space representation by VINAR$_S(p)$ model:

$$\boldsymbol{Y}_k = \sum_{i=0}^{p} A_i^k \circ \boldsymbol{Y}_{k-i} + \boldsymbol{\varepsilon}_k$$

where $\{A_i^k\circ\}$ are $S \times S$ matricial binomial thinning operators defined as $A_i^k\circ = (a_{r,s}^{k,i}\circ)$, $i = 0, 1, \ldots, p$, where $a_{r,s}^{k,0}\circ := \alpha_{r,r-s}^k\circ$ if $r > s$ and $0\circ$ otherwise (i.e., if $r \le s$), and $a_{r,s}^{k,i}\circ := \alpha_{r,iS+r-s}^k\circ$ for all $i = 1, \ldots, p$ and $r, s = 1, \ldots, S$, where $\alpha_{s,i}^k\circ := 0\circ$ if $i > p_s$ for all $k \in \mathbb{Z}$.

Main theorem

Assumptions. Suppose that $\mathsf{E}\|\varepsilon_0\| < \infty$ and $\rho(A_0 + A_1 + \ldots + A_p) < 1$, where A_i denotes the mean matrix of $A_i^k \circ$, $i = 0, 1, \ldots, p$.

The VINAR$_S(p)$ model has a unique non-anticipative, strictly stationary solution $\{\boldsymbol{Y}_k\}$ with finite mean vector. The $\mathbb{N}_0^S$-valued stochastic process $\{\boldsymbol{Y}_k\}$ is $\{\mathcal{G}_k\}$-adapted, ergodic pth order homogeneous Markov chain.

The PINAR$_S(\boldsymbol{p})$ model has a unique non-anticipative, periodically strictly stationary solution $\{Y_t\}$ with finite mean and period S. The $\mathbb{N}_0$-valued stochastic process $\{Y_t\}$ is $\{\mathcal{H}_t\}$-adapted, periodic inhomogeneous Markov chain with orders p_s, $s = 1, \ldots, S$.

Examples

Consider the $\text{PINAR}(1)_S$ model. The characteristic polynomial of this model simplifies to $P(z) = z^S - \prod_{j=1}^S \alpha_j$ and the condition $\rho(A_0 + A_1) < 1$ is equivalent to $\prod_{j=1}^S \alpha_j < 1$.

Consider the case $S = 2$, with $p_1 = p_2 = 2$, i.e., the $\text{PINAR}_2(2, 2)$ model. Then

$$A_0 = \begin{bmatrix} 0 & 0 \\ \alpha_{2,1} & 0 \end{bmatrix}, \quad A_1 = \begin{bmatrix} \alpha_{1,2} & \alpha_{1,1} \\ 0 & \alpha_{2,2} \end{bmatrix}, \quad A_0 + A_1 = \begin{bmatrix} \alpha_{1,2} & \alpha_{1,1} \\ \alpha_{2,1} & \alpha_{2,2} \end{bmatrix}$$

The characteristic polynomial is given by

$$P(z) = (z - \alpha_{1,2})(z - \alpha_{2,2}) - \alpha_{1,1}\alpha_{2,1}$$

It can be seen that $\alpha_{1,2} + \alpha_{2,2} - \alpha_{1,2}\alpha_{2,2} + \alpha_{1,1}\alpha_{2,1} < 1$ is a necessary and sufficient stationarity condition. Note that this condition can be rewritten as $\alpha_{1,1}\alpha_{2,1} < (1 - \alpha_{1,2})(1 - \alpha_{2,2})$. See Darolles et al. (2019)

Yule-Walker estimation

Parameters of $\text{VINAR}_S(p)$: A_i's, and $\boldsymbol{\lambda} = \mathsf{E}(\boldsymbol{\varepsilon})$, $\Lambda = \mathsf{Var}(\boldsymbol{\varepsilon})$
First-order or Markovian state-space representation results in $\text{VINAR}_{Sp}(1)$ model with parameters: A_0, A_1, $\boldsymbol{\lambda}$, and Λ
Let $\boldsymbol{\mu} = \mathsf{E}(\boldsymbol{Y})$, $\Sigma = \mathsf{Var}(\boldsymbol{Y}) = \Gamma(0)$ and $\Gamma = \Gamma(1)$
We have the Yule-Walker equations:

$$\boldsymbol{\mu} = (I - A_0 - A_1)^{-1}\boldsymbol{\lambda}$$

$$\Sigma - \Gamma\Sigma^{-1}\Gamma^\top = (I - A_0)^{-1} D \left((I - A_0)^{-1}\right)^\top$$

$$\Gamma\Sigma^{-1} = (I - A_0)^{-1} A_1$$

where $D := \Lambda + \mathrm{diag}((V_0 + V_1)\boldsymbol{\mu})$ and V_0, V_1 are the Bernoulli variance of A_0, A_1. Number of equations: $(3/2)S(S+1)$
Scenario $S1$: the immigration is uncorrelated within a period
2nd equation: Cholesky factorization of a Schur complement
Scenario $S2$: the immigration is correlated within a period, the autoregressive orders are fixed S
3rd equation: LU factorization if A_1 is upper triangular

Result of the YW estimation for Pickup data under S1

col	1	2	3	4	5	6	7
	0	0	0	0	0	0	0
	0.164	0	0	0	0	0	0
	0.029	0.270	0	0	0	0	0
$\widehat{A}_0$	0.224	0.253	0.437	0	0	0	0
	0.096	-0.058	0.281	0.213	0	0	0
	0.072	-0.026	0.012	0.163	0.275	0	0
	-0.022	-0.003	0.024	0.016	-0.014	0.041	0
	-0.006	0.15	0.138	-0.037	0.165	0.155	-0.141
	-0.042	-0.065	-0.032	0.05	0.232	0.068	-0.226
	0.063	0.078	0.034	-0.034	0.078	0.221	0.204
$\widehat{A}_1$	0.002	0.025	-0.05	0.113	0.113	-0.001	-0.548
	0.127	-0.173	-0.072	0.247	0.014	0.234	-0.361
	0.122	0.121	0.013	-0.151	0.065	0.126	0.306
	-0.002	0.018	0.035	-0.017	-0.028	0.066	-0.131
$\widehat{\lambda}$	4.501	10.72	5.623	2.159	6.648	1.261	0.712

Result of the YW estimation for Pickup data under S2

col	1	2	3	4	5	6	7
$\widehat{A}_1 + \widehat{A}_0$	0.165	0.114	0.009	-0.067	0.151	0.261	0.085
	0.882	-0.22	-0.057	0.316	-0.014	0.119	-0.523
	0.58	-0.703	0.098	0.183	0.161	0.25	-0.505
	0.509	-0.503	-0.254	0.26	0.307	0.249	-0.889
	0.315	0.057	-0.024	0.067	0.285	0.148	-0.161
	0.371	0.005	0.252	0.003	0.208	0.093	0.192
	-0.24	0.027	0.083	-0.159	-0.213	0.655	-0.312
$\widehat{\lambda}$	3.581	2.095	10.309	12.971	17.621	-0.009	1.6176
$\widehat{\Lambda}$	34.79	-31.62	-17.22	-4.83	20.22	-11.55	8.56
	-31.62	75.94	64.14	53.66	-13.3	6.73	-6.25
	-17.22	64.14	129.02	94.83	16.87	-6.22	-4.84
	-4.83	53.66	94.83	137.49	22.87	2.47	2.16
	20.22	-13.3	16.87	22.87	69.41	-3.85	5.6
	-11.55	6.73	-6.22	2.47	-3.85	29.58	-23.68
	8.56	-6.25	-4.84	2.16	5.6	-23.68	29.1

Conclusions

- Real count time series are presented which possess periodicity and seasonality.
- A periodic INAR model is proposed.
- YW method under two scenarios estimates the model parameters.
- Monte Carlo simulations verify the effectiveness of the proposed methodology.
- The proposed model is successfully fitted to real data.

Thank you for your attention! [1]

References:

[1] Filho, Reisen, Bondon, Ispány, Melo and Serpa (2021) A periodic and seasonal statistical model for non-negative integer-valued time series with an application to dispensed medications in respiratory diseases. In: Appl Math Model , pp. 545-558.
[2] Ispány, Bondon, Reisen and Filho (2024) Existence of a periodic and seasonal non-negative integer-valued autoregressive process. J Time Ser Anal.

[3] Ispány, Bondon, Reisen: On existence of periodic non-negative integer-valued ARMA processes, manuscript

[1] This research was supported by the Project no. TKP2020-NKA-04 which has been implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020-4.1.1-TKP2020 funding scheme.