



UNIVERSITY OF KRAGUJEVAC
FACULTY OF SCIENCE

Bogdan A. Pirković

**CONTRIBUTION TO THE THEORY
OF RANDOM ENVIRONMENT
INTEGER-VALUED
AUTOREGRESSIVE PROCESSES**

Doctoral Dissertation

Kragujevac, 2022



UNIVERZITET U KRAGUJEVCU
PRIRODNO-MATEMATIČKI FAKULTET

Bogdan Pirković

**DOPRINOS TEORIJI CELOBROJNIH
AUTOREGRESIVNIH PROCESA
U SLUČAJNOJ OKOLOINI**

doktorska disertacija

Kragujevac, 2022

Author
Name and surname: Bogdan Pirković
Date and place of birth: 01/09/1988, Kragujevac, Serbia
Current employment: system analyst at Faculty of Science, University of Kragujevac
Doctoral Dissertation
Title: Contribution to the theory of random environment integer-valued autoregressive processes
No. of pages: 117
No. of images: 28
No. of bibliographic data: 50
Institution and place of work: Faculty of science, Kragujevac
Scientific area(UDC): Mathematical statistics (519.2)
Mentor: dr Aleksandar Nastić, full professor, Faculty of Science and Mathematics, University of Niš
Grade and Dissertation Defense
Topic Application Date: 30/03/2021
Decision number and date of acceptance of the doctoral dissertation topic: IV-01-682/10, dated on 16/09/2021
Commission for evaluation of the scientific merit of the topic and the eligibility of the candidate:
<ol style="list-style-type: none"> 1. dr Miroslav Ristić, full professor, Faculty of Science and Mathematics, University of Niš 2. dr Predrag Popović, assistant professor, Faculty of Civil Engineering and Architecture, University of Niš 3. dr Slađana Dimitrijević, assistant professor, Faculty of Science, University of Kragujevac
Commission for evaluation and defense of doctoral dissertation:
<ol style="list-style-type: none"> 1. dr Božidar Popović, associate professor, Faculty of Science and Mathematics, University of Montenegro 2. dr Marija Milošević, full professor, Faculty of Science and Mathematics, University of Niš 3. dr Predrag Popović, associate professor, Faculty of Civil Engineering and Architecture, University of Niš 4. dr Miodrag Đorđević, associate professor, Faculty of Science and Mathematics, University of Niš 5. dr Slađana Dimitrijević, assistant professor, Faculty of Science, University of Kragujevac
Date of Dissertation Defense:

A word from the author

After some time, you learn the subtle difference between holding a hand and imprisoning a soul, and you start to learn... and you start to accept defeat with the head up high and open eyes, and you learn to build all roads on today, because the terrain of tomorrow is too insecure for plans...and the future has its own way of falling apart in half. And you learn that if it's too much, even the warmth of the sun can burn.

So you plant your own garden and embellish your own soul, instead of waiting for someone to bring flowers to you. And you learn that you can actually bear hardship; that you are actually strong, and you are actually worthy, and you learn and learn...and so every day.

Over time you learn that real friends are few and whoever doesn't fight for them, sooner or later, will find himself surrounded only with false friendships. Over time you learn the word spoken in moments of anger continues hurting throughout a lifetime. Over time you learn that everyone can apologize, but forgiveness is an attribute solely of great souls. Over time you realize that every experience lived, with each person, is unrepeatable. Over time you comprehend that rushing things or forcing them to happen causes the finale to be different from expected. Over time you realize that in fact the best was not the future, but the moment you were living just that instant.

But unfortunately, only over time...

Jorge Luis Borges

In addition to the already mentioned, over time I have learned to show gratitude, to say thank you. The first portion of gratitude goes to prof. dr Aleksandar Nastić. I was so fortunate to have such an astonishing PhD supervisor. His greatness is reflected not only through a mathematical excellence, but through a great character as well. Not only he had been a remarkable mentor, he had been a remarkable person. He offered me something that many never get, a second chance. I am grateful for all the knowledge, patience and wisdom he shared with me through all these years.

Many thanks go to the members of the commission for evaluation and defense of doctoral dissertation: prof. dr Božidar Popović, prof. dr Marija Milošević, prof. dr Predrag Popović and prof. dr Miodrag Đorđević for great attention they paid to this dissertation and for useful advices and suggestions they gave.

Special thanks go to doc. dr Slađana Dimitrijević. As a witness to my complete mathematical development from early teenage days onwards, she has contributed to this development in so many ways. It would be pointless to list them all. Nevertheless, the

contribution she made to this dissertation through careful reading and constructive tips must be highlighted.

My endless "thank you" goes to entire Pirković family, especially to my parents, mom Ljubinka and the late dad Aleksandar. It was not an easy task to be my parent and make a man of me. But this is just what they did, with all their love, support and selflessness. I spent many nights figuring out a proper way to repay them. But this is just impossible. I'll be in their debt 'till the day I die.

Last, but not the least, I would like to thank all those good people around me, who kept standing beside without any justifiable reason, who did not leave even when they had every right to do so. They are few, but so special. If you recognize yourself in these lines my dear friend, be sure your courage, loyalty and persistence are noticed. And trust me, great times are about to start.

Abstract

This dissertation has 2 basic goals. The first goal is to construct the random environment *INAR* time series that can take both positive and negative values. The realization of such goal would create new possibilities in integer-valued data modeling. In addition, since the environment state estimation of each individual realization is a crucial step in real-life data modeling using models in random environment, the goal is to adapt existing clustering techniques in order to make the environment state estimates more accurate. Both goals, if realized, would represent an original authorial contribution to the integer-valued time series analysis.

The dissertation contains 4 chapters. Chapter 1 is the introductory one and provides a historical overview of the *INAR* models development. Also, this chapter offers important theorems and distributions known from before, necessary to adduce proofs in subsequent chapters. Relying on results given in [15], Chapter 2 discusses possibilities of extracting and predicting latent components of the true *INAR* time series with skewed *Skellam* marginal distribution. In Chapter 3, a construction of the new non-stationary random environment *INAR* model with values over entire \mathbb{Z} is given. Unknown model parameters are estimated using adapted estimation techniques. The efficiency of estimates is tested on simulated data. A quality of the introduced model is examined on appropriate real-life data. In Chapter 4, the K-means clustering technique adaptation is provided, in order to make it suitable for estimating environment states of realizations corresponding to the generalized random environment *INAR* time series. The adaptation efficiency is tested on simulated and real-life data and compared to clustering results obtained using standard K-means.

Key words: *INAR*(1), *DLINAR*(1), thinning operator, random environment, discrete Laplace marginals, geometric marginals, K-means technique, state estimation, Markov chain

Abstrakt

Ova disertacija ima 2 cilja. Najpre, cilj disertacije je konstrukcija novih *INAR* vremenskih serija u slučajnoj okolini koji mogu uzeti kako pozitivne, tako i negativne vrednosti. Uspešna realizacija ovog cilja donela bi nove mogućnosti u modeliranju celobrojnih nizova podataka. Dodatno, kako je ocena stanja okoline svake realizacije ključni korak u modeliranju stvarnih procesa pomoću novouvedenih modela u slučajnoj okolini, cilj disertacije je prilagođavanje postojećih metoda klasterovanja sa namerom da ocene stanja budu što preciznije. Oba navedena cilja bi, u slučaju realizacije, predstavljala originalan doprinos autora analizi celobrojnih vremenskih serija.

Disertacija sadrži 4 glave. Glava 1 je uvodnog karaktera i daje istorijski pregled razvoja *INAR* modela. Takođe, ova glava nudi neke bitne teoreme i raspodele poznate od ranije, neophodne za izvođenje dokaza u narednim glavama. Oslanjajući se na rezultate date u [15], u Glavi 2 su razmotrene mogućnosti identifikovanja i predviđanja latentnih komponenti *INAR* vremenske serije sa asimetričnom Skelamovom marginalnom raspodelom. U Glavi 3 pristupa se konstrukciji novog nestacionarnog *INAR* modela u slučajnoj okolini koji može uzeti vrednosti na čitavom skupu \mathbb{Z} . Nepoznati parametri modela ocenjeni su pomoću prilagođenih tehnika ocenjivanja. Efikasnost ocena je testirana na simuliranim podacima. Kvalitet modela ispitan je na odgovarajućim realnim nizovima podataka. U Glavi 4 pristupa se adaptaciji K-means tehnike klasterovanja, sa ciljem da se ona prilagodi ocenjivanju stanja okoline realizacija koje odgovaraju uopštenoj *INAR* vremenskoj seriji u slučajnoj okolini. Efikasnost adaptacije testirana je na simuliranim podacima i upoređena sa rezultatima klasterovanja dobijenim pomoću standardne K-means tehnike.

Ključne reči: *INAR*(1), *DLINAR*(1), tining operator, slučajna okolina, diskretna Laplasova marginalna raspodela, geometrijska marginalna raspodela, K-means tehnika, ocena stanja, lanac Markova

Contents

1	<i>INAR</i> models-from the very beginnings to the present day	1
1.1	Introduction	1
1.2	Evolution of <i>INAR</i> models	2
1.2.1	Binomial thinning operator	2
1.2.2	<i>INAR</i> models based on the binomial thinning operator	3
1.2.3	Other models based on the counting sequence of Bernoulli trails	6
1.2.4	Negative binomial thinning operator	7
1.2.5	<i>INAR</i> models based on the negative binomial thinning operator	8
1.2.6	Mixed <i>INAR</i> models	10
1.2.7	Models with both positive and negative values	12
1.2.8	Random environment process	15
1.2.9	<i>INAR</i> models based on the random environment process	18
1.3	Important distributions	20
1.3.1	Skellam distribution	21
1.3.2	Discrete Laplace distribution	22
1.4	Important theorems	23
2	Extracting and predicting latent components of the skewed <i>TINAR</i>(1) time series	25
2.1	Symmetric <i>TINAR</i> (1) time series	25
2.1.1	Construction of the model	26
2.1.2	Properties of the model	27
2.1.3	Symmetric <i>TINAR</i> (1) model with negative lag-one autocorrelation	31
2.1.4	Parameter estimation of the symmetric <i>TINAR</i> (1) model	32
2.2	Skewed <i>TINAR</i> (1) time series	33
2.2.1	Construction of the model	33
2.2.2	Properties of the model	34
2.2.3	Parameter estimation of the skewed <i>TINAR</i> (1) model	36
2.3	Latent components of the skewed <i>TINAR</i> (1) time series-extraction and prediction	38
2.4	Simulation study	43
2.5	Application to real-life data	45
3	Random Environment Integer-Valued Autoregressive model with discrete Laplace marginal distribution	48
3.1	Construction of the model	49
3.2	Properties of the model	53
3.2.1	The k-step ahead conditional expectation	54

3.2.2	Correlation structure	55
3.3	Parameter estimation of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model	57
3.3.1	Yule-Walker estimation	58
3.3.2	Conditional least squares estimation	61
3.4	Simulations study-Estimates quality analysis	62
3.4.1	The case of two states	62
3.4.2	The case of three states	63
3.4.3	Estimation results	63
3.5	Application to real-life data	66
3.5.1	Fitting quality evaluation	68
3.5.2	Forecasting accuracy assessment	70
4	Random Environment Estimation (<i>RENES</i>) Method for Generalized Random Environment <i>INAR</i> Models of Higher Order	71
4.1	Construction of the new <i>RENES</i> method	74
4.2	Simulation study—the choice of model parameters	77
4.3	Simulation study — simulation results and selection of the <i>RENES</i> method parameters	78
4.4	Application to real-life data	92
	Conclusion	96
	Bibliography	98
	Appendix	102
	Appendix A	102
	Appendix B	103
	Appendix C	115
	Biography	

Chapter 1

INAR models-from the very beginnings to the present day

1.1 Introduction

Many natural or social phenomena often have different outcomes, despite the same circumstances in which they take place. These outcomes might be observed as realizations of some random variables, and thus become suitable to be described by usage of particular statistical model.

Let one observe the outcomes of a phenomenon of interest in equal time intervals. Then, the obtained outcomes represent the realization of one time series. In order to define the term "time series", a more general term from which this new term derives must be mentioned first. To that purpose, let $(\Omega, \mathcal{F}, \mathcal{P}_m)$ be a probability space, where Ω represents the set of all possible outcomes of the experiment, \mathcal{F} represents σ -algebra of events made of outcomes from Ω and \mathcal{P}_m represents a probability measure.

DEFINITION 1.1.1 ([7], Definition 1.2.1). *A stochastic process is a family of random variables $\{X_t, t \in T\}$ defined on a probability space $(\Omega, \mathcal{F}, \mathcal{P}_m)$, where T is an index set.*

Different types of stochastic processes can be noticed, depending on the cardinality of the index set. A stochastic process is said to be a continuous-time stochastic process if the index set T is uncountably infinite set. On the other hand, a stochastic process is said to be in discrete time, if the index set T has finite or countable number of elements. In particular, a stochastic process whose index set is a subset of the set of integers \mathbb{Z} is of special importance for this research.

DEFINITION 1.1.2. *A time series is a stochastic process $\{X_t, t \in \mathbb{Z}\}$ defined on the probability space $(\Omega, \mathcal{F}, \mathcal{P}_m)$.*

Remark. In probability theory, the term "time series" usually represents a stochastic process in discrete time. However, this term may denote the sequence of observed values of the stochastic process, that is, a realization of the stochastic process. By agreement, the same term is used in literature to denote the stochastic process itself and its realization. This agreement will be followed in this dissertation as well.

The basic problem is to choose the proper time series model, which would be a good representation of the observed real-life data. The data can occur, for instance, as a result of

registering the values of the particular phenomenon, or as a result of counting realizations of the phenomenon in corresponding time intervals. Many factors have an effect on mentioned choice, among them the very nature of the observed data. For instance, the data might consist of both positive or negative integers. On the other hand, the data values might be strictly nonnegative. This is usually the case when observed values are obtained as a result of counting the elements of the population. The aforementioned data can often be found in many fields, including medicine, economics, finance, telecommunications, criminology, sports. In order to model such data as well as possible, mathematicians used at first autoregressive time series with continuous marginal distributions. This gave acceptable results only when it comes to phenomena that generate extremely high realization values, where round-off error is negligible. On the other hand, in situations when observations registered over time are not that high (lower than 10^6), previously mentioned models cannot be used successfully. Slightly better results were achieved by involving Markov chains into the modeling procedure, as described in [9]. Unusually big number of parameters was the key obstacle to this idea. Few years later, several Discrete Autoregressive models of Moving Average (*DARMA*), based on well known *ARMA* models, were defined in [21], [22] and [23]. These models gave even better results. In mid 1980s, [34] and [2] introduced in different ways an Integer-valued Autoregressive model of order 1 (*INAR*(1)), based on the binomial thinning operator. This newly introduced model gave remarkable improvement in modeling the data that represents cardinality of the set whose elements, with respect to previous observation, might survive or not.

The choice of the thinning operator, as well as the marginal distribution, determines the very essence and the application potential of the *INAR* models. Over time, a large number of variations and generalizations of *INAR* models, in terms of their orders, thinning operators and marginal distributions, have emerged. The next section follows the development of this kind of models over decades, starting from the *INAR*(1) model introduced by [34] and [2] all the way to the present day.

1.2 Evolution of *INAR* models

As mentioned earlier, the first *INAR* model was introduced in mid 1980s by [34] and [2]. Following the approach given by Al-Osh and Alzaid, one concludes that the binomial thinning operator lays in the center of this idea. Thus, an essence and properties of this important operator should be brought closer to the reader.

1.2.1 Binomial thinning operator

Binomial thinning operator was introduced by [48] in the following way. Let X be an integer-valued random variable. For all $\alpha \in [0, 1]$, the operator " $\alpha \circ$ " is defined as

$$(1.1) \quad \alpha \circ X = \sum_{i=1}^X W_i,$$

where $\{W_i\}$, named as counting sequence, represents the sequence of independent and identically distributed (i.i.d.) Bernoulli trails, independent of X , distributed as $P(W_i = 1) = 1 - P(W_i = 0) = \alpha$. Thus, the distribution of $\alpha \circ X|X = x$ is binomial, with

distribution parameters x and α . Bearing in mind that the distribution of the variable $\alpha \circ X|X$ determines the name of the thinning operator, one may say " $\alpha \circ$ " is the binomial thinning operator. Further, some essential properties of the binomial thinning operator will be exposed.

Properties of the binomial thinning operator

Let $\alpha, \beta \in [0, 1]$ and let " $\alpha \circ$ ", " $\beta \circ$ " be operators defined as $\alpha \circ X = \sum_{i=1}^X W_i^{(1)}$, $\beta \circ X = \sum_{i=1}^X W_i^{(2)}$, where $\{W_i^{(1)}\}, \{W_i^{(2)}\}$ represent counting sequences of i.i.d. Bernoulli trails, with distribution parameters α and β , respectively. In this case, it holds:

1. $0 \circ X \stackrel{a.s.}{=} 0$;
2. $1 \circ X \stackrel{a.s.}{=} X$;
3. $\alpha \circ (\beta \circ X) = (\alpha\beta) \circ X$, $\alpha, \beta \in (0, 1)$, where counting sequences involved in $\alpha \circ X$ and $\beta \circ X$ are mutually independent;
4. $E(\alpha \circ X) = \alpha E(X)$;
5. $E(\alpha \circ X)^2 = \alpha^2 E(X^2) + \alpha(1 - \alpha)E(X)$;
6. $\alpha \circ (X + Y) \stackrel{d}{=} \alpha \circ X + \alpha \circ Y$, whereby the notation " $\stackrel{d}{=}$ " is used when random variables on both sides of the equality sign have the same distribution;
7. $Cov(X, \alpha \circ Y) = \alpha Cov(X, Y)$, where thinning operator " $\alpha \circ$ " is independent of random variable X ;
8. $E(\alpha \circ X|X) = \alpha X$, $\alpha \in (0, 1)$;
9. $E((\alpha \circ X)^2|X) = \alpha^2 X^2 + \alpha(1 - \alpha)X$, $\alpha \in (0, 1)$.

Proofs of these properties can be found in [46], [2], [13] and [14].

1.2.2 INAR models based on the binomial thinning operator

Al-Osh and Alzaid used predefined binomial thinning operator to introduce $INAR(1)$ models. These models have proven suitable for describing the data related to the counting of population elements, where elements can survive or vanish from the population with certain probability. This feature is a direct corollary of the fact that the binomial thinning operator is based on the counting sequence of Bernoulli trails, i.e. random variables with two possible realizations, 0 and 1.

DEFINITION 1.2.1 ([2]). *Time series $\{X_n\}$, given as*

$$(1.2) \quad X_n = \alpha \circ X_{n-1} + \varepsilon_n, \quad n \in \mathbb{Z},$$

is $INAR(1)$ time series, where $\alpha \in (0, 1)$, " $\alpha \circ$ " is defined with (1.1), $\{\varepsilon_n\}$ represents an innovation sequence of i.i.d. nonnegative integer-valued random variables with mathematical expectation μ_ε and finite variance σ_ε^2 , such that X_m and ε_n are independent random variables for all $m < n$.

An application of such defined models is related to the following interpretation. A size of the system X_n in moment n is consisted of two components: (i) the number of elements of the system survived from the previous observation in moment $n - 1$, denoted as $\alpha \circ X$, whereby the surviving probability of each element is α ; (ii) the number of new elements entered the system during the interval $(n - 1, n]$, denoted as ε_n . Hence, there is a widely used name for $\{\varepsilon_n\}$ - "innovation sequence".

Due to the importance and role that such defined models have played in the progress of statistical modeling, the list of crucial properties proven by [2] and [46] will be presented. First of all, bearing in mind an assumption that random variables involved in $INAR(1)$ time series are identically distributed, it holds that

$$\begin{aligned} E(X_n) &= \frac{\mu_\varepsilon}{1 - \alpha}, \quad \alpha \neq 1, \\ Var(X_n) &= \frac{\alpha\mu_\varepsilon + \sigma_\varepsilon^2}{1 - \alpha^2}, \quad \alpha \neq 1. \end{aligned}$$

Further, the autocovariance function of $\{X_n\}$ at lag k is

$$\gamma_k \equiv Cov(X_n, X_{n+k}) = \alpha^k \gamma_0, \quad k \geq 0,$$

whereas the autocorrelation function at lag k is of the form

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \alpha^k, \quad k \geq 0.$$

The form of the $INAR(1)$ autocorrelation function is identical to the one that correspond to standard autoregressive $AR(1)$ models, which is the proof of their interconnection. Furthermore, the k -step ahead conditional expectation can be written as

$$E(X_{n+k}|X_n) = \alpha^k X_n + \mu_\varepsilon \left(\frac{1 - \alpha^k}{1 - \alpha} \right), \quad k \geq 1.$$

For k -step ahead conditional variance, it holds

$$Var(X_{n+k}|X_n) = \alpha^k (1 - \alpha^k) X_n + \mu_\varepsilon \frac{\alpha (1 - \alpha^k) (1 - \alpha^{k-1})}{1 - \alpha^2} + \sigma_\varepsilon^2 \frac{1 - \alpha^{2k}}{1 - \alpha^2}.$$

It becomes easy to prove that $E(X_{n+k}|X_n) \rightarrow E(X_n)$ and $Var(X_{n+k}|X_n) \rightarrow Var(X_n)$ when $k \rightarrow \infty$. To summarize, for large number of steps ahead, conditional properties asymptotically approaches to unconditional properties.

Al-Osh and Alzaid gave in [2] one concretization of the $INAR(1)$ models. In Definition 1.2.1, authors specified the distribution of the innovation process $\{\varepsilon_n\}$ to be Poisson with distribution parameter λ , $\lambda > 0$, denoted as $Po(\lambda)$. In this case, it becomes easy to prove that the marginal distribution of the time series $\{X_n\}$ is Poisson as well, but with different distribution parameter $\frac{\lambda}{1 - \alpha}$. The newly acquired time series was named the first-order

Poisson Integer-valued Autoregressive (*PoINAR*(1)) time series. Using formulas given above, properties of the *PoINAR*(1) time series were easily derived. Markov property, strong stationarity and ergodicity were proven as well. In addition, the following statement holds.

Theorem 1.2.1 ([3]). *PoINAR*(1) is the only strongly stationary *INAR*(1) time series with finite variance and linear backward regression.

Few years latter, another important concretization of *INAR*(1) models is given in [1]. Starting from Definition 1.2.1, authors assumed geometric marginal distribution of the sequence $\{X_n\}$, with distribution parameter q , $q \in (0, 1)$, denoted as *Geom*(q). Then, it was proven that

$$\varepsilon_n \stackrel{d}{=} \begin{cases} 0, & \text{w.p. } \alpha \\ G_n, & \text{w.p. } 1 - \alpha, \end{cases}$$

where G_n is a random variable with *Geom*(q) distribution. This time series was called the first-order Geometric Integer-valued Autoregressive (*GINAR*(1)) time series. The notation w.p. represents a shorten form of the term "with probability". One might say that the random variable ε_n is a mixture of the constant random variable 0 and the random variable with corresponding geometric distribution. Thus, ε_n can be written as $\varepsilon_n = I_n G_n$, whereby I_n represents the Bernoulli trail with probability of success $1 - \alpha$. Similar as before, Markov property, strong stationarity and ergodicity of the *GINAR*(1) time series were proven. Other properties were derived as special cases of corresponding *INAR*(1) properties.

Remark. As shown in [35], a useful alternative parametrization can be introduced here. Namely, one can replace parameter q with $\frac{P}{1+P}$, where $P > 0$. The additional value obtained by the introduction of this new parametrization is reflected in more simple recording of many statistical properties. For instance, the new parametrization might be used in unknown parameter estimation, when its positive sides come to the fore.

In order to model more successfully the nonnegative integer-valued autoregressive time series of more complex correlation structure, i.e. the time series with significant dependence between more distant elements, *INAR* models of higher order were introduced. Several approaches appeared, but the most significant between them was the one introduced in [13]. Authors defined the new nonnegative integer-valued autoregressive time series as a generalization of *INAR*(1), introduced by [2]. Their analysis did not refer to any particular distribution, but to the generalized model of order p . Using the binomial thinning operator defined with (1.1), authors introduced the *INAR*(p) time series as

$$(1.3) \quad X_n = \sum_{l=1}^p \alpha_l \circ X_{n-l} + \varepsilon_n,$$

where $\alpha_l \in (0, 1)$, $l = 1, 2, \dots, p$, $\{\varepsilon_n\}$ are i.i.d. random variables with mathematical expectation μ_ε and variance σ_ε^2 , independent of all counting sequences $\{W_{l,i}\}$, such that $P(W_{l,i} = 1) = 1 - P(W_{l,i} = 0) = \alpha_l$. All counting sequences are mutually independent as well. Authors also proved that a stationary solution of the equation (1.3) exists only if the equation

$$\lambda^p - \sum_{l=1}^p \alpha_l \lambda^{p-l} = 0$$

has its roots out of the unit circle. In addition, the weak stationarity is meant by the term "stationary".

1.2.3 Other models based on the counting sequence of Bernoulli trails

After previously presented time series, some significant *INAR* models that have come up by generalization or modification of the thinning operator based on the counting sequences of Bernoulli trails will be mentioned in brief.

To make one such model, a very efficient operator which can be interpreted as a composition of two thinning operators was introduced in [1]. Namely, authors defined operator " $\alpha *$ " as

$$\alpha * X = \sum_{i=1}^{N(X)} U_i,$$

such that:

- $\{U_i\}$ is a sequence of i.i.d. random variables with $Geom\left(\frac{\alpha}{1+\alpha}\right)$ distribution, independent of X and $N(X)$,
- if x is a fixed realization of the random variable X , then $N(x)$ represents the random variable with binomial $\mathcal{B}(x, \alpha p)$ distribution, $0 \leq p \leq 1$.

After that, they introduced a Negative Binomial Iterative Integer-valued Autoregressive (*NBIINAR*(1)) time series $\{X_n\}$ of order 1, based on the new thinning operator, in a following way:

$$X_n = \alpha * X_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables with negative binomial $\mathcal{NB}\left(\nu, \frac{\alpha}{1+\alpha}\right)$ distribution, $\nu > 0$ and ε_n is independent of $\alpha * X_{n-1}$. In addition, if one supposes that X_0 is $\mathcal{NB}\left(\nu, \frac{\alpha(1-p)}{1+\alpha(1-p)}\right)$ distributed, then the newly defined time series has $\mathcal{NB}\left(\nu, \frac{\alpha(1-p)}{1+\alpha(1-p)}\right)$ marginal distribution. This distribution is especially suitable for over-dispersed count data modeling.

Shortly afterwards, the question of modeling time series with much more complex survival mechanism (than the one described by standard binomial thinning operator) suddenly arose. Namely, in many real-life problems the survival probability of population elements varies. For that purpose, a time series with generalized Poisson marginal distributions is defined in [4], whereby survival probability is a linear function of previously counted elements. First, authors defined the quasi-binomial thinning operator " $\rho_{\theta, \lambda} \circ$ " such that $\rho_{\theta, \lambda} \circ X$ has quasi-binomial $\mathcal{QB}\left(\rho, \frac{\theta}{\lambda}, x\right)$ distribution, given $X = x$, $x \in \mathbb{N}_0$, whereby $\lambda > 0$, $\lambda, \rho \in (0, 1)$ and $\frac{x\theta}{\lambda} < \min\{\rho, 1 - \rho\}$. Using aforementioned thinning operator, a Generalized Poisson Quasi-binomial Integer-valued Autoregressive (*GPQINAR*(1)) time series of order 1 was introduced in the following way:

$$X_n = \rho_{\theta, \lambda} \circ X_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where $\lambda > 0$, $\lambda, \rho \in (0, 1)$, $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables with generalized Poisson $GPo(1 - \rho\lambda, \theta)$ distribution and the counting sequence involved in $\rho_{\theta, \lambda} \circ X_m$, as well as the random variable ε_n , are independent of X_m for all $m < n$. As the name itself suggests, marginal distribution is $GPo(\lambda, \theta)$, with probability mass function

$$P(X = x) = \frac{\lambda(\lambda + \theta x)^{x-1} e^{-\lambda - \theta x}}{x!}, \quad x = 0, 1, \dots,$$

where $\lambda > 0$, $\max\{-1, -\frac{\lambda}{m}\} < \theta \leq 1$, $m > 4$.

The problem of modeling time series with inconstant survival probability led to the introduction of the model characterized by the fact that its thinning parameter $\alpha \in [0, 1]$ has been substituted by the random variable ϕ , whose realizations belong to $[0, 1]$. This kind of concept was suggested at first in [50]. Shortly afterwards, a special case with $Beta(\alpha, \beta)$ distributed random variable ϕ was discussed in [49], where the following definition can be found.

DEFINITION 1.2.2 ([49]). *Let X be a nonnegative integer-valued random variable and let ϕ be a random variable with realizations in $[0, 1]$. We say that random variable $\phi \circ X$ is obtained by usage of random coefficient thinning operator, if " \circ " represents a binomial thinning operator independent of ϕ and X .*

Following the results given in [24], author analyzed the case where ϕ has $Beta(\alpha, \beta)$ distribution, and in that case random variable $\phi \circ X$, given $X = x$, has beta-binomial $\mathcal{BB}(x; \alpha, \beta)$ distribution. Then, the definition of the Negative Binomial Random Coefficient Integer-valued Autoregressive ($NBRCINAR(1)$) time series of order 1 appeared in a form:

$$X_n = \phi_n \circ X_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where X_n is $\mathcal{NB}(n, p)$ distributed for all $n \in \mathbb{N}$, $\{\varepsilon_n\}$ represents an innovation sequence of i.i.d. integer-valued random variables with $\mathcal{NB}(n(1 - \rho), p)$ distribution, $\{\phi_n\}$ is a sequence of i.i.d. random variables with $Beta(n\rho, n(1 - \rho))$ distribution, independent of $\{\varepsilon_n\}$, as well as of $\{X_m\}_{m < n}$. It is assumed that n and $n\rho$ are in \mathbb{N}_0 , while $p \in [0, 1]$.

1.2.4 Negative binomial thinning operator

Time series based on the binomial thinning operator are suitable for modeling the data whose elements can contribute to the total sum only with 0 or 1. However, in real-life problems, the observed element can generate several new elements. In that case, counting series of Bernoulli trails are not suitable to describe such data. The need to construct more useful counting series, as well as models based on it, suddenly occurred. A new approach to this problem was suggested in [43]. Authors defined a negative binomial thinning operator " $\alpha *$ " as:

$$(1.4) \quad \alpha * X = \sum_{i=1}^X U_i,$$

where $\alpha \in (0, 1)$ and $\{U_i\}$ represents the sequence of i.i.d. geometric distributed random variables with distribution parameter $\frac{\alpha}{1+\alpha}$. Hence, the random variable $\alpha * X$, given $X = x$, has negative binomial distribution with distribution parameters x and $\frac{\alpha}{1+\alpha}$. Again, bearing in mind that the distribution of the variable $\alpha * X | X$ determines the name of the thinning operator, " $\alpha *$ " is called a negative binomial thinning operator.

Properties of the negative binomial thinning operator

The most important properties of the negative binomial thinning operator defined above will be listed. Namely, let $\alpha, \beta \in (0, 1)$ and let us define operators " $\alpha *$ " and " $\beta *$ " as

$\alpha * X = \sum_{i=1}^X U_i^{(1)}$, $\beta * X = \sum_{i=1}^X U_i^{(2)}$, whereby $\{U_i^{(1)}\}$ and $\{U_i^{(2)}\}$ are assumed to be counting sequences of i.i.d. random variables with $Geom\left(\frac{\alpha}{1+\alpha}\right)$ and $Geom\left(\frac{\beta}{1+\beta}\right)$ distributions, respectively. Now, it holds:

1. $0 * X = 0$;
2. $1 * X \neq X$;
3. $\alpha * (\beta * X) \neq (\alpha\beta) * X$;
4. $E(\alpha * X) = \alpha E(X)$;
5. $E(\alpha * X)^2 = \alpha^2 E(X^2) + \alpha(1 + \alpha)E(X)$;
6. $E((\alpha * X)Y) = \alpha E(XY)$, if the counting sequence involved in $\alpha * X$ is independent of X and Y ;
7. $E(\alpha * X|X) = \alpha X$, if the counting sequence involved in $\alpha * X$ is independent of X ;
8. $E(\alpha * X|X)^2 = \alpha^2 X^2 + \alpha(1 + \alpha)X$, if the counting sequence involved in $\alpha * X$ is independent of X .

Complete list of properties of the negative binomial thinning operator, alongside with corresponding proofs, can be found in [35].

1.2.5 INAR models based on the negative binomial thinning operator

The negative binomial thinning operator was used by [43] to define a new time series with geometric marginal distribution. Geometric distribution is suitable to cover the case of over-dispersion, i.e. the fact that the variance of the count data is considerably larger than the mean. This new time series, referred to as a New Geometric Integer-valued Autoregressive (*NGINAR*(1)) time series of order 1, is defined as follows.

DEFINITION 1.2.3 ([43]). *We say that nonnegative integer-valued autoregressive time series of order 1 with geometric marginal distribution (*NGINAR*(1)) is a sequence $\{X_n\}$ which satisfies the equation*

$$X_n = \alpha * X_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where $\alpha \in (0, 1)$, operator " $\alpha *$ " is defined by (1.4), $\{X_n\}$ is a sequence of random variables with geometric distribution with the distribution parameter $\frac{\mu}{1+\mu}$, $\mu > 0$, $\{\varepsilon_n\}$ represents an innovation sequence of nonnegative integer-valued i.i.d. random variables, such that ε_n is independent of X_{n-k} for all $k > 0$ and independent of counting sequence involved in $\alpha * X_{n-1}$.

Unlike the random variable with Bernoulli distribution, whose realizations may take only 0 or 1, a random variable with geometric distribution is able to generate any nonnegative integer. This fact brought new possibilities to the time series modeling. Namely, the time series defined within Definition 1.2.3 could be used to model not just the phenomenon when the set of elements registered in one moment is only contained of elements survived from the previous observation, but the phenomenon whose elements are able to interact with each other and thus create new elements. Again, the population size can be enlarged through newly arrived elements, which is reflected in the model through innovation sequence.

Regarding the importance of the $NGINAR(1)$ time series, a list of important properties of the model proven in [43] will be presented. Due to the fact that the marginal distribution of the model is $Geom\left(\frac{\mu}{1+\mu}\right)$, $\mu > 0$, and the fact that $\alpha * X_{n-1}$, given $X_{n-1} = x_{n-1}$, has $\mathcal{NB}\left(x_{n-1}, \frac{\alpha}{1+\alpha}\right)$ distribution, it is possible to prove that

$$\varepsilon_n \stackrel{d}{=} \begin{cases} Geom\left(\frac{\alpha}{1+\alpha}\right), & w.p. \frac{\alpha\mu}{\mu-\alpha}, \\ Geom\left(\frac{\mu}{1+\mu}\right), & w.p. 1 - \frac{\alpha\mu}{\mu-\alpha}, \end{cases}$$

i.e. the distribution of the innovation random variable can be represented as a mixture of two geometric distributions with distribution parameters $\frac{\alpha}{1+\alpha}$ and $\frac{\mu}{1+\mu}$. Let us mention here that random variable ε_n is well-defined for $\alpha \in \left(0, \frac{\mu}{1+\mu}\right]$. Next, regarding the assumption that random variables involved in the $NGINAR(1)$ time series are identically distributed, we have

$$\begin{aligned} E(X_n) &= \mu, \\ Var(X_n) &= \mu(1 + \mu). \end{aligned}$$

Further, the autocovariance function of $\{X_n\}$ at lag k is

$$\gamma_k \equiv Cov(X_n, X_{n+k}) = \alpha^k \gamma_0, \quad k \geq 0,$$

whereas the autocorrelation function at lag k is of the form

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \alpha^k, \quad k \geq 0.$$

Furthermore, the k -step ahead conditional expectation is

$$E(X_{n+k}|X_n) = \alpha^k X_n + \frac{1 - \alpha^k}{1 - \alpha} \mu_\varepsilon,$$

where $\mu_\varepsilon = E(\varepsilon_n)$. Particularly, for $k = 1$, the one-step ahead conditional expectation is of the form

$$E(X_{n+1}|X_n) = \alpha X_n + \mu_\varepsilon.$$

Finally, Markov property, property of strong stationarity and property of ergodicity are proven as well.

Soon after that, a time series based on the negative binomial thinning operator, with more general marginal distribution was discussed in [45]. Namely, authors assigned the

negative binomial distribution as a marginal distribution of the model. In that way, they defined a Negative Binomial Integer-valued Autoregressive ($NBINAR(1)$) time series of order 1 as

$$X_n = \alpha * X_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where $\alpha \in (0, 1)$, operator " $\alpha *$ " is defined by (1.4), $\{X_n\}$ is a sequence of random variables with negative binomial $\mathcal{NB}\left(\theta, \frac{\mu}{1+\mu}\right)$ marginal distribution, that is,

$$P(X_n = i) = \frac{\Gamma(\theta + i)}{\Gamma(\theta)i!} \frac{\mu^i}{(1 + \mu)^{\theta+i}}, \quad \theta > 0, \quad \mu > 0, \quad i = 0, 1, 2, \dots,$$

and $\{\varepsilon_n\}$ represents an innovation sequence of nonnegative integer-valued i.i.d. random variables, such that ε_n is independent of X_{n-k} for all $k > 0$ and independent of counting sequence involved in $\alpha * X_{n-1}$.

Bearing in mind the marginal distribution of the model, as well as the distribution of $\alpha * X_{n-1}$, one can determine the distribution of the innovation random variable ε_n . According to [45], $\varepsilon_n \stackrel{d}{=} Y_n + Z_n$, where Y_n and Z_n are mutually independent random variables, the distribution of Y_n is $\mathcal{NB}\left(\theta, \frac{\alpha}{1+\alpha}\right)$ and

$$Z_n = \sum_{l=1}^N \left(\frac{\alpha(1+\mu)}{\mu} \right)^{R_l} \circ V_l,$$

where N has $Po\left(-\theta \ln \frac{\alpha(1+\mu)}{\mu}\right)$ distribution, the distribution of R_l is $\mathcal{U}(0, 1)$, the distribution of V_l is $Geom\left(\frac{\mu}{1+\mu}\right)$, " $\sum_{l=1}^N \left(\frac{\alpha(1+\mu)}{\mu} \right)^{R_l} \circ$ " is a notation for binomial thinning operator and N, R_l and V_l are mutually independent. Again, the random variable ε_n is well-defined for $\alpha \in \left(0, \frac{\mu}{1+\mu}\right]$.

Similar as before, Markov property, property of strong stationarity and property of ergodicity were proven. In addition, for $\theta = 1$, the $NBINAR(1)$ time series is equivalent to the $NGINAR(1)$, introduced in [43].

1.2.6 Mixed $INAR$ models

In practice, sometimes is needed to describe elements whose character varies. These elements are passive in certain moment, which means that they do not interact with other elements of the population. In other moments, observed elements may be very active, which means they may interact with other population elements. In order to describe phenomena made of such elements, a model based on mixture of binomial and negative binomial thinning operator was constructed in [36]. Namely, for $\alpha, \beta \in (0, 1)$, let " $\alpha \circ_n$ " and " $\beta *_n$ " be notations for binomial and negative binomial thinning operator, where index n indicates the moment of operator application. Operators " $\alpha \circ_n$ " and " $\beta *_n$ " are defined by (1.1) and (1.4), respectively. With these notations on the menu, authors introduced a Mixed Geometric Integer-valued Autoregressive ($MGINAR(1)$) time series of order 1 with $Geom\left(\frac{\mu}{1+\mu}\right)$ marginal distribution as

$$X_n = \begin{cases} \alpha \circ_n X_{n-1} + \varepsilon_n, & w.p. \quad p, \\ \beta *_n X_{n-1} + \varepsilon_n, & w.p. \quad 1 - p, \end{cases}$$

where $p \in [0, 1]$ and the following conditions are satisfied:

- (i) $\{\varepsilon_n\}$ is a sequence of nonnegative integer-valued i.i.d. random variables;
- (ii) the random variable ε_n is independent of X_m , $\alpha \circ_{m+1} X_m$ and $\beta *_{m+1} X_m$ for all $m < n$;
- (iii) counting sequences involved in $\alpha \circ_{n+1} X_n$ and $\beta *_{n+1} X_n$ are mutually independent;
- (iv) counting sequences involved in $\alpha \circ_{n+1} X_n$ and $\beta *_{n+1} X_n$ are independent of all X_{n-1}, X_{n-2}, \dots ;
- (v) $P(\alpha \circ_{n+1} X_n = i, \beta *_{n+1} X_n = j | X_n = x, H_{n-1}) = P(\alpha \circ_{n+1} X_n = i, \beta *_{n+1} X_n = j | X_n = x)$, where H_{n-1} represents the process history;
- (vi) given X_n , random variables $\alpha \circ_{n+1} X_n$ and $\beta *_{n+1} X_n$ are mutually independent.

If $\alpha\mu < \beta(1 + \mu) < \mu$, then the distribution of the random variable ε_n is a mixture of three geometric distributions, i.e.

$$(1.5) \quad \varepsilon_n \stackrel{d}{=} \begin{cases} \text{Geom}\left(\frac{\mu}{1+\mu}\right), & w.p. A_1 \equiv \frac{\mu(\alpha-1)(\beta-\mu+\mu\beta)}{(\mu-a)(\mu-b)}, \\ \text{Geom}\left(\frac{a}{1+a}\right), & w.p. A_2 \equiv \frac{(\alpha\mu-a)(\beta-a+\mu\beta)}{(\mu-a)(b-a)}, \\ \text{Geom}\left(\frac{b}{1+b}\right), & w.p. A_3 \equiv \frac{(\alpha\mu-b)(\beta-b+\mu\beta)}{(\mu-b)(a-b)}, \end{cases}$$

where a and b are roots of the equation $x^2 - (\beta + \alpha\mu + \beta\mu p - \alpha\mu p)x + (1 - p)\alpha\beta\mu = 0$ and $a < b$.

The predefined first-order model considers the case when a significant change in the activity of the observed population elements really exists, but the direction of change isn't specified. However, additional information could be available. It might happen that elements start being passive at the beginning of their existence, but become active in the next moment. In order to describe this kind of dependance between elements, the second-order Mixed Geometric Integer-valued Autoregressive (*MGINAR*(2)) time series was defined in [36] in the following way:

$$X_n = \begin{cases} \alpha \circ_n X_{n-1} + \varepsilon_n, & w.p. p, \\ \beta *_{n-1} X_{n-2} + \varepsilon_n, & w.p. 1 - p, \end{cases}$$

where instead of conditions (ii), (v) and (vi), mentioned within the previous model, it holds:

- (ii_a) the random variable ε_n is independent of X_m , $\alpha \circ_{m+1} X_m$ and $\beta *_{m+2} X_m$ for all $m < n$;
- (v_a) $P(\alpha \circ_{n+1} X_n = i, \beta *_{n+2} X_n = j | X_n = x, H_{n-1}) = P(\alpha \circ_{n+1} X_n = i, \beta *_{n+2} X_n = j | X_n = x)$, where H_{n-1} represents history of the process;
- (vi_a) given X_n , random variables $\alpha \circ_{n+1} X_n$ and $\beta *_{n+2} X_n$ are mutually independent.

For $\alpha\mu < \beta(1 + \mu) < \mu$, $\{X_n\}$ is well-defined time series and the distribution of its innovation process is given by (1.5).

At the same time, a generalization of the $MGINAR(2)$ time series was developed in [44], in order to describe a more complex situation when elements start being passive at the beginning of their existence, but become active in the several forthcoming moments. Thus, authors introduced a p -order Mixed Integer-valued Autoregressive ($MINAR(p)$) time series in the following way:

$$X_n = \begin{cases} \alpha \circ_n X_{n-1} + \varepsilon_n, & w.p. \phi_1, \\ \alpha *_n X_{n-2} + \varepsilon_n, & w.p. \phi_2, \\ \alpha *_n X_{n-3} + \varepsilon_n, & w.p. \phi_3, \\ \vdots \\ \alpha *_n X_{n-p} + \varepsilon_n, & w.p. \phi_p, \end{cases}$$

where $\sum_{l=1}^p \phi_l = 1$, $\phi_l \geq 0$, $\alpha \in (0, 1)$, and the following conditions are satisfied:

- a) $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables, such that $E(\varepsilon_n) = \mu_\varepsilon$ and $Var(\varepsilon_n) = \sigma_\varepsilon^2 < \infty$;
- b) the random variable ε_n is independent of X_m , $\alpha \circ_{m+1} X_m$ and $\alpha *_m X_m$, for all $m < n$ and all $l = 2, 3, \dots, p$;
- c) counting sequences involved in thinning operators applied in the moment n are mutually independent;
- d) counting sequences involved in thinning operators applied in the moment n are independent of X_{n-1}, X_{n-2}, \dots

In order to describe the model completely, the distribution of the innovation process was also provided in [44]. Namely, if $0 < \alpha < \frac{\mu}{1+\mu}$, the distribution of ε_n is given as

$$\varepsilon_n \stackrel{d}{=} \begin{cases} Geom\left(\frac{\mu}{1+\mu}\right), & w.p. A_1 \equiv \frac{\mu(\alpha-1)(\alpha-\mu+\mu\alpha)}{(\mu-a)(\mu-b)}, \\ Geom\left(\frac{a}{1+a}\right), & w.p. A_2 \equiv \frac{(\alpha\mu-a)(\alpha-a+\mu\alpha)}{(\mu-a)(b-a)}, \\ Geom\left(\frac{b}{1+b}\right), & w.p. A_3 \equiv \frac{(\alpha\mu-b)(\alpha-b+\mu\alpha)}{(\mu-b)(a-b)}, \end{cases}$$

where a and b are roots of the equation $x^2 - \alpha(1 + \mu)x + \alpha^2\mu(1 - p) = 0$ and $a < b$.

All models listed in this subsection are based on the mixture of binomial and negative binomial thinning operator. In addition, an existence of such defined models is proven. The property of strong stationarity, as well as the property of ergodicity are confirmed for all mixed models listed here.

1.2.7 Models with both positive and negative values

All models mentioned above are limited in the sense that they generate only nonnegative integer values. On the other hand, realizations of the real-life phenomena can frequently take negative values. These phenomena are not suitable to be described with any of predefined models. This is why the researchers tried to create new models which overcome newly arisen problem.

A step forward in this direction was made by [15]. Namely, author defined a time series with symmetric Skellam marginal distribution using results presented in [16]. First of all, author introduced a new thinning operator in the following way:

$$\alpha \star Z_n | Z_n \stackrel{d}{=} (\alpha \circ X_n - \alpha \circ Y_n) | (X_n - Y_n),$$

where $\alpha \in (0, 1)$, " $\alpha \circ$ " is defined by (1.1) and X_n, Y_n are random variables with the same $Po\left(\frac{\mu}{1-\alpha}\right)$ distributions. Counting sequences involved in $\alpha \circ X_n$ and $\alpha \circ Y_n$ are mutually independent and independent of X_n, Y_n and Z_n . After that, author defined the first-order True Integer-valued Autoregressive ($TINAR(1)$) time series as

$$Z_n = \alpha \star Z_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

with symmetric $Skellam\left(\frac{\mu}{1-\alpha}\right)$ marginal distribution, where $\{\varepsilon_n\}$ represents a sequence of i.i.d. integer-valued random variables with symmetric $Skellam(\mu)$ distribution, such that ε_n and Z_{n-k} are independent for all $k > 0$. Since the $TINAR(1)$ represents the topic of Chapter 2, properties of this model will be omitted here.

Another interesting model was created by [8]. Authors used the signed thinning operator, already defined in [31] as

$$F \circ X = \begin{cases} \operatorname{sgn}(X) \sum_{i=1}^{|X|} Y_i, & X \neq 0 \\ 0, & X = 0, \end{cases}$$

where elements of the counting sequence, i.e. $Y_i, i > 0$, are independent of X and are i.i.d. random variables with the distribution characterized by the probability mass function F . With such defined operator on the table, the first-order Signed Integer-valued Autoregressive ($SINAR(1)$) time series was introduced by [8] in the following way:

$$X_n = F \circ X_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where $\{\varepsilon_n\}$ represents an innovation sequence of i.i.d. random variables with mathematical expectation μ_ε and the finite variance σ_ε^2 , independent of the counting sequence involved in " $F \circ$ ", such that ε_n and X_{n-k} are independent for all $k > 0$. Given the distribution of the innovation sequence, authors find the way to determine the marginal distribution of such defined time series $\{X_n\}$.

In order to successfully model the real-life data with both positive and negative values, one more interesting idea was introduced in [5]. First, authors defined so called "expanded" thinning operator as

$$S_{\alpha, \theta}(Z) = \operatorname{sgn}(Z) \sum_{i=1}^{|Z|} Y_i + \sum_{i=1}^{W(Z)} B_i,$$

where $\alpha \in (0, 1)$, $\theta \geq 0$, $\{Y_i\}$ is a sequence of i.i.d. Bernoulli trials with probability of success α , independent of the sequence $\{B_i\}$ and random variables Z and $W(Z)$. Also, $\{B_i\}$ is a sequence of i.i.d. random variables distributed as

$$B_i : \begin{pmatrix} -1 & 0 & 1 \\ \alpha(1-\alpha) & 1-2\alpha(1-\alpha) & \alpha(1-\alpha) \end{pmatrix},$$

independent of Z and $W(Z)$. Finally, the random variable $W(Z)$ satisfies condition that $W(Z)|Z = z$ has Bessel distribution with distribution parameters $|z|$ and θ . With such defined thinning operator, authors came into position to introduce the first-order Poisson Difference Integer-Valued Autoregressive ($PDINAR(1)$) time series in the following way:

$$Z_n = \delta S_{\alpha, \theta}(Z_{n-1}) + \varepsilon_n, \quad n \in \mathbb{N},$$

where parameter δ takes values from the set $\{-1, 1\}$, $\{\varepsilon_n\}$ is a sequence of i.i.d. random variables with skewed $Skellam(\theta_1, \theta_2)$ distribution, such that ε_n is independent of Z_{n-k} for all $k > 0$.

A significant step forward happened in 2016. when a time series with discrete Laplace marginal distribution was defined by [37]. Namely, authors constructed the time series basing it on the new thinning operator, defined as

$$(1.6) \quad \alpha \odot Z_n | Z_n \stackrel{d}{=} (\alpha * X_n - \alpha * Y_n) | (X_n - Y_n),$$

where $\alpha \in (0, 1)$, " $\alpha *$ " is defined by (1.4) and X_n and Y_n are random variables with the same $Geom\left(\frac{\mu}{1+\mu}\right)$ distribution. Using this newly defined thinning operator, authors constructed a new time series, referred to as the first-order Discrete Laplace Integer-valued Autoregressive ($DLINAR(1)$) time series, in the following way:

$$(1.7) \quad Z_n = \alpha \odot Z_{n-1} + e_n, \quad n \in \mathbb{N},$$

where $\{Z_n\}$ represents a time series with discrete Laplace distribution, while $\{e_n\}$ is an innovation sequence of i.i.d. random variables, such that e_n and Z_{n-k} are mutually independent for all $k > 0$. Further, authors proved that $\{Z_n\}$ is well-defined time series for $0 < \alpha \leq \frac{\mu}{1+\mu}$. Furthermore, it holds that

$$e_n \stackrel{d}{=} \begin{cases} DL\left(\frac{\mu}{1+\mu}\right), & w.p. \left(1 - \frac{\alpha\mu}{\mu-\alpha}\right)^2, \\ SDL\left(\frac{\mu}{1+\mu}, \frac{\alpha}{1+\alpha}\right), & w.p. \frac{\alpha\mu}{\mu-\alpha} \left(1 - \frac{\alpha\mu}{\mu-\alpha}\right), \\ SDL\left(\frac{\alpha}{1+\alpha}, \frac{\mu}{1+\mu}\right), & w.p. \frac{\alpha\mu}{\mu-\alpha} \left(1 - \frac{\alpha\mu}{\mu-\alpha}\right), \\ DL\left(\frac{\alpha}{1+\alpha}\right), & w.p. \left(\frac{\alpha\mu}{\mu-\alpha}\right)^2, \end{cases}$$

where " SDL " represents the notation of the skewed discrete Laplace distribution. In addition, Markov property, property of strong stationarity, property of ergodicity and positive correlation between elements of the sequence $\{Z_n\}$ are proven. Besides this, the definition of the time series with negatively correlated elements is given as well.

Several generalizations of this concept appeared. For instance, authors discussed in [6] the case when marginal distribution of the sequence $\{Z_n\}$ is skewed discrete Laplace, that is, the case when marginal distributions of $NGINAR(1)$ time series involved in the definition of $\{Z_n\}$ are geometric with different distribution parameters. However, thinning parameters that appear beside X_n and Y_n remained the same again.

Even more generalized case was introduced by [11]. Namely, author discussed the case when marginal distribution of the time series $\{Z_n\}$ is again skewed discrete Laplace,

but thinning parameters appearing beside X_n and Y_n are different. This new model is referred to as the first-order Skewed Discrete Laplace Integer-valued Autoregressive (*SDLINAR*(1)) time series. The most of the properties confirmed for *DLINAR*(1) have also been proven in the case of *SDLINAR*(1).

More complex real-life problems required appropriate mathematical model with more complex correlation structure, that is, time series with a significant dependence between more distant elements. As a response to this kind of problems, researches introduced *INAR*(p) models with values in \mathbb{Z} . First of all, an integer-valued autoregressive time series of higher order was defined in [27] using the newly introduced signed binomial thinning operator. This operator is defined in the following way:

$$\alpha \odot X = \text{sgn}(\alpha)\text{sgn}(X) \sum_{i=1}^{|X|} W_i,$$

where elements of the counting sequence are i.i.d. Bernoulli trials with distribution parameter $|\alpha|$, $\alpha \in [-1, 1]$. Also, elements of the counting sequence are independent of X . With such defined thinning operator, authors introduced in [27] a non-stationary integer-valued autoregressive model of order p as

$$X_n = \sum_{l=1}^p \alpha_l \odot X_{n-l} + \epsilon_n, \quad n \geq p,$$

where $\alpha_l \in (-1, 1)$, $l = 1, 2, \dots, p$ and $\{\epsilon_n\}$ is an innovation sequence of i.i.d. random variables with finite mean and finite variance, such that ϵ_n is not correlated with X_{n-k} , $k \geq 1$. Elements of the counting sequences involved in $\alpha_l \odot X_{n-l}$, $l = 1, 2, \dots, p$, are mutually independent and independent of X_n .

Another interesting approach was given in [25], where authors managed to define an integer-valued autoregressive time series using the signed binomial thinning operator, given in [31]. Time series, referred to as a p -order Signed Integer-valued Autoregressive (*SINAR*(p)) time series, was defined as

$$X_n = \sum_{l=1}^p F_l \circ X_{n-l} + \xi_n, \quad n \geq p,$$

where " $F_l \circ$ " represents the signed binomial thinning operator given in [31] and $\{\xi_n\}$ is an innovation sequence of i.i.d. random variables with finite expectation and finite variance. Elements of the innovation sequence $\{\xi_n\}$ are independent of elements of counting sequences. Also, counting sequences involved in $F_l \circ X_{n-l}$, $l = 1, 2, \dots, p$, are mutually independent.

1.2.8 Random environment process

The most of aforementioned models are stationary, which means that they are suitable to describe the phenomena with approximately constant characteristics over time. However, the real-life data often deviates from this assumption and shows some non-stationary

properties. In this case, aforementioned models did not give satisfactory results. Obviously, the non-stationary *INAR* models had to be introduced somehow. It was desirable to simplify these models in the way to become as similar as possible to the stationary *INAR* models. One of the ways to do so, given in [38], is to define a concept of time series in random environment. The essence of the concept is to introduce an auxiliary process, i.e. a random environment process, in order to allow the primary time series to have a different behavior in each environment.

Each phenomenon, observed by researchers, takes place under conditions determined by the environment, which directly affects the registered values of the observed phenomenon. It sounds reasonable to suppose that any change in those conditions may lead to a change in registered values. Each state of environment conditions allows primary time series to have one particular distribution, that is, to accumulate realizations in the range around one specific value. The number of states, denoted as r , may be enlarged or reduced, depending on the analyzed problem. However, it is recommended not to enlarge the number of states r too much, despite the fact that a higher number of states increases the flexibility of the model. Yet, a higher number of states entail a higher number of model parameters, which complicates working with such model. A formal definition of the random environment process follows.

DEFINITION 1.2.4 ([38], Definition 1). *A sequence of random variables $\{Z_n\}$, $n \in \mathbb{N}_0$, is called the r states random environment process, for $r \in \mathbb{N}$, if it is a Markov chain with values in $E_r = \{1, 2, \dots, r\}$. More generally, $\{Z_n\}$, $n \in \mathbb{N}_0$, is the random environment process, if it is the r states random environment process, for some $r \in \mathbb{N}$.*

One can notice here that the set E_r is consisted of the first r natural numbers. These numbers do not represent any observed values, i.e. they do not represent any measurable characteristic of the environment. There is no experiment nor observation which measure the environment itself. The goal of the research is not the random environment modeling, but the influence of the random environment on the primary process being modeled. Bearing that in mind, it becomes clear that the properties of the environment are not of importance and that its only role is to provide the introduction of non-stationarity into a particular *INAR* model. But, one may wonder how to determine the values of the random environment process, if those are not obtained by experiment. Well, the procedure is quite easy. By observing given realizations, it is possible to determine the number of different sets where realizations are grouped in, so in accordance with that, one may assume that there is an equal number of different environment states. If r states are registered, then one may claim that the random environment process is taking values in $E_r = \{1, 2, \dots, r\}$. The notation $Z_n = s$ means that the environment is in the s -th state, and in accordance with that, the s -th distribution from the family of distributions $\{F(x, \theta_q), q \in E_r\}$ will be associated to the *INAR* model in moment n . To sum up, the values found in the set E_r are not of importance, but the fact that there are r different values, and that each value is associated with exactly one state and each state with exactly one distribution of the observed process.

One important property of the random environment process needs to be highlighted. As given in [38], let $z_n = z_{n+1} = \dots = z_{n+k} = s$, $k \in \mathbb{N}$, $s \in E_r$, where z_i , $i = n, n+1, \dots, n+k$ represent realizations of the random environment process. In addition, let $z_{n-1} \neq z_n$ and $z_{n+k} \neq z_{n+k+1}$. Elements of the subsequence $X_n, X_{n+1}, \dots, X_{n+k}$

all correspond to the state s , while adjacent elements correspond to some other states. Hence, this subsequence might be observed as a stationary piece of the given time series, regarding the absence of the state change. Thus, random environment *INAR* models might be observed as a piecewise stationary. Consequently, some properties of the stationary *INAR* time series may be applied piecewise. Given conclusion brings a great simplification to the analysis of random environment *INAR* models. It should be mentioned that the lengths of stationary pieces are determined by the transition probability matrix of the r states random environment process. To clarify, the transition probability matrix reveals the probabilities of transitioning from one state to another in two consecutive observations. More precisely, for all $q, s \in E_r$, transition probability p_{qs} from state q to state s is defined as $p_{qs} = P(Z_{n+1} = s | Z_n = q)$. Now, higher values on the main diagonal of the transition probability matrix mean that the environment state will more likely remain the same for a while, which implies longer pieces correspond to the same state.

In order to apply *INAR* models based on the random environment process, the sequence $\{z_n\}$ of realizations of the random environment process needs to be determined first. For that purpose, clustering is used. After determining the number of clusters r , observed realizations are supposed to be grouped into r different clusters and each cluster is observed as one specific state. If the realization x_n is in the s -th cluster, one may consider that $z_n = s$. In that way, each realization is associated with one particular state. This completely determines the sequence $\{z_n\}$. Let's say that the sample of size N is registered and some predictions need to be made. Before predicting the value of X_{N+1} , one has to find the way to predict Z_{N+1} , given realizations z_1, z_2, \dots, z_N . In order to do so, an estimate of the transition probability matrix is going to be used. Namely, transition probabilities can be estimated in a usual way, as a quotient of the number of favorable outcomes and the total number of outcomes. For instance, as an estimate of the transition probability from state q to state s one can take a quotient of the number of all transitions from q to s registered in the sample and the total number of transitions from state q . Using estimated transition probability matrix thus obtained, for given $Z_N = z_N$, one can generate the prediction value of Z_{N+1} , and then the value of Z_{N+2} based on $Z_{N+1} = z_{N+1}$ etc. In this way, the sequence $\{z_n\}$ of an arbitrary length can be created.

Described interconnection between the *INAR* time series and the random environment process might be generalized. So far, environment states had influence only on the distribution of the observed model. Nevertheless, one can expand the field of influence of the environment states and suppose their influence on the order of the model, or even on the value of the corresponding thinning parameter. Generalization might be also carried out in terms of the time series dimension. For instance, a bivariate random environment process would be useful to involve the non-stationarity into the bivariate *INAR* time series, where two different environments exist, such that each of them affects one component of the time series. The following definition introduces the bivariate random environment process.

DEFINITION 1.2.5 ([30]). *A sequence of bivariate random variables $\{(W_n, Q_n)\}$, $n \in \mathbb{N}_0$, indexed by the set of nonnegative integer numbers \mathbb{N}_0 is called a bivariate (r_1, r_2) states random environment process, for r_1, r_2 from the set of positive integer numbers \mathbb{N} , if $\{W_n\}$ and $\{Q_n\}$, $n \in \mathbb{N}_0$ are the r_1 and the r_2 states random environment processes, respectively. More generally, $\{(W_n, Q_n)\}$, $n \in \mathbb{N}_0$, is bivariate random environment process if it is*

bivariate (r_1, r_2) states random environment process, for some $r_1, r_2 \in \mathbb{N}$.

1.2.9 INAR models based on the random environment process

An idea of the random environment process, presented in the preceding text, served to develop *INAR* models in a completely new direction. The first *INAR* model based on the random environment process was introduced in [38]. The main assumption made by authors is that the environment conditions have an influence on the marginal distribution of the *INAR* time series. More precisely, the registered value z_n of the random environment process in moment n determines the marginal distribution parameter μ_{z_n} in the same moment from the supposed set of marginal parameter values $\mathcal{M} = \{\mu_1, \mu_2, \dots, \mu_r\}$, $r \in \mathbb{N}$. Under the additional assumption that the marginal distribution is geometric, the r states random environment *INAR* time series of order 1 (*RrNGINAR*(1)) was defined in [38] as

$$X_n(z_n) = \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), \quad n \in \mathbb{N},$$

where $\alpha \in (0, 1)$, " $\alpha *$ " is defined by (1.4), $\{z_n\}$ is a realization of the process $\{Z_n\}$ and $\{\varepsilon_n(q, s)\}$, $n \in \mathbb{N}$, $q, s \in E_r$, are sequences of i.i.d. random variables, such that:

- (1) $\{Z_n\}, \{\varepsilon_n(1, 1)\}, \{\varepsilon_n(1, 2)\}, \dots, \{\varepsilon_n(r, r)\}$ are mutually independent sequences of random variables;
- (2) $X_n(u)$ is independent of Z_m and $\varepsilon_m(q, s)$ for all $n < m$ and all $q, s, u \in E_r$.

The notation $X_n(z_n)$ is here to clarify the fact that the distribution of X_n depends on z_n , that is,

$$P(X_n(z_n) = x) = \frac{\mu_{z_n}^x}{(1 + \mu_{z_n})^{x+1}}, \quad x \in \mathbb{N}_0,$$

where $\mu_{z_n} \in \mathcal{M}$. Several properties have been proven, among them the distribution of the innovation process. If $0 \leq \alpha \leq \min \left\{ \frac{\mu_j}{1 + \mu_i}, i, j \in E_r \right\}$, then for fixed $z_{n-1} = q$ and $z_n = s$, $q, s \in E_r$,

$$(1.8) \quad \varepsilon_n(q, s) \stackrel{d}{=} \begin{cases} \text{Geom} \left(\frac{\mu_s}{1 + \mu_s} \right), & w.p. \ 1 - \frac{\alpha \mu_q}{\mu_s - \alpha}, \\ \text{Geom} \left(\frac{\alpha}{1 + \alpha} \right), & w.p. \ \frac{\alpha \mu_q}{\mu_s - \alpha}. \end{cases}$$

This model served as a prototype of introducing the non-stationarity into *INAR* models. Several generalizations appeared in forthcoming years, but all of them were based on the same foundations. One of those generalizations was given in [39], where authors modified assumptions a bit. Namely, it was assumed that environment conditions, beside the marginal distribution parameter, have an effect on the order of the time series, i.e. the realization z_n of the random environment process in moment n also determines the order of the model in the same moment. More precisely, authors defined two similar models, depending on the order's growth within a subsequence of consecutive elements corresponding to the same state. To explain more appropriately, let us introduce here the set $\mathcal{P} = \{p_1, p_2, \dots, p_r\}$ of maximal orders for all states.

In the first case, authors assumed that the element $X_n(z_n)$ in moment n is allowed to take the largest possible order p_n , not greater than maximal order p_{z_n} , provided

$X_{n-1}(z_{n-1}), X_{n-2}(z_{n-2}), \dots, X_{n-p_n}(z_{n-p_n})$ belong to the same state. To realize that, authors introduced in [39] an *INAR* time series with r states random environment guided geometric marginal distributions ($RrNGINAR_{max}(p)$) in the following way:

$$X_n(z_n) = \begin{cases} \alpha * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & w.p. \phi_1^{(p_n)}, \\ \alpha * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & w.p. \phi_2^{(p_n)}, \\ \vdots \\ \alpha * X_{n-p_n}(z_{n-p_n}) + \varepsilon_n(z_{n-p_n}, z_n), & w.p. \phi_{p_n}^{(p_n)}, \end{cases}$$

where $p_n = \min\{p_{z_n}, p_n^*\}$, $p_n^* = \max\{k \in \{1, 2, \dots, n\} : z_{n-1} = z_{n-2} = \dots = z_{n-k}\}$, and the following conditions are satisfied:

- (1) $\phi_l^{(p_n)} \geq 0$, $l \in \{1, 2, \dots, p_n\}$, $\sum_{l=1}^{p_n} \phi_l^{(p_n)} = 1$;
- (2) $\alpha \in (0, 1)$ and " $\alpha *$ " is defined by (1.4);
- (3) for fixed $q, s \in E_r$, $\{\varepsilon_n(q, s)\}$ is a sequence of i.i.d. random variables;
- (4) $\{Z_n\}, \{\varepsilon_n(1, 1)\}, \{\varepsilon_n(1, 2)\}, \dots, \{\varepsilon_n(r, r)\}$ are mutually independent sequences of random variables;
- (5) $X_n(u)$ is independent of Z_m and $\varepsilon_m(q, s)$ for all $n < m$ and all $q, s, u \in E_r$.

Behavior of the model needs to be highlighted additionally. When the state change occurs, the order of the defined time series takes value 1. Afterwards, the order value starts increasing by 1 in each subsequent moment, until it reaches its maximum value predicted for that state p_{z_n} . Then, it remains at maximum until the next state change.

In the second case, assumptions were modified. Namely, after the state change occurs, the order becomes 1 again, but it doesn't gradually increase over time. After taking value 1, the order of the model remains the same until the conditions to take maximum value p_{z_n} are fulfilled. Precisely, the only two possible order values corresponding to the particular state s are 1 and p_s . Bearing in mind everything mentioned here, authors introduced an *INAR* time series with r states random environment guided geometric marginal distributions ($RrNGINAR_1(p)$) in the same way as it was done with $RrNGINAR_{max}(p)$, except the definition of p_n , which was in the case of $RrNGINAR_1(p)$ time series defined as

$$p_n = \begin{cases} p_{z_n}, & p_n^* \geq p_{z_n}, \\ 1, & p_n^* < p_{z_n}. \end{cases}$$

Again, p_n^* is defined as $p_n^* = \max\{k \in \{1, 2, \dots, n\} : z_{n-1} = z_{n-2} = \dots = z_{n-k}\}$, and conditions (1)-(5) still hold. For both predefined models, the distribution of the innovation process is equivalent to the one of the $RrNGINAR(1)$ model. Provided $0 \leq \alpha \leq \min\left\{\frac{\mu_j}{1+\mu_i}, i, j \in E_r\right\}$, the distribution of the innovation process is given by (1.8) for fixed $z_{n-1} = q$ and $z_n = s$, $q, s \in E_r$.

Even more generalized models were defined by [29]. Beside the marginal distribution parameter and the order of the model, authors supposed that environment conditions have an effect on the thinning operator value as well. Precisely, the registered value z_n of the random environment process in moment n determines the thinning parameter value

α_{z_n} in the same moment from the supposed set of parameter values $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$. Depending on the behavior of the model order after the state change, authors defined two models with r states random environment guided geometric marginal distributions. First of them, referred to as $RrNGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$, was defined as

$$(1.9) \quad X_n(z_n) = \begin{cases} \alpha_{z_n} * X_{n-1}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), & w.p. \phi_{1, P_n}^{(z_n)}, \\ \alpha_{z_n} * X_{n-2}(z_{n-2}) + \varepsilon_n(z_{n-2}, z_n), & w.p. \phi_{2, P_n}^{(z_n)}, \\ \vdots \\ \alpha_{z_n} * X_{n-P_n}(z_{n-P_n}) + \varepsilon_n(z_{n-P_n}, z_n), & w.p. \phi_{P_n, P_n}^{(z_n)}, \end{cases}$$

where $P_n = \min\{p_{z_n}, p_n^*\}$, $p_n^* = \max\{k \in \{1, 2, \dots, n\} : z_{n-1} = z_{n-2} = \dots = z_{n-k}\}$, and conditions (1)-(5), set within the previous model, are satisfied. The second model, referred to as $RrNGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$, was defined in an almost identical manner. An exception is the definition of P_n , which was given as

$$P_n = \begin{cases} p_{z_n}, & p_n^* \geq p_{z_n}, \\ 1, & p_n^* < p_{z_n}. \end{cases}$$

Finally, let one supposes that $z_{n-1} = q$ and $z_n = s$ for some $q, s \in E_r$. If $0 \leq \alpha_s \leq \frac{\mu_s}{1 + \max_{q \in E_r} \mu_q}$, then the distribution of the random variable $\varepsilon_n(q, s)$ can be written as a mixture of two geometrically distributed random variables with means μ_s and α_s , as follows:

$$\varepsilon_n(q, s) \stackrel{d}{=} \begin{cases} \text{Geom}\left(\frac{\mu_s}{1 + \mu_s}\right), & w.p. 1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}, \\ \text{Geom}\left(\frac{\alpha_s}{1 + \alpha_s}\right), & w.p. \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}. \end{cases}$$

This conclusion holds for both, $RrNGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$ and $RrNGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models.

Regarding the generalization in terms of the model dimension, one idea stands out in importance. The idea was carried out in [30], where the first-order bivariate $INAR(1)$ time series with geometric marginal distributions, based on bivariate (r_1, r_2) states random environment process, was defined. Components of the bivariate $INAR(1)$ time series are related through the bivariate random environment process in the way that they depend on both of the random environment components, so it becomes impossible to separate them completely into two $RrNGINAR(1)$ time series and analyze them separately. This is why the model is referred to as a Crossed Bivariate $INAR(1)$ time series with (r_1, r_2) states random environment guided geometric marginal distributions, or abbreviated $CBRNGINAR(1)$.

1.3 Important distributions

In this section, two important discrete distributions are discussed and their crucial properties are presented. These distributions, alongside with their properties, will be widely used in the following chapters.

1.3.1 Skellam distribution

Even though Skellam distribution got its name by John Skellam¹, it was introduced first in its symmetric form by [20]. Current name was given a decade later. One says that the random variable Z has a symmetric Skellam distribution with distribution parameter μ , $\mu \geq 0$, if its probability mass function is of the form

$$(1.10) \quad p(k; \mu) = P(Z = k) = e^{-2\mu} I_{|k|}(2\mu), \quad k \in \mathbb{Z},$$

where $I_{|k|}$ is the modified Bessel function of the first kind, given as

$$(1.11) \quad I_{|k|}(x) = \sum_{l=0}^{\infty} \frac{\left(\frac{x}{2}\right)^{2l+|k|}}{l! \Gamma(l + |k| + 1)}.$$

The fact that the random variable Z has Skellam distribution with distribution parameter μ will be denoted as $Z : Skellam(\mu)$. One interesting fact in regard to the symmetric Skellam distribution is given in [20], and represents a connection between symmetric Skellam distribution and Poisson distribution. Namely, it is shown that the random variable Z with symmetric Skellam distribution can be represented in distribution as a difference between two i.i.d. random variables with identically parameterized Poisson distributions.

A generalization of the symmetric Skellam distribution was introduced in [47], where a skewed Skellam distribution with distribution parameters μ and ν , $\mu, \nu \geq 0$, was defined and denoted as $Skellam(\mu, \nu)$. The corresponding probability mass function is of the form:

$$(1.12) \quad p(k; \mu, \nu) = P(Z = k) = e^{-(\mu+\nu)} \left(\frac{\mu}{\nu}\right)^{\frac{k}{2}} I_{|k|}(2\sqrt{\mu\nu}), \quad k \in \mathbb{Z},$$

where Z represents the skewed Skellam distributed random variable with distribution parameters μ and ν , and $I_{|k|}$ is given by (1.11). Analogously to the symmetric case, the random variable Z can be represented in distribution as a difference between two independent random variables X and Y with differently parameterized Poisson distributions and distribution parameters μ and ν respectively, i.e.

$$Z \stackrel{d}{=} X - Y.$$

Bearing in mind the shape of the probability mass function of the $Skellam(\mu, \nu)$ distributed random variable given by (1.12), a formula for moment-generating function (MGF) of the mentioned random variable might be easily obtained. The MGF of the random variable Z with $Skellam(\mu, \nu)$ distribution is of the form:

$$(1.13) \quad M_Z(s) = e^{-(\mu+\nu)+\mu e^s + \nu e^{-s}}.$$

Using properties of the MGF, it becomes easy to determine raw moments of the random variable Z :

$$\begin{aligned} m_1 &= E(Z) = \mu - \nu, \\ m_2 &= E(Z^2) = \mu + \nu + (\mu - \nu)^2, \\ m_3 &= E(Z^3) = (\mu - \nu) (1 + 3(\mu + \nu) + (\mu - \nu)^2). \end{aligned}$$

Consequently, $Var(Z) = m_2 - m_1^2 = \mu + \nu$. Corresponding results for symmetric Skellam distribution are easy to obtain by equalizing parameters μ and ν .

¹John Gordon Skellam (1914 – 1979)

1.3.2 Discrete Laplace distribution

In this subsection, a discrete Laplace distribution will be presented and some of its properties will be revealed. Discrete Laplace distribution was introduced by [19] as a discrete equivalent of the Laplace distribution of the continuous type. The random variable Z is said to be discrete Laplace distributed with distribution parameter $p \in (0, 1)$, if its probability mass function is of the form

$$P(Z = z) = \frac{1-p}{1+p} p^{|z|}, \quad z \in \mathbb{Z}.$$

The fact that the random variable Z has discrete Laplace distribution with distribution parameter p will be denoted as $Z : DL(p)$. Since a generalization of the discrete Laplace distribution occurred later, this form of distribution sometimes have a prefix "symmetric". In [19], a connection between discrete Laplace distribution and geometric distribution was also revealed. Similar as it was the case with symmetric Skellam distribution, authors showed that the random variable Z with discrete Laplace distribution can be represented in distribution as a difference between two i.i.d. geometrically distributed random variables with the same distribution parameter p . Regarding this fact, a tiny change in notation of the distribution parameter will be implemented. Namely, the discrete Laplace distribution will be denoted as $DL\left(\frac{\mu}{1+\mu}\right)$, $\mu > 0$, where μ represents mathematical expectation of the aforementioned random variables with geometric distributions. In that case, one would say the random variable Z is discrete Laplace distributed if its probability mass function is defined as

$$(1.14) \quad P(Z = z) = \frac{1}{1+2\mu} \left(\frac{\mu}{1+\mu}\right)^{|z|}, \quad z \in \mathbb{Z}, \quad \mu > 0.$$

A skewed discrete Laplace distribution with parameters $\mu > 0$ and $\nu > 0$ was introduced in [28] as a generalization of the (symmetric) discrete Laplace distribution. The new distribution was denoted as $SDL\left(\frac{\mu}{1+\mu}, \frac{\nu}{1+\nu}\right)$. The probability mass function of such introduced distribution is given as

$$(1.15) \quad P(Z = z) = \begin{cases} \frac{1}{1+\mu+\nu} \left(\frac{\mu}{1+\mu}\right)^z, & z \geq 0, \\ \frac{1}{1+\mu+\nu} \left(\frac{\nu}{1+\nu}\right)^{-z}, & z < 0. \end{cases}$$

Similar as in the symmetric case, a random variable with $SDL\left(\frac{\mu}{1+\mu}, \frac{\nu}{1+\nu}\right)$ distribution can be represented in distribution as a difference between two independent random variables with $Geom\left(\frac{\mu}{1+\mu}\right)$ and $Geom\left(\frac{\nu}{1+\nu}\right)$ distributions, respectively. Bearing in mind the probability mass function of the random variable with skewed discrete Laplace distribution, it becomes easy to obtain its characteristic function, which is of the form

$$(1.16) \quad \varphi_Z(t) = \frac{1}{(1+\mu - \mu e^{it})(1+\nu - \nu e^{-it})}.$$

Using the well known properties of the characteristic function, some numerical characteristics of the random variable Z with $SDL\left(\frac{\mu}{1+\mu}, \frac{\nu}{1+\nu}\right)$ distribution are easily obtained, that is,

$$E(Z) = \mu - \nu, \quad Var(Z) = \mu(1+\mu) + \nu(1+\nu).$$

Also it would be convenient to present the expression for $E(|Z|)$, which will be useful for some later calculations:

$$(1.17) \quad E(|Z|) = \frac{\mu(1 + \mu) + \nu(1 + \nu)}{1 + \mu + \nu}.$$

And again, as it was the case with Skellam distribution, corresponding results for (symmetric) discrete Laplace distribution are easy to obtain by equalizing parameters μ and ν .

1.4 Important theorems

Several important theorems, crucial for proving numerous properties of the time series with discrete Laplace marginal distributions, will be presented in this section. These theorems represent results of various authors, but the most of them have been already mentioned in the text given above.

Theorem 1.4.1 ([37], Corollary 2.1). *For thinning operator " $\alpha \odot$ ", defined by (1.6), it holds:*

- (a) $\alpha \odot Z_{n-1} \stackrel{d}{=} \alpha * X_{n-1} - \alpha * Y_{n-1}$;
- (b) $E(\alpha \odot Z_{n-1}) = 0$;
- (c) $Var(\alpha \odot Z_{n-1}) = 2\alpha\mu(1 + 2\alpha + \alpha\mu)$;
- (d) $0 \odot Z_{n-1} \stackrel{d}{=} 0$;
- (e) $1 \odot Z_{n-1} \stackrel{d}{\neq} Z_{n-1}$.

Theorem 1.4.2 ([37], Theorem 2.3). *The conditional expectation and conditional variance of the random variable $\alpha \odot Z_{n-1}$ for given Z_{n-1} are respectively given as*

$$\begin{aligned} E(\alpha \odot Z_{n-1} | Z_{n-1}) &= \alpha Z_{n-1}, \\ Var(\alpha \odot Z_{n-1} | Z_{n-1}) &= \alpha(1 + \alpha)|Z_{n-1}| + \frac{2\alpha(1 + \alpha)\mu^2}{1 + 2\mu}. \end{aligned}$$

Theorem 1.4.3 ([37], Theorem 2.4). *Let Z, X and Y be random variables with $DL\left(\frac{\mu}{1+\mu}\right)$, $Geom\left(\frac{\mu}{1+\mu}\right)$ and $Geom\left(\frac{\mu}{1+\mu}\right)$ distributions, respectively. Let $\{D_l, l \geq 1\}$ be a sequence of independent random variables with $DL\left(\frac{\alpha}{1+\alpha}\right)$ distributions and suppose that random variables $Z, X, Y, D_l, l \geq 1$, and the random variables involved in $\alpha * |Z|$ are independent. Then,*

$$(1.18) \quad \alpha \odot Z \stackrel{d}{=} \text{sgn}(Z)(\alpha * |Z|) + \sum_{l=1}^{\min\{X, Y\}} D_l,$$

where $\sum_{l=1}^{\min\{X, Y\}} D_l = 0$ when $\min\{X, Y\} = 0$ and sgn is the sign function.

Theorem 1.4.4 ([37], Theorem 3.3). *DLINAR(1) time series $\{Z_n\}$ given by (1.7) is positively correlated time series with autocorrelation function at lag k given as $\rho_k = \text{Corr}(Z_n, Z_{n-k}) = \alpha^k$, $k \geq 0$.*

Theorem 1.4.5 ([33], Continuous Mapping Theorem). *Let $\{\mathbf{X}_n\}$ be a sequence of k -dimensional random variables and let $g : \mathbb{R}^k \rightarrow \mathbb{R}^l$ be a continuous mapping. Then,*

$$\mathbf{X}_n \rightarrow \mathbf{X} \Rightarrow g(\mathbf{X}_n) \rightarrow g(\mathbf{X}), \quad n \rightarrow +\infty,$$

where the statement holds for convergence in distribution, convergence in probability, and almost sure convergence.

Theorem 1.4.6 ([13], Theorem 4.1). *Let $\{X_n\}$ be the INAR(p) time series, $p \geq 1$ and let $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_p)$ be the vector of thinning parameters of the model, such that*

$$\boldsymbol{\Gamma} \boldsymbol{\alpha} = \boldsymbol{\rho},$$

where $\boldsymbol{\Gamma} = [\rho_{|i-j|}]_{p \times p}$, $\boldsymbol{\rho} = (\rho_1, \rho_2, \dots, \rho_p)'$ and $\rho_k = \text{Corr}(X_n, X_{n-k})$, $k = 1, 2, \dots, p$. Then,

$$\hat{\boldsymbol{\alpha}}^{YW} = \hat{\boldsymbol{\Gamma}}^{-1} \hat{\boldsymbol{\rho}}$$

is a strongly consistent estimator of the thinning parameter vector $\boldsymbol{\alpha}$.

Chapter 2

Extracting and predicting latent components of the skewed $TINAR(1)$ time series

As mentioned in Chapter 1, $TINAR(1)$ time series described by [15] was regarded as a great step forward within the theory of integer-valued autoregressive time series. Following the results given in [16] and [15], it's not complicated to conclude that time series $\{Z_n\}$ with symmetric or skewed Skellam marginal distribution can be defined in distribution as a difference between two independent nonnegative time series that affect the values of $\{Z_n\}$ in opposite directions, i.e. $Z_n = X_n - Y_n$, $n \geq 0$. These two hidden time series with Poisson marginal distributions, named latent components, are in focus of the chapter. Namely, a very reasonable question arises. Is it possible to identify latent components when realizations of the time series $\{Z_n\}$ are familiar?

The following text discusses this issue in the case of $TINAR(1)$ time series with skewed Skellam marginal distribution. The chapter represents a unique combination of well known facts introduced by several researchers and recent achievements given in [12]. Bearing in mind the fact that $TINAR(1)$ underlies the mentioned issue, the first part of the chapter is dedicated to the symmetric $TINAR(1)$ time series, introduced by [15]. In addition, singularities of the skewed $TINAR(1)$ time series are presented as well. The second part of the chapter analyzes the possibilities of identifying and predicting latent components $\{X_n\}$ and $\{Y_n\}$, depending on realizations of the skewed $TINAR(1)$ time series $\{z_n\}$. Results are based on the expression for calculating mathematical expectation of the random variable X_{n+k} , given Z_n . Expressions for latent components extraction and prediction will be the functions of the following arguments: the realization sequence $\{z_n\}$ of the skewed $TINAR(1)$ and Yule-Walker YW estimates of its marginal distribution parameters. All YW estimates will be counted upon the random sample Z_1, Z_2, \dots, Z_N of size N . Except expressions for extracting and predicting latent components, the fitting quality of such obtained results will be assessed. Also, the application of this type of modeling on real-life data will be presented.

2.1 Symmetric $TINAR(1)$ time series

This section is dedicated to the construction of the symmetric $TINAR(1)$ time series, which is built on the sample space of integers including positive and negative values.

Marginal distribution is symmetric Skellam, whose properties have already been described in Section 1.3. The time series is built using innovations that come from the symmetric distribution, which gives many advantages including the possibility of negative correlation modeling. It should be emphasized that the results given within this section are mostly taken from [15]. In addition, some calculations are performed in the original way.

2.1.1 Construction of the model

The construction of the model is begun with definition of the new thinning operator. The thinning operator is defined in quite interesting way by using the equality in distribution rather than the classical equality, which was usually the case. But first of all, let Z_n be a random variable with symmetric *Skellam* $(\frac{\mu}{1-\alpha})$ distribution. Bearing in mind the properties of the aforementioned distribution, random variable Z_n can be represented in distribution as a difference between two i.i.d. random variables with *Po* $(\frac{\mu}{1-\alpha})$ distributions. Thus, let $Z_n \stackrel{d}{=} X_n - Y_n$, where X_n and Y_n represent two latent Poisson random variables. It is obvious that the random variable Z_n defined in this way can take either positive or negative integer values, depending on the interrelationship of its latent components. The higher values of the component X_n lead to the positive values of Z_n , and vice versa, the higher values of the component Y_n lead to the negative values of Z_n . Furthermore, let " $\alpha \circ$ " be a denotement of the binomial thinning operator, defined with (1.1). Bearing that in mind, the thinning operator " $\alpha \star$ " can be defined in a following way:

$$(2.1) \quad (\alpha \star Z_n) | Z_n \stackrel{d}{=} (\alpha \circ X_n - \alpha \circ Y_n) | (X_n - Y_n),$$

for $\alpha \in (0, 1)$. Counting series involved in $\alpha \circ X_n$ and $\alpha \circ Y_n$ are mutually independent and independent of random variables X_n , Y_n and Z_n .

Now, one can focus attention on calculating conditional probabilities of the random variable $\alpha \star Z_n$, given Z_n , i.e. on calculating transition probabilities $P(\alpha \star Z_n = w | Z_n = z)$. Transition probabilities can be used for clarifying some properties of the operator " $\alpha \star$ " and the *TINAR*(1) time series itself. First of all, regarding the distribution of the latent variables X_n and Y_n , one can calculate that

$$(2.2) \quad \begin{aligned} P(X_n = z + y, Y_n = y) &= P(X_n = z + y) \cdot P(Y_n = y) = \frac{e^{-\frac{\mu}{1-\alpha}} \left(\frac{\mu}{1-\alpha}\right)^{z+y}}{(z+y)!} \cdot \frac{e^{-\frac{\mu}{1-\alpha}} \left(\frac{\mu}{1-\alpha}\right)^y}{y!} \\ &= \frac{e^{-2\frac{\mu}{1-\alpha}} \left(\frac{\mu}{1-\alpha}\right)^{z+2y}}{(z+y)!y!}, \end{aligned}$$

where $x! = 0$ for $x \leq 0$. The second, conditioning on latent components leads to the following:

$$(2.3) \quad \begin{aligned} P(\alpha \star Z_n = w | X_n = z + y, Y_n = y) &= P(\alpha \circ X_n - \alpha \circ Y_n = w | X_n = z + y, Y_n = y) \\ &= \sum_{l=0}^y \binom{y}{l} \alpha^l (1-\alpha)^{y-l} \times \\ &\times \binom{z+y}{w+l} \alpha^{w+l} (1-\alpha)^{z+y-w-l} \\ &= \sum_{l=0}^y \binom{y}{l} \binom{z+y}{w+l} \alpha^{w+2l} (1-\alpha)^{z+2y-w-2l}. \end{aligned}$$

Without any loss of generality, let $w \in \mathbb{Z}$ and $z > 0$. Finally, by combining the results obtained in expressions (2.2) and (2.3), the following form of the transition probability holds:

$$\begin{aligned}
P(\alpha \star Z_n = w | Z_n = z) &= P(\alpha \circ X_n - \alpha \circ Y_n = w | X_n - Y_n = z) \\
&= \sum_{y=0}^{\infty} P(\alpha \circ X_n - \alpha \circ Y_n = w | X_n = z + y, Y_n = y) \times \\
&\quad \times P(X_n = z + y, Y_n = y) \\
&= \sum_{y=0}^{\infty} \frac{e^{-2\frac{\mu}{1-\alpha}} \left(\frac{\mu}{1-\alpha}\right)^{z+2y}}{(z+y)!y!} \times \\
&\quad \times \sum_{l=0}^y \binom{y}{l} \binom{z+y}{w+l} \alpha^{w+2l} (1-\alpha)^{z+2y-w-2l} \\
&= \sum_{y=0}^{\infty} \frac{e^{-2\frac{\mu}{1-\alpha}} \mu^{z+2y}}{(z+y)!y!} \sum_{l=0}^y \binom{y}{l} \binom{z+y}{w+l} \left(\frac{\alpha}{1-\alpha}\right)^{w+2l}.
\end{aligned}$$

Given $z < 0$, it is easy to see that, because of symmetry, $P(\alpha \star Z_n = w | Z_n = z) = P(\alpha \star Z_n = -w | Z_n = -z)$. Although one might find the expression for calculating transition probabilities way complicated, some properties of the following process cannot be clarified without direct interference of this conditional probability function.

It is a proper occasion now to introduce a stationary autoregressive time series with symmetric *Skellam* $\left(\frac{\mu}{1-\alpha}\right)$ marginal distribution.

DEFINITION 2.1.1. Let $\{Z_n\}$, $n \geq 0$ be a time series with symmetric *Skellam* $\left(\frac{\mu}{1-\alpha}\right)$ marginal distribution, defined as

$$(2.4) \quad Z_n = \alpha \star Z_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where Z_0 represents a random variable with symmetric *Skellam* $\left(\frac{\mu}{1-\alpha}\right)$ distribution, $\{\varepsilon_n, n \geq 1\}$ represents a sequence of i.i.d. integer-valued random variables with symmetric *Skellam* (μ) distributions, such that ε_n and Z_{n-l} are independent for all $l \geq 1$ and the thinning operator " $\alpha \star$ " is defined by (2.1). From now on, this time series will be denoted as a symmetric True Integer-valued Autoregressive time series of order 1, or abbreviated, symmetric *TINAR*(1).

2.1.2 Properties of the model

Many properties of the symmetric *TINAR*(1) time series will be discussed bellow. At the beginning, it would be very helpful to represent *TINAR*(1) in distribution as a difference between two independent Poisson *INAR*(1) time series. For that purpose, the distribution of $\alpha \star Z_{n-1}$, $n \geq 1$ must be determined first. In accordance with (1.13), the *MGF* of the random variable Z with *Skellam* (μ) distribution is of the form

$$M_Z(s) = e^{\mu(e^s + e^{-s} - 2)}.$$

For that reason and due to the fact that ε_n is independent of all elements of the counting series involved in $\alpha \star Z_{n-1}$, one has

$$M_{\alpha \star Z_{n-1}}(s) = \frac{M_{Z_n}(s)}{M_{\varepsilon_n}(s)} = \frac{e^{\frac{\mu}{1-\alpha}(e^s + e^{-s} - 2)}}{e^{\mu(e^s + e^{-s} - 2)}} = e^{\left(\frac{\mu}{1-\alpha} - \mu\right)(e^s + e^{-s} - 2)} = e^{\frac{\alpha\mu}{1-\alpha}(e^s + e^{-s} - 2)},$$

which represent the *MGF* of the random variable with *Skellam* $(\frac{\alpha\mu}{1-\alpha})$ distribution, i.e. $\alpha \star Z_{n-1}$ is *Skellam* $(\frac{\alpha\mu}{1-\alpha})$ distributed. Now, let

$$\begin{aligned} X_n &= \alpha \circ X_{n-1} + \xi_n, \quad n \in \mathbb{N}, \\ Y_n &= \alpha \circ Y_{n-1} + \eta_n, \quad n \in \mathbb{N} \end{aligned}$$

be two independent Poisson *INAR*(1) time series with the same *Po* $(\frac{\mu}{1-\alpha})$ marginal distributions. Thus, $\{\xi_n\}$ and $\{\eta_n\}$ are mutually independent processes with the same *Po* (μ) distributions. The distribution of $\alpha \circ X_{n-1}$ can be obtained in a similar way as it was done with the distribution of $\alpha \star Z_{n-1}$. Namely, bearing in mind the form of *MGF* of the random variable X with *Po* (μ) distribution,

$$M_X(s) = e^{\mu(e^s-1)},$$

it holds that

$$M_{\alpha \circ X_{n-1}}(s) = \frac{M_{X_n}(s)}{M_{\xi_n}(s)} = \frac{e^{\frac{\mu}{1-\alpha}(e^s-1)}}{e^{\mu(e^s-1)}} = e^{(\frac{\mu}{1-\alpha}-\mu)(e^s-1)} = e^{\frac{\alpha\mu}{1-\alpha}(e^s-1)}.$$

To be precise, this means that $\alpha \circ X_{n-1}$ has Poisson distribution with distribution parameter $\frac{\alpha\mu}{1-\alpha}$. The same can be proven for $\alpha \circ Y_{n-1}$. Then, the distribution of $\alpha \circ X_{n-1} - \alpha \circ Y_{n-1}$ is *Skellam* $(\frac{\alpha\mu}{1-\alpha})$, i.e. $\alpha \circ X_{n-1} - \alpha \circ Y_{n-1} \stackrel{d}{=} \alpha \star Z_{n-1}$. On the other hand, it is obvious that $\xi_n - \eta_n \stackrel{d}{=} \varepsilon_n$. To sum up,

$$X_n - Y_n = (\alpha \circ X_{n-1} - \alpha \circ Y_{n-1}) + (\xi_n - \eta_n) \stackrel{d}{=} \alpha \star Z_{n-1} + \varepsilon_n = Z_n.$$

Hence, $E(Z_n) = E(X_n) - E(Y_n) = 0$ and $Var(Z_n) = Var(X_n) + Var(Y_n) = 2Var(X_n) = 2\frac{\mu}{1-\alpha}$. Beside this, $E(\varepsilon_n) = 0$ and $Var(\varepsilon_n) = 2\mu$.

To reveal more properties of the symmetric *TINAR*(1) time series, results given in [2] and [13] are going to be used. Namely, authors gave in [2] two conditions that imply stationarity of the *INAR*(1) time series: $E(X_n) = \frac{\lambda}{1-\alpha}$ and $Var(X_n) = \frac{\alpha\lambda + \sigma^2}{1-\alpha^2}$. Here, λ and σ^2 represent mathematical expectation and variance of the innovation process. Moreover, if the innovation process has Poisson distribution, then one may claim the *INAR*(1) time series is strongly stationary. Poisson *INAR*(1) time series satisfies all of these conditions. The first and third condition are evidently satisfied. To prove the satisfaction of the second condition, one must take into account the fact that $E(\xi_n) = Var(\xi_n) = \mu$, where $\{\xi_n\}$ represents the innovation process of the Poisson *INAR*(1) time series. Thus,

$$Var(X_n) = \frac{\mu}{1-\alpha} = \frac{\mu(1+\alpha)}{1-\alpha^2} = \frac{\alpha\mu + \mu}{1-\alpha^2},$$

which proves that Poisson *INAR*(1) time series satisfies the second condition as well. Regarding everything mentioned above, one may claim Poisson *INAR*(1) is the strongly stationary time series, that is, latent components $\{X_n\}$ and $\{Y_n\}$ are strongly stationary. This means that k -dimensional vectors $(X_{n_1}, X_{n_2}, \dots, X_{n_k})$ and $(X_{n_1+h}, X_{n_2+h}, \dots, X_{n_k+h})$, as well as vectors $(Y_{n_1}, Y_{n_2}, \dots, Y_{n_k})$ and $(Y_{n_1+h}, Y_{n_2+h}, \dots, Y_{n_k+h})$, have the same multivariate joint distributions for all $k, n_1, n_2, \dots, n_k \in \mathbb{N}$ and for all $h \in \mathbb{Z}$. Hence,

$$\begin{aligned} (Z_{n_1+h}, Z_{n_2+h}, \dots, Z_{n_k+h}) &= (X_{n_1+h} - Y_{n_1+h}, X_{n_2+h} - Y_{n_2+h}, \dots, X_{n_k+h} - Y_{n_k+h}) \\ &= (X_{n_1+h}, X_{n_2+h}, \dots, X_{n_k+h}) - (Y_{n_1+h}, Y_{n_2+h}, \dots, Y_{n_k+h}) \\ &\stackrel{d}{=} (X_{n_1}, X_{n_2}, \dots, X_{n_k}) - (Y_{n_1}, Y_{n_2}, \dots, Y_{n_k}) \\ &= (Z_{n_1}, Z_{n_2}, \dots, Z_{n_k}), \end{aligned}$$

which proves the strong stationarity of $\{Z_n\}$. Further, the property of ergodicity for $INAR(p)$ time series is proven in [13] under certain conditions. For $\alpha \in (0, 1)$, all $INAR(1)$ time series satisfy these conditions, including Poisson $INAR(1)$. Since the symmetric $TINAR(1)$ time series represents a difference between two independent Poisson $INAR(1)$ time series, it must be ergodic as well.

In addition, the following properties of the $INAR(1)$ time series $\{X_n\}$ appear in [2]:

$$\begin{aligned} X_n &\stackrel{d}{=} \sum_{l=0}^{\infty} \alpha \circ^{(l)} \xi_{n-l}, \\ (X_n, X_{n-k}) &\stackrel{d}{=} \left(\alpha \circ^{(k)} X_{n-k} + \sum_{l=0}^{k-1} \alpha \circ^{(l)} \xi_{n-l}, X_{n-k} \right), \\ E(X_n) &= \alpha E(X_{n-1}) + E(\xi_n) = \alpha^n E(X_0) + E(\xi_n) \sum_{l=0}^{n-1} \alpha^l, \\ Var(X_n) &= \alpha^{2n} Var(X_0) + (1 - \alpha) \sum_{l=1}^n \alpha^{2l-1} E(X_{n-l}) + \sigma^2 \sum_{l=1}^n \alpha^{2(l-1)}, \\ \gamma_k &= Cov(X_{n-k}, X_n) = \alpha^k Var(X_{n-k}) = \alpha^k \gamma_0, \end{aligned}$$

where $\sigma^2 = Var(\xi_n)$ and " $\alpha \circ^{(l)}$ " is defined as $\alpha \circ^{(0)}(X) \stackrel{def}{=} X$, $\alpha \circ^{(l)} X \stackrel{def}{=} \alpha \circ (\alpha \circ^{(l-1)} X)$, i.e. the notation " $\alpha \circ^{(l)}$ " represents l consecutive applications of the operator " $\alpha \circ$ ". In accordance with properties mentioned above, corresponding formulas for symmetric $TINAR(1)$ time series can be obtained.

Theorem 2.1.1. *For symmetric $TINAR(1)$ time series $\{Z_n\}$ it holds:*

- i) $Z_n \stackrel{d}{=} \sum_{l=0}^{\infty} \alpha \star^{(l)} \varepsilon_{n-l}$;
- ii) $(Z_n, Z_{n-k}) \stackrel{d}{=} \left(\alpha \star^{(k)} Z_{n-k} + \sum_{l=0}^{k-1} \alpha \star^{(l)} \varepsilon_{n-l}, Z_{n-k} \right)$;
- iii) $E(Z_n) = \alpha^n E(Z_0)$;
- iv) $Var(Z_n) = \alpha^{2n} Var(Z_0) + (1 - \alpha) \sum_{l=1}^n \alpha^{2l-1} (E(X_{n-l}) + E(Y_{n-l})) + 2\sigma^2 \sum_{l=1}^n \alpha^{2(l-1)}$;
- v) $\gamma_k = \alpha^k \gamma_0$.

Proof.

i) First,

$$Z_n \stackrel{d}{=} X_n - Y_n \stackrel{d}{=} \sum_{l=0}^{\infty} \alpha \circ^{(l)} \xi_{n-l} - \sum_{l=0}^{\infty} \alpha \circ^{(l)} \eta_{n-l} = \sum_{l=0}^{\infty} (\alpha \circ^{(l)} \xi_{n-l} - \alpha \circ^{(l)} \eta_{n-l}).$$

It has been already mentioned that the distribution of $\alpha \circ X_n$ is $Po\left(\frac{\alpha\mu}{1-\alpha}\right)$ when X_n is $Po\left(\frac{\mu}{1-\alpha}\right)$ distributed. In general, if random variable X has Poisson distribution with distribution parameter p , it is easy to show that $\alpha \circ X$ also has Poisson distribution,

but this time with distribution parameter αp . Regarding this fact, the distribution of $\alpha \circ \xi_{n-l}$ is $Po(\alpha\mu)$. Bearing in mind that

$$\alpha \circ^{(l)} \xi_{n-l} = \underbrace{\alpha \circ (\alpha \circ (\dots \alpha \circ (\xi_{n-l}) \dots))}_{l \text{ times}},$$

the distribution of $\alpha \circ^{(l)} \xi_{n-l}$ is Poisson with distribution parameter $\alpha^l \mu$. The same can be shown for $\alpha \circ^{(l)} \eta_{n-l}$. Consequently, the distribution of $\alpha \circ^{(l)} \xi_{n-l} - \alpha \circ^{(l)} \eta_{n-l}$ is *Skellam* ($\alpha^l \mu$). On the other hand, the distribution of $\alpha \star \varepsilon_{n-l}$ is Skellam, with distribution parameter $\alpha \mu$. In the same way as it was done in previous lines, it can be shown that $\alpha \star^{(l)} \varepsilon_{n-l}$ has *Skellam* ($\alpha^l \mu$) distribution. Thus, $\alpha \circ^{(l)} \xi_{n-l} - \alpha \circ^{(l)} \eta_{n-l} \stackrel{d}{=} \alpha \star^{(l)} \varepsilon_{n-l}$. Summa summarum,

$$Z_n \stackrel{d}{=} \sum_{l=0}^{\infty} (\alpha \circ^{(l)} \xi_{n-l} - \alpha \circ^{(l)} \eta_{n-l}) \stackrel{d}{=} \sum_{l=0}^{\infty} \alpha \star^{(l)} \varepsilon_{n-l}.$$

ii)

$$\begin{aligned} (Z_n, Z_{n-k}) &\stackrel{d}{=} (X_n - Y_n, X_{n-k} - Y_{n-k}) \\ &= (X_n, X_{n-k}) - (Y_n, Y_{n-k}) \\ &\stackrel{d}{=} \left(\alpha \circ^{(k)} X_{n-k} + \sum_{l=0}^{k-1} \alpha \circ^{(l)} \xi_{n-l}, X_{n-k} \right) \\ &\quad - \left(\alpha \circ^{(k)} Y_{n-k} + \sum_{l=0}^{k-1} \alpha \circ^{(l)} \eta_{n-l}, Y_{n-k} \right) \\ &= \left((\alpha \circ^{(k)} X_{n-k} - \alpha \circ^{(k)} Y_{n-k}) \right. \\ &\quad \left. + \sum_{l=0}^{k-1} (\alpha \circ^{(l)} \xi_{n-l} - \alpha \circ^{(l)} \eta_{n-l}), X_{n-k} - Y_{n-k} \right). \end{aligned}$$

For the same reason as it was the case in (i), $\alpha \circ^{(k)} X_{n-k}$ and $\alpha \circ^{(k)} Y_{n-k}$ have the same Poisson distributions with distribution parameter $\alpha^k \frac{\mu}{1-\alpha}$. Hence, $\alpha \circ^{(k)} X_{n-k} - \alpha \circ^{(k)} Y_{n-k}$ has *Skellam* ($\alpha^k \frac{\mu}{1-\alpha}$) distribution, which is the distribution of $\alpha \star^{(k)} Z_{n-k}$, i.e. $\alpha \circ^{(k)} X_{n-k} - \alpha \circ^{(k)} Y_{n-k} \stackrel{d}{=} \alpha \star^{(k)} Z_{n-k}$. Same as in (i), $\alpha \circ^{(l)} \xi_{n-l} - \alpha \circ^{(l)} \eta_{n-l} \stackrel{d}{=} \alpha \star^{(l)} \varepsilon_{n-l}$. With the well known fact that $X_{n-k} - Y_{n-k} \stackrel{d}{=} Z_{n-k}$, it finally holds that

$$(Z_n, Z_{n-k}) \stackrel{d}{=} \left(\alpha \star^{(k)} Z_{n-k} + \sum_{l=0}^{k-1} \alpha \star^{(l)} \varepsilon_{n-l}, Z_{n-k} \right).$$

iii) Regarding the characteristics of the binomial thinning operator, it holds that

$$\begin{aligned} E(Z_n) &= E(X_n) - E(Y_n) = \alpha E(X_{n-1}) + \mu - \alpha E(Y_{n-1}) - \mu \\ &= \alpha E(X_{n-1} - Y_{n-1}) = \alpha E(Z_{n-1}) = \dots = \alpha^n E(Z_0). \end{aligned}$$

iv) Here,

$$\begin{aligned}
\text{Var}(Z_n) &= \text{Var}(X_n) + \text{Var}(Y_n) \\
&= \alpha^{2n} \text{Var}(X_0) + (1 - \alpha) \sum_{l=1}^n \alpha^{2l-1} E(X_{n-l}) + \sigma^2 \sum_{l=1}^n \alpha^{2(l-1)} \\
&+ \alpha^{2n} \text{Var}(Y_0) + (1 - \alpha) \sum_{l=1}^n \alpha^{2l-1} E(Y_{n-l}) + \sigma^2 \sum_{l=1}^n \alpha^{2(l-1)} \\
&= \alpha^{2n} (\text{Var}(X_0) + \text{Var}(Y_0)) + (1 - \alpha) \sum_{l=1}^n \alpha^{2l-1} (E(X_{n-l}) + E(Y_{n-l})) \\
&+ 2\sigma^2 \sum_{l=1}^n \alpha^{2(l-1)} \\
&= \alpha^{2n} \text{Var}(Z_0) + (1 - \alpha) \sum_{l=1}^n \alpha^{2l-1} (E(X_{n-l}) + E(Y_{n-l})) \\
&+ 2\sigma^2 \sum_{l=1}^n \alpha^{2(l-1)}.
\end{aligned}$$

v) Finally, from the independence of X_i and Y_j , $i, j = 0, 1, 2, \dots, n$, it holds

$$\begin{aligned}
\gamma_k &= \text{Cov}(Z_{n-k}, Z_n) \\
&= \text{Cov}(X_{n-k} - Y_{n-k}, X_n - Y_n) \\
&= E((X_{n-k} - Y_{n-k}) \cdot (X_n - Y_n)) - E(X_{n-k} - Y_{n-k})E(X_n - Y_n) \\
&= E(X_{n-k}X_n) - E(X_{n-k})E(X_n) + E(Y_{n-k}Y_n) - E(Y_{n-k})E(Y_n) \\
&= \text{Cov}(X_{n-k}, X_n) + \text{Cov}(Y_{n-k}, Y_n) \\
&= \alpha^k (\text{Var}(X_{n-k}) + \text{Var}(Y_{n-k})) \\
&= \alpha^k \gamma_0. \quad \square
\end{aligned}$$

□

In particular, $\gamma_k = 2\alpha^k \frac{\mu}{1-\alpha}$. Also, one can easily calculate the following:

$$\text{Corr}(Z_{n-k}, Z_n) = \frac{\text{Cov}(Z_{n-k}, Z_n)}{\sqrt{\text{Var}(Z_{n-k})\text{Var}(Z_n)}} = \frac{\gamma_k}{\gamma_0} = \alpha^k.$$

One useful property, necessary for future deriving, will be also proven here. Namely, from $\gamma_k = \alpha^k \gamma_0$, it holds that

$$\alpha \gamma_0 = \gamma_1 = \text{Cov}(Z_{n-1}, Z_n) = \text{Cov}(Z_{n-1}, \alpha \star Z_{n-1} + \varepsilon_n).$$

Since random variables Z_{n-1} and ε_n are independent, the following can be obtained:

$$(2.5) \quad \text{Cov}(Z_{n-1}, \alpha \star Z_{n-1}) = \alpha \gamma_0 = \alpha \text{Cov}(Z_{n-1}, Z_{n-1}).$$

2.1.3 Symmetric $TINAR(1)$ model with negative lag-one autocorrelation

It is possible to define a stationary symmetric $TINAR(1)$ time series with negative lag-one autocorrelation in the following way:

$$(2.6) \quad Z_n^{(1)} = \alpha \star \left(-Z_{n-1}^{(1)} \right) + \zeta_n,$$

where $Z_0^{(1)}$ is a random variable with *Skellam* $\left(\frac{\mu}{1-\alpha}\right)$ distribution, $\{\zeta_n\}$ represents a sequence of i.i.d. random variables such that ζ_n and $Z_{n-l}^{(1)}$ are mutually independent for all $l \geq 1$ and $\zeta_n \stackrel{d}{=} (-1)^n(\xi_n - \eta_n)$. Thinning operator " $\alpha \star$ " is, as usual, defined by (2.1).

Obviously, if $Z_{n-1}^{(1)} \stackrel{d}{=} X_{n-1} - Y_{n-1}$ and $\zeta_{n-1} \stackrel{d}{=} \xi_{n-1} - \eta_{n-1}$, then $\zeta_n \stackrel{d}{=} \eta_n - \xi_n$ and

$$\begin{aligned} Z_n^{(1)} &= \alpha \star \left(-Z_{n-1}^{(1)}\right) + \zeta_n \\ &\stackrel{d}{=} \alpha \circ Y_{n-1} - \alpha \circ X_{n-1} + \eta_n - \xi_n \\ &= Y_n - X_n. \end{aligned}$$

This proves the fact that the marginal distribution of the newly defined time series is the same as the marginal distribution of the time series introduced by (2.4). Also, differences between latent time series are alternating, i.e.

$$Z_n^{(1)} \stackrel{d}{=} \begin{cases} X_n - Y_n, & n = 0, 2, 4, \dots, \\ Y_n - X_n, & n = 1, 3, 5, \dots \end{cases}$$

For time series defined by (2.6), the same or similar properties might be calculated as it was the case with time series defined by (2.4). The first, $E\left(Z_n^{(1)}\right) = 0$ and $Var\left(Z_n^{(1)}\right) = Var(Z_n) = 2\frac{\mu}{1-\alpha}$. Further,

$$\begin{aligned} E\left(Z_n^{(1)}\right) &= E(X_n) - E(Y_n) = \alpha(E(X_{n-1}) - E(Y_{n-1})) = -\alpha E\left(Z_{n-1}^{(1)}\right) \\ &= \dots = (-\alpha)^n E\left(Z_0^{(1)}\right). \end{aligned}$$

In addition,

$$\gamma_1 = Cov\left(Z_{n-1}^{(1)}, Z_n^{(1)}\right) = Cov\left(Z_{n-1}^{(1)}, \alpha \star \left(-Z_{n-1}^{(1)}\right) + \zeta_n\right).$$

From independence of ζ_n and $Z_{n-1}^{(1)}$ and equality (2.5), it holds

$$\begin{aligned} \gamma_1 &= Cov\left(Z_{n-1}^{(1)}, \alpha \star \left(-Z_{n-1}^{(1)}\right)\right) = \alpha Cov\left(Z_{n-1}^{(1)}, -Z_{n-1}^{(1)}\right) \\ &= -\alpha Cov\left(Z_{n-1}^{(1)}, Z_{n-1}^{(1)}\right) = -\alpha \gamma_0. \end{aligned}$$

Finally, $Corr\left(Z_{n-1}^{(1)}, Z_n^{(1)}\right) = \frac{\gamma_1}{\gamma_0} = -\alpha < 0$, which proves the assumption of negative lag-one autocorrelation.

2.1.4 Parameter estimation of the symmetric *TINAR*(1) model

In order to estimate unknown parameters μ and α of the symmetric *TINAR*(1) model, the *YW* estimation method is going to be used. For that purpose, let Z_1, Z_2, \dots, Z_N be a sample of size N . Since the equality $\gamma_k = \alpha^k \gamma_0$ had been already confirmed, it becomes easy to obtain the *YW* estimator of the parameter α :

$$\hat{\alpha}^{YW} = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{\sum_{n=1}^{N-1} Z_n Z_{n+1}}{\sum_{n=1}^N Z_n^2}.$$

Bearing in mind the formula for marginal variance, $\gamma_0 = \text{Var}(Z_n) = 2\frac{\mu}{1-\alpha}$, it is trivial to obtain the estimator of the parameter μ :

$$\hat{\mu}^{YW} = \frac{1 - \hat{\alpha}^{YW}}{2} \hat{\gamma}_0 = \frac{1 - \hat{\alpha}^{YW}}{2N} \sum_{n=1}^N Z_n^2.$$

The asymptotic behavior of these estimators is not in the focus here and won't be analyzed. To reveal their asymptotic distribution, see [15].

2.2 Skewed *TINAR*(1) time series

In this section, a skewed *TINAR*(1) time series will be analyzed. Results given within the section represent a generalization of the results from [15]. This applies in particular to the first two subsections. The last subsection relies on results given in [12].

As well as the symmetric *TINAR*(1), the skewed *TINAR*(1) time series may consist of both positive and negative integer values. The main difference comes from the marginal distribution of the time series and the distribution of the innovation process. Also, the thinning operator had to be adapted a bit, which is about to be seen in the text bellow. The newly introduced time series represent a generalization of the time series introduced in the previous section.

2.2.1 Construction of the model

Again, the construction of the model will start with introduction of the new thinning operator. For that purpose, let Z_n be a random variable with skewed *Skellam* $\left(\frac{\mu}{1-\alpha}, \frac{\nu}{1-\beta}\right)$ distribution. According to the distribution properties given in Section 1.3, this random variable can be represented in distribution as a difference between two random variables X_n and Y_n with $Po\left(\frac{\mu}{1-\alpha}\right)$ and $Po\left(\frac{\nu}{1-\beta}\right)$ distribution respectively, i.e. $Z_n \stackrel{d}{=} X_n - Y_n$. Depending on the interrelationship between X_n and Y_n , the random variable Z_n can take both positive and negative values. Let " $\alpha \circ$ " be a binomial thinning operator, introduced by (1.1). A new thinning operator will be defined in the following way:

$$(2.7) \quad ((\alpha, \beta) \star Z_n) \stackrel{d}{=} (\alpha \circ X_n - \beta \circ Y_n) | (X_n - Y_n),$$

where $\alpha \in (0, 1)$, counting series involved in $\alpha \circ X_n$ and $\beta \circ Y_n$ are mutually independent and independent of random variables X_n, Y_n and Z_n .

Calculating the transition probabilities $P((\alpha, \beta) \star Z_n = w | Z_n = z)$ might be again carried out through the three-step procedure. Namely, same as it was the case with $P(\alpha \star Z_n = w | Z_n = z)$, one first has that

$$(2.8) \quad \begin{aligned} P(X_n = z + y, Y_n = y) &= e^{-\frac{\mu}{1-\alpha}} \frac{\left(\frac{\mu}{1-\alpha}\right)^{z+y}}{(z+y)!} e^{-\frac{\nu}{1-\beta}} \frac{\left(\frac{\nu}{1-\beta}\right)^y}{y!} \\ &= e^{-\frac{\mu}{1-\alpha} - \frac{\nu}{1-\beta}} \frac{\left(\frac{\mu}{1-\alpha}\right)^{z+y} \left(\frac{\nu}{1-\beta}\right)^y}{(z+y)! y!}. \end{aligned}$$

The second, conditioning on latent components, it holds

$$\begin{aligned}
P((\alpha, \beta) \star Z_n = w | X_n = z + y, Y_n = y) &= P(\alpha \circ X_n - \beta \circ Y_n = w | X_n = z + y, Y_n = y) \\
(2.9) \qquad \qquad \qquad &= \sum_{l=0}^y \binom{y}{l} \beta^l (1 - \beta)^{y-l} \cdot \binom{z+y}{w+l} \alpha^{w+l} (1 - \alpha)^{z+y-w-l}.
\end{aligned}$$

By combining results obtained in (2.8) and (2.9), one gets the following form of the transition probability for all $z > 0$ and for all $w \in \mathbb{Z}$:

$$\begin{aligned}
P((\alpha, \beta) \star Z_n = w | Z_n = z) &= P(\alpha \circ X_n - \beta \circ Y_n = w | X_n - Y_n = z) \\
&= \sum_{y=0}^{\infty} P(\alpha \circ X_n - \beta \circ Y_n = w | X_n = z + y, Y_n = y) \times \\
&\quad \times P(X_n = z + y, Y_n = y) \\
&= \sum_{y=0}^{\infty} e^{-\frac{\mu}{1-\alpha} - \frac{\nu}{1-\beta}} \frac{\left(\frac{\mu}{1-\alpha}\right)^{z+y} \left(\frac{\nu}{1-\beta}\right)^y}{(z+y)! y!} \times \\
&\quad \times \sum_{l=0}^y \binom{y}{l} \binom{z+y}{w+l} \alpha^{w+l} (1 - \alpha)^{z+y-w-l} \beta^l (1 - \beta)^{y-l} \\
&= \sum_{y=0}^{\infty} e^{-\frac{\mu}{1-\alpha} - \frac{\nu}{1-\beta}} \frac{\mu^{z+y} \nu^y}{(z+y)! y!} \times \\
&\quad \times \sum_{l=0}^y \binom{y}{l} \binom{z+y}{w+l} \left(\frac{\alpha}{1-\alpha}\right)^{w+l} \left(\frac{\beta}{1-\beta}\right)^l.
\end{aligned}$$

With thinning operator described in (2.7), one is able to introduce a more generalized *TINAR*(1) time series.

DEFINITION 2.2.1. *Let $\{Z_n\}$, $n \geq 0$ be a time series with skewed Skellam $\left(\frac{\mu}{1-\alpha}, \frac{\nu}{1-\beta}\right)$ marginal distribution, defined as follows:*

$$(2.10) \qquad \qquad \qquad Z_n = (\alpha, \beta) \star Z_{n-1} + \varepsilon_n, \quad n \in \mathbb{N},$$

where Z_0 is a random variable with skewed Skellam $\left(\frac{\mu}{1-\alpha}, \frac{\nu}{1-\beta}\right)$ distribution, $\{\varepsilon_n, n \geq 1\}$ represents a sequence of i.i.d. integer-valued random variables with skewed Skellam (μ, ν) distribution, such that ε_n and Z_{n-1} are independent for all $n \geq 1$ and the thinning operator " $(\alpha, \beta) \star$ " is defined by (2.7). The time series introduced here will be denoted as a skewed *TINAR*(1) time series of order 1.

2.2.2 Properties of the model

Similar as before, the skewed *TINAR*(1) time series can be represented in distribution as a difference between two independent differently parameterized Poisson *INAR*(1) time series. Namely, the distribution of $(\alpha, \beta) \star Z_{n-1}$ is skewed Skellam $\left(\frac{\alpha\mu}{1-\alpha}, \frac{\beta\nu}{1-\beta}\right)$. This can be proven in the same way as it was done in the previous section with distribution of

$\alpha \star Z_{n-1}$, bearing in mind that the *MGF* of the random variable Z with *Skellam* (μ, ν) distribution is given by (1.13). Furthermore, let

$$\begin{aligned} X_n &= \alpha \circ X_{n-1} + \xi_n, \quad n \in \mathbb{N}, \\ Y_n &= \beta \circ Y_{n-1} + \eta_n, \quad n \in \mathbb{N} \end{aligned}$$

be two independent Poisson *INAR*(1) time series with $Po\left(\frac{\mu}{1-\alpha}\right)$ and $Po\left(\frac{\nu}{1-\beta}\right)$ marginal distributions respectively. Obviously, $\{\xi_n\}$ and $\{\eta_n\}$ are mutually independent innovation processes with $Po(\mu)$ and $Po(\nu)$ distributions respectively. Random variable $\xi_n - \eta_n$ is *Skellam* (μ, ν) distributed, and hence $\xi_n - \eta_n \stackrel{d}{=} \varepsilon_n$. As it was proven earlier, the distribution of $\alpha \circ X_{n-1}$ is $Po\left(\frac{\alpha\mu}{1-\alpha}\right)$, and similarly to this, the distribution of $\beta \circ Y_{n-1}$ is $Po\left(\frac{\beta\nu}{1-\beta}\right)$. Thus, $\alpha \circ X_{n-1} - \beta \circ Y_{n-1}$ is *Skellam* $\left(\frac{\alpha\mu}{1-\alpha}, \frac{\beta\nu}{1-\beta}\right)$ distributed, which means that $\alpha \circ X_{n-1} - \beta \circ Y_{n-1} \stackrel{d}{=} (\alpha, \beta) \star Z_{n-1}$. Finally,

$$X_n - Y_n = (\alpha \circ X_{n-1} - \beta \circ Y_{n-1}) + (\xi_n - \eta_n) \stackrel{d}{=} (\alpha, \beta) \star Z_{n-1} + \varepsilon_n = Z_n.$$

In accordance with this equality, it holds that $E(Z_n) = E(X_n) - E(Y_n) = \frac{\mu}{1-\alpha} - \frac{\nu}{1-\beta}$ and $Var(Z_n) = Var(X_n) + Var(Y_n) = \frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}$.

Discussing the symmetric variant of the *TINAR*(1) time series, it has been already proven that Poisson *INAR*(1) is the strongly stationary and ergodic time series for $\alpha \in (0, 1)$. Since the skewed *TINAR*(1) time series is a difference of two independent and differently parameterized Poisson *INAR*(1) time series, it must also be strongly stationary and ergodic. The process characterization may be continued relying on the properties of the *INAR*(1) time series given in [2].

Theorem 2.2.1. *For skewed TINAR(1) time series $\{Z_n\}$ it holds:*

- i) $Z_n \stackrel{d}{=} \sum_{l=0}^{\infty} (\alpha, \beta) \star^{(l)} \varepsilon_{n-l}$;
- ii) $(Z_n, Z_{n-k}) \stackrel{d}{=} \left((\alpha, \beta) \star^{(k)} Z_{n-k} + \sum_{l=0}^{k-1} (\alpha, \beta) \star^{(l)} \varepsilon_{n-l}, Z_{n-k} \right)$;
- iii) $E(Z_n) = \alpha^n E(X_0) + \mu \sum_{l=0}^{n-1} \alpha^l - \beta^n E(Y_0) - \nu \sum_{l=0}^{n-1} \beta^l$;
- iv)

$$\begin{aligned} Var(Z_n) &= \alpha^{2n} Var(X_0) + (1-\alpha) \sum_{l=1}^n \alpha^{2l-1} E(X_{n-l}) + \sigma^2 \sum_{l=1}^n \alpha^{2(l-1)} \\ &\quad + \beta^{2n} Var(Y_0) + (1-\beta) \sum_{l=1}^n \beta^{2l-1} E(Y_{n-l}) + \sigma^2 \sum_{l=1}^n \beta^{2(l-1)}; \end{aligned}$$

$$v) \gamma_k = \alpha^k \gamma_0^{(X)} + \beta^k \gamma_0^{(Y)}.$$

Proof.

- i) Similar as earlier, the notation " $(\alpha, \beta) \star^{(l)}$ " represents l consecutive applications of the operator " $(\alpha, \beta) \star$ ". Bearing in mind all the facts regarding the distribution of the random variable Z_n and its latent components X_n and Y_n , it holds

$$Z_n \stackrel{d}{=} X_n - Y_n \stackrel{d}{=} \sum_{l=0}^{\infty} \alpha \circ^{(l)} \xi_{n-l} - \sum_{l=0}^{\infty} \beta \circ^{(l)} \eta_{n-l} = \sum_{l=0}^{\infty} (\alpha \circ^{(l)} \xi_{n-l} - \beta \circ^{(l)} \eta_{n-l}).$$

It has been already shown that the distribution of $\alpha \circ^{(l)} \xi_{n-l}$ is Poisson with distribution parameter $\alpha^l \mu$. Analogously, the distribution of $\beta \circ^{(l)} \eta_{n-l}$ is Poisson with distribution parameter $\beta^l \nu$. Consequently, the distribution of $\alpha \circ^{(l)} \xi_{n-l} - \beta \circ^{(l)} \eta_{n-l}$ is *Skellam* $(\alpha^l \mu, \beta^l \nu)$. On the other hand, the distribution of $(\alpha, \beta) \star \varepsilon_{n-l}$ is *Skellam* $(\alpha \mu, \beta \nu)$. Following the proof of Theorem 2.1.1, it can be proven that $(\alpha, \beta) \star^{(l)} \varepsilon_{n-l}$ also has *Skellam* $(\alpha^l \mu, \beta^l \nu)$ distribution. Thus, $\alpha \circ^{(l)} \xi_{n-l} - \beta \circ^{(l)} \eta_{n-l} \stackrel{d}{=} (\alpha, \beta) \star^{(l)} \varepsilon_{n-l}$, which implies that

$$Z_n \stackrel{d}{=} \sum_{l=0}^{\infty} (\alpha, \beta) \star^{(l)} \varepsilon_{n-l}.$$

- ii) From the proof of Theorem 2.1.1 one knows that

$$\begin{aligned} (Z_n, Z_{n-k}) &\stackrel{d}{=} (X_n, X_{n-k}) - (Y_n, Y_{n-k}) \\ &\stackrel{d}{=} \left((\alpha \circ^{(k)} X_{n-k} - \beta \circ^{(k)} Y_{n-k}) \right. \\ &\quad \left. + \sum_{l=0}^{k-1} (\alpha \circ^{(l)} \xi_{n-l} - \beta \circ^{(l)} \eta_{n-l}), X_{n-k} - Y_{n-k} \right). \end{aligned}$$

In accordance with facts already proven in (i), the distributions of $\alpha \circ^{(k)} X_{n-k}$ and $\beta \circ^{(k)} Y_{n-k}$ are Poisson with distribution parameters $\alpha^k \frac{\mu}{1-\alpha}$ and $\beta^k \frac{\nu}{1-\beta}$ respectively. Hence, $\alpha \circ^{(k)} X_{n-k} - \beta \circ^{(k)} Y_{n-k}$ has *Skellam* $(\alpha^k \frac{\mu}{1-\alpha}, \beta^k \frac{\nu}{1-\beta})$ distribution, which is as well the distribution of $(\alpha, \beta) \star^{(k)} Z_{n-k}$. Hence, $\alpha \circ^{(k)} X_{n-k} - \beta \circ^{(k)} Y_{n-k} \stackrel{d}{=} (\alpha, \beta) \star^{(k)} Z_{n-k}$. Same as in (i), $\alpha \circ^{(l)} \xi_{n-l} - \beta \circ^{(l)} \eta_{n-l} \stackrel{d}{=} (\alpha, \beta) \star^{(l)} \varepsilon_{n-l}$. Using that $X_{n-k} - Y_{n-k} \stackrel{d}{=} Z_{n-k}$, it finally holds

$$(Z_n, Z_{n-k}) \stackrel{d}{=} \left((\alpha, \beta) \star^{(k)} Z_{n-k} + \sum_{l=0}^{k-1} (\alpha, \beta) \star^{(l)} \varepsilon_{n-l}, Z_{n-k} \right).$$

Proofs for *iii*), *iv*) and *v*) are obtained by direct application of the properties given in [2].
□

2.2.3 Parameter estimation of the skewed *TINAR*(1) model

This subsection is based on results published in [12]. In order to estimate unknown parameters of the model, the *YW* method is going to be used. Let Z_1, Z_2, \dots, Z_N be a

sample of size N . As an initial step, the mean, variance and autocovariances of order one and two are provided, that is

$$\begin{aligned} E(Z_n) &= \frac{\mu}{1-\alpha} - \frac{\nu}{1-\beta}, \\ \text{Var}(Z_n) &= \frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}, \\ \gamma_1 &= \alpha \frac{\mu}{1-\alpha} + \beta \frac{\nu}{1-\beta}, \\ \gamma_2 &= \alpha^2 \frac{\mu}{1-\alpha} + \beta^2 \frac{\nu}{1-\beta}. \end{aligned}$$

Further, the system of four equations is generated by equalizing true moments with corresponding sample moments. Furthermore, by solving the system thus obtained, one gets the following estimators:

$$\begin{aligned} \hat{\mu} &= \frac{(\bar{S}^2 - \hat{\gamma}_1)(\bar{S}^2 + \bar{Z}) - \sqrt{(\hat{\gamma}_2 \bar{S}^2 - \hat{\gamma}_1^2) \left((\bar{S}^2)^2 - \bar{Z}^2 \right)}}{2\bar{S}^2} \\ &= \frac{(1 - \hat{\rho}_1) \left(1 + \frac{\bar{Z}}{\bar{S}^2} \right) - \sqrt{(\hat{\rho}_2 - \hat{\rho}_1^2) \left(1 - \left(\frac{\bar{Z}}{\bar{S}^2} \right)^2 \right)}}{2} \\ \hat{\nu} &= \frac{(\bar{S}^2 - \hat{\gamma}_1)(\bar{S}^2 - \bar{Z}) + \sqrt{(\hat{\gamma}_2 \bar{S}^2 - \hat{\gamma}_1^2) \left((\bar{S}^2)^2 - \bar{Z}^2 \right)}}{2\bar{S}^2} \\ &= \frac{(1 - \hat{\rho}_1) \left(1 - \frac{\bar{Z}}{\bar{S}^2} \right) - \sqrt{(\hat{\rho}_2 - \hat{\rho}_1^2) \left(1 - \left(\frac{\bar{Z}}{\bar{S}^2} \right)^2 \right)}}{2} \\ \hat{\alpha} &= \frac{\hat{\gamma}_1 (\bar{S}^2 + \bar{Z}) + \sqrt{(\hat{\gamma}_2 \bar{S}^2 - \hat{\gamma}_1^2) \left((\bar{S}^2)^2 - \bar{Z}^2 \right)}}{(\bar{S}^2)^2 + \bar{S}^2 \bar{Z}} \\ &= \frac{\hat{\rho}_1 \left(1 + \frac{\bar{Z}}{\bar{S}^2} \right) + \sqrt{(\hat{\rho}_2 - \hat{\rho}_1^2) \left(1 - \left(\frac{\bar{Z}}{\bar{S}^2} \right)^2 \right)}}{1 + \frac{\bar{Z}}{\bar{S}^2}} \\ \hat{\beta} &= \frac{\hat{\gamma}_1 (\bar{S}^2 - \bar{Z}) - \sqrt{(\hat{\gamma}_2 \bar{S}^2 - \hat{\gamma}_1^2) \left((\bar{S}^2)^2 - \bar{Z}^2 \right)}}{(\bar{S}^2)^2 - \bar{S}^2 \bar{Z}} \\ &= \frac{\hat{\rho}_1 \left(1 - \frac{\bar{Z}}{\bar{S}^2} \right) + \sqrt{(\hat{\rho}_2 - \hat{\rho}_1^2) \left(1 - \left(\frac{\bar{Z}}{\bar{S}^2} \right)^2 \right)}}{1 - \frac{\bar{Z}}{\bar{S}^2}}, \end{aligned}$$

where $\bar{S}^2 = \frac{1}{N} \sum_{i=1}^N (Z_i - \bar{Z})$ represents the sample variance and $\hat{\rho}_i = \frac{\gamma_i}{\bar{S}^2}$, $i = 1, 2$. Again, properties of the estimators such obtained haven't been analyzed, because these are not in the focus of this dissertation, regarding the intention to extract and predict latent

components of the skewed $TINAR(1)$ model. The asymptotic behavior of these estimators might be an interesting topic for the future research.

2.3 Latent components of the skewed $TINAR(1)$ time series-extraction and prediction

Following results given in [12], possibility of extracting and one-step ahead predicting latent components $\{X_n\}$ and $\{Y_n\}$, upon realizations of the time series $\{Z_n\}$, will be examined in this section.

First of all, for any time series $\{Z_n\}$ defined in distribution as a difference between two independent time series $\{X_n\}$ and $\{Y_n\}$,

$$Z_n \stackrel{d}{=} X_n - Y_n,$$

it will be shown in following lines that

$$(2.11) \quad E(X_{n+k}^r | Z_n = z) = \begin{cases} \frac{\sum_{l=0}^{\infty} E(X_{n+k}^r | X_n = l+z, Y_n = l) P(X_n = l+z) P(Y_n = l)}{P(Z_n = z)}, & z > 0, \\ \frac{\sum_{l=0}^{\infty} E(X_{n+k}^r | X_n = l, Y_n = l-z) P(X_n = l) P(Y_n = l-z)}{P(Z_n = z)}, & z \leq 0, \end{cases}$$

where $E(X_{n+k}^r | Z_n = z)$ is a conditional moment of X_{n+k} of order r , $r \in \mathbb{N}$, given $Z_n = z$. Namely, let $z > 0$. In that case,

$$\begin{aligned} E(X_{n+k}^r | Z_n = z) &= \sum_{j=0}^{\infty} j^r P(X_{n+k} = j | Z_n = z) = \sum_{j=0}^{\infty} j^r \frac{P(X_{n+k} = j, Z_n = z)}{P(Z_n = z)} \\ &= \frac{1}{P(Z_n = z)} \sum_{j=0}^{\infty} j^r \sum_{l=0}^{\infty} P(X_{n+k} = j, X_n = l+z, Y_n = l) \\ &= \frac{1}{P(Z_n = z)} \sum_{j=0}^{\infty} j^r \sum_{l=0}^{\infty} P(X_{n+k} = j | X_n = l+z, Y_n = l) \\ &\quad \times P(X_n = l+z) P(Y_n = l). \end{aligned}$$

Exchanging the order of summation, it holds

$$\begin{aligned} E(X_{n+k}^r | Z_n = z) &= \frac{1}{P(Z_n = z)} \sum_{l=0}^{\infty} P(X_n = l+z) P(Y_n = l) \\ &\quad \times \sum_{j=0}^{\infty} j^r P(X_{n+k} = j | X_n = l+z, Y_n = l) \\ &= \frac{1}{P(Z_n = z)} \sum_{l=0}^{\infty} E(X_{n+k}^r | X_n = l+z, Y_n = l) \\ &\quad \times P(X_n = l+z) P(Y_n = l). \end{aligned}$$

Corresponding equality when $z \leq 0$ can be obtained in the same way. Equivalently to this,

$$(2.12) \quad E(Y_{n+k}^r | Z_n = z) = \begin{cases} \frac{\sum_{l=0}^{\infty} E(Y_{n+k}^r | X_n=l+z, Y_n=l) P(X_n=l+z) P(Y_n=l)}{P(Z_n=z)}, & z > 0, \\ \frac{\sum_{l=0}^{\infty} E(Y_{n+k}^r | X_n=l, Y_n=l-z) P(X_n=l) P(Y_n=l-z)}{P(Z_n=z)}, & z \leq 0. \end{cases}$$

Now, let $\{Z_n\}$ be a *TINAR*(1) time series of order 1 with skewed *Skellam* $\left(\frac{\mu}{1-\alpha}, \frac{\nu}{1-\beta}\right)$ marginal distribution and let $\{X_n\}$ and $\{Y_n\}$ be two independent Poisson *INAR*(1) time series with $Po\left(\frac{\mu}{1-\alpha}\right)$ and $Po\left(\frac{\nu}{1-\beta}\right)$ marginal distributions respectively. Let these two time series be latent components of $\{Z_n\}$ and let

$$Z_n \stackrel{d}{=} X_n - Y_n.$$

In order to extract and predict $\{X_n\}$ and $\{Y_n\}$ upon realization of $\{Z_n\}$, conditional expectations $E(X_{n+k} | Z_n = z)$ and $E(Y_{n+k} | Z_n = z)$ for $k = 0, 1$ are going to be calculated. For this purpose, the usage of (2.11) and (2.12) for $r = 1$ provides appropriate results. Namely, considering the fact that the marginal distribution of the time series $\{Z_n\}$ is skewed *Skellam* $\left(\frac{\mu}{1-\alpha}, \frac{\nu}{1-\beta}\right)$, i.e.,

$$(2.13) \quad P(Z_n = z) = e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)} \left(\frac{\mu(1-\beta)}{\nu(1-\alpha)}\right)^{\frac{z}{2}} I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right),$$

where $I_z(x)$ represents the modified Bessel function of the first kind given by (1.11), the following calculation can be performed. Relying on (2.11) for $z > 0$ and $k = 0$, it holds that

$$\begin{aligned} E(X_n | Z_n = z) &= \frac{1}{P(Z_n = z)} \sum_{l=0}^{\infty} E(X_n | X_n = l + z, Y_n = l) P(X_n = l + z) P(Y_n = l) \\ &= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z)!} E(X_n | X_n = l + z, Y_n = l) \\ &= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z)!} (l+z) \\ &= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z-1)!}. \end{aligned}$$

Using (2.13) and (1.11), one confirms that

$$(2.14) \quad E(X_n | Z_n = z) = \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{z-1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right)}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right)}.$$

Similar to this, for $z \leq 0$, it can be obtained that

$$(2.15) \quad E(X_n|Z_n = z) = \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{-z+1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{-z} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}.$$

Shortened, (2.14) and (2.15) may be pooled in

$$E(X_n|Z_n = z) = \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{|z-1|} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{|z|} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}.$$

Since $z = E(Z_n|Z_n = z) = E(X_n - Y_n|Z_n = z) = E(X_n|Z_n = z) - E(Y_n|Z_n = z)$, it is easy to see that $E(Y_n|Z_n = z) = E(X_n|Z_n = z) - z$. Then, it holds for $z > 0$ that

$$\begin{aligned} E(Y_n|Z_n = z) &= E(X_n|Z_n = z) - z = \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{z-1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} - z \\ &= \frac{\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} \left(\sum_{l=0}^{\infty} \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)^{2l+z-1}}{l!(l+z-1)!} - \sum_{l=0}^{\infty} z \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)^{2l+z-1}}{l!(l+z)!} \right) \\ &= \frac{\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} \left(\sum_{l=0}^{\infty} \frac{(l+z) \left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)^{2l+z-1}}{l!(l+z)!} - \sum_{l=0}^{\infty} z \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)^{2l+z-1}}{l!(l+z)!} \right) \\ &= \frac{\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} \sum_{l=1}^{\infty} \frac{l \left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)^{2l+z-1}}{l!(l+z)!}. \end{aligned}$$

By involving the substitution $j = l - 1$ and using (1.11), one finds that

$$(2.16) \quad \begin{aligned} E(Y_n|Z_n = z) &= \frac{\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} \sum_{j=0}^{\infty} \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)^{2j+z+1}}{j!(j+z+1)!} \\ &= \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{z+1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}. \end{aligned}$$

Similar to this, for $z \leq 0$ it can be obtained that

$$(2.17) \quad E(Y_n|Z_n = z) = \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{-z-1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{-z} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}.$$

Again, (2.16) and (2.17) can be united in

$$E(Y_n|Z_n = z) = \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{|z+1|} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{|z|} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}.$$

Now, using the equality (2.11) for $k = 1$ and $r = 1$, one gets the expression for $E(X_{n+1}|Z_n = z)$. Let $z > 0$. Then,

$$\begin{aligned}
E(X_{n+1}|Z_n = z) &= \frac{1}{P(Z_n = z)} \sum_{l=0}^{\infty} E(X_{n+1}|X_n = l+z, Y_n = l) P(X_n = l+z) P(Y_n = l) \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z)!} E(X_{n+1}|X_n = l+z, Y_n = l) \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z)!} (\alpha(l+z) + \mu) \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \left(\sqrt{\frac{\frac{\mu}{1-\alpha}}{\frac{\nu}{1-\beta}}} \right)^z \\
&\quad \times \left[\alpha \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \sum_{l=0}^{\infty} \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right)^{2l+z-1}}{l!(l+z-1)!} + \mu \sum_{l=0}^{\infty} \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right)^{2l+z}}{l!(l+z)!} \right] \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \left(\sqrt{\frac{\frac{\mu}{1-\alpha}}{\frac{\nu}{1-\beta}}} \right)^z \alpha \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} I_{z-1} \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right) \\
&\quad + \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \left(\sqrt{\frac{\frac{\mu}{1-\alpha}}{\frac{\nu}{1-\beta}}} \right)^z \mu I_z \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right).
\end{aligned}$$

Finally, the usage of (2.13) provides that

$$(2.18) \quad E(X_{n+1}|Z_n = z) = \alpha \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{z-1} \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_z \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} + \mu.$$

Analogously to this, for $z \leq 0$,

$$(2.19) \quad E(X_{n+1}|Z_n = z) = \alpha \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{-z+1} \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{-z} \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} + \mu.$$

One may unite (2.18) and (2.19) in

$$E(X_{n+1}|Z_n = z) = \alpha \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{|z-1|} \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{|z|} \left(2 \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} + \mu.$$

For the other latent component $\{Y_n\}$, equality (2.12) is exploited in order to get one-step ahead conditional expectation, given Z_n . When $z > 0$, the following can be derived:

$$\begin{aligned}
E(Y_{n+1}|Z_n = z) &= \frac{1}{P(Z_n = z)} \sum_{l=0}^{\infty} E(Y_{n+1}|X_n = l+z, Y_n = l) P(X_n = l+z) P(Y_n = l) \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z)!} E(Y_{n+1}|X_n = l+z, Y_n = l) \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \sum_{l=0}^{\infty} \frac{\left(\frac{\mu}{1-\alpha}\right)^{l+z} \left(\frac{\nu}{1-\beta}\right)^l}{l!(l+z)!} (\beta l + \nu) \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \left(\sqrt{\frac{\mu}{1-\alpha}} \right)^z \\
&\quad \times \left[\beta \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \sum_{l=0}^{\infty} \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right)^{2l+z+1}}{l!(l+z+1)!} + \nu \sum_{l=0}^{\infty} \frac{\left(\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}}\right)^{2l+z}}{l!(l+z)!} \right] \\
&= \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \left(\sqrt{\frac{\mu}{1-\alpha}} \right)^z \beta \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} I_{z+1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right) \\
&\quad + \frac{e^{-\left(\frac{\mu}{1-\alpha} + \frac{\nu}{1-\beta}\right)}}{P(Z_n = z)} \left(\sqrt{\frac{\mu}{1-\alpha}} \right)^z \nu I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right).
\end{aligned}$$

Using (2.13), it holds

$$(2.20) \quad E(Y_{n+1}|Z_n = z) = \beta \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{z+1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_z \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} + \nu.$$

For $z \leq 0$, it won't be hard to show that

$$(2.21) \quad E(Y_{n+1}|Z_n = z) = \beta \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{-z-1} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{-z} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} + \nu.$$

Finally, (2.20) and (2.21) are affiliated in

$$E(Y_{n+1}|Z_n = z) = \beta \sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \frac{I_{|z+1|} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)}{I_{|z|} \left(2\sqrt{\frac{\mu}{1-\alpha} \frac{\nu}{1-\beta}} \right)} + \nu.$$

At the very end, it is possible to formulate statistics that can be used in order to extract or predict latent components of the skewed Skellam distributed time series. Statistics are formulated as follows:

$$\begin{aligned}
\hat{X}_n &= \sqrt{\frac{\hat{\mu}}{1-\hat{\alpha}} \frac{\hat{\nu}}{1-\hat{\beta}}} \frac{I_{|Z_n-1|} \left(2\sqrt{\frac{\hat{\mu}}{1-\hat{\alpha}} \frac{\hat{\nu}}{1-\hat{\beta}}} \right)}{I_{|Z_n|} \left(2\sqrt{\frac{\hat{\mu}}{1-\hat{\alpha}} \frac{\hat{\nu}}{1-\hat{\beta}}} \right)}, \\
\hat{Y}_n &= \sqrt{\frac{\hat{\mu}}{1-\hat{\alpha}} \frac{\hat{\nu}}{1-\hat{\beta}}} \frac{I_{|Z_n+1|} \left(2\sqrt{\frac{\hat{\mu}}{1-\hat{\alpha}} \frac{\hat{\nu}}{1-\hat{\beta}}} \right)}{I_{|Z_n|} \left(2\sqrt{\frac{\hat{\mu}}{1-\hat{\alpha}} \frac{\hat{\nu}}{1-\hat{\beta}}} \right)},
\end{aligned}$$

$$\begin{aligned}\widehat{X}_{n+1} &= \widehat{\alpha} \sqrt{\frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{\widehat{\nu}}{1-\widehat{\beta}}} \frac{I_{|Z_n-1|} \left(2\sqrt{\frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{\widehat{\nu}}{1-\widehat{\beta}}} \right)}{I_{|Z_n|} \left(2\sqrt{\frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{\widehat{\nu}}{1-\widehat{\beta}}} \right)} + \widehat{\mu}, \\ \widehat{Y}_{n+1} &= \widehat{\beta} \sqrt{\frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{\widehat{\nu}}{1-\widehat{\beta}}} \frac{I_{|Z_n+1|} \left(2\sqrt{\frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{\widehat{\nu}}{1-\widehat{\beta}}} \right)}{I_{|Z_n|} \left(2\sqrt{\frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{\widehat{\nu}}{1-\widehat{\beta}}} \right)} + \widehat{\nu}.\end{aligned}$$

One interesting fact can be noticed here. Namely, for $\mu = \nu$ and $\alpha = \beta$, statistics for extracting and predicting latent components are of the form:

$$\begin{aligned}\widehat{X}_n &= \frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{I_{|Z_n-1|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)}{I_{|Z_n|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)}, \\ \widehat{Y}_n &= \frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{I_{|Z_n+1|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)}{I_{|Z_n|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)}, \\ \widehat{X}_{n+1} &= \widehat{\alpha} \frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{I_{|Z_n-1|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)}{I_{|Z_n|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)} + \widehat{\mu}, \\ \widehat{Y}_{n+1} &= \widehat{\alpha} \frac{\widehat{\mu}}{1-\widehat{\alpha}} \frac{I_{|Z_n+1|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)}{I_{|Z_n|} \left(\frac{2\widehat{\mu}}{1-\widehat{\alpha}} \right)} + \widehat{\mu},\end{aligned}$$

which are exactly the formulas proposed in [10] for extracting and predicting latent components of the symmetric *TINAR*(1) time series. To summarize, formulas obtained in this section represent a generalization of the formulas obtained in the case of symmetric *TINAR*(1).

2.4 Simulation study

To verify the effectiveness of the proposed method for extracting and predicting latent components of the skewed *TINAR*(1) time series, the idea given in [12] is followed. Namely, simulations of size $N = 5000$ are created for each of two independent Poisson *INAR*(1) time series. Based on these values, realizations of the skewed *TINAR*(1) time series are generated as a difference of the values acquired in Poisson *INAR*(1) simulations. By means of realizations thus obtained, parameters μ , ν , α and β are estimated using proposed *YW* method. A proposed latent components modeling can be performed now using parameter estimates and simulated realizations of the skewed *TINAR*(1) model. The quality of modeling is estimated by calculating root mean squares (*RMS*) of differences between simulated latent components and their reconstructions and predictions. The quality of parameter estimates will not be analyzed, since the behavior of these estimates is not in the focus of this research, but the convenience of proposed models for extraction and prediction. This is why the corresponding *RMS*-s for extractions and predictions are discussed only. Four skewed *TINAR*(1) models are simulated with different combinations of parameters. Small values, large values, close and distant values are all considered. Parameters $\mu = 0.6$, $\nu = 0.9$, $\alpha = 0.2$, $\beta = 0.7$ are used for the first

model simulation. As may be seen, values of μ and ν are relatively small and close to each other. The second model simulation is based on relatively different parameter values $\mu = 2$, $\nu = 0.5$, $\alpha = 0.6$, $\beta = 0.3$. In order to simulate the third model, parameters $\mu = 4$, $\nu = 8$, $\alpha = 0.2$, $\beta = 0.4$ are used. Values of μ and ν are quite large and different this time. The fourth model is close to be a symmetric one, since the values of μ and ν are the same and the values of α and β are very close, i.e. $\mu = \nu = 3$, $\alpha = 0.6$, $\beta = 0.5$.

Table 2.1 presents results of the parameter estimation, as well as the corresponding RMS -s for latent components extractions and predictions. One can conclude that obtained extractions behave quite good, that is, all RMS values in penultimate column are small enough. Figure 2.1 shows trajectories of the originally simulated Poisson $INAR(1)$ components and their reconstructions extracted from skewed $TINAR(1)$ models. Considering the impossibility to present graphically the entire simulated Poisson $INAR(1)$ samples, only initial subsamples of sizes 100 are shown for each of the observed cases. The high fitting capacity of extracted latent components may be noticed.

Table 2.1: RMS values and YW parameter estimates for skewed $TINAR(1)$ simulations.

Exact values				YW estimates				RMS	$RMS(1)$
μ	ν	α	β	$\hat{\mu}$	$\hat{\nu}$	$\hat{\alpha}$	$\hat{\beta}$		
0.6	0.9	0.2	0.7	0.594	0.756	0.270	0.741	0.774	1.029
2	0.5	0.6	0.3	1.865	0.596	0.638	0.276	0.859	1.874
4	8	0.2	0.4	4.498	7.324	0.075	0.447	1.870	3.403
3	3	0.6	0.5	2.530	3.404	0.680	0.431	1.799	3.380

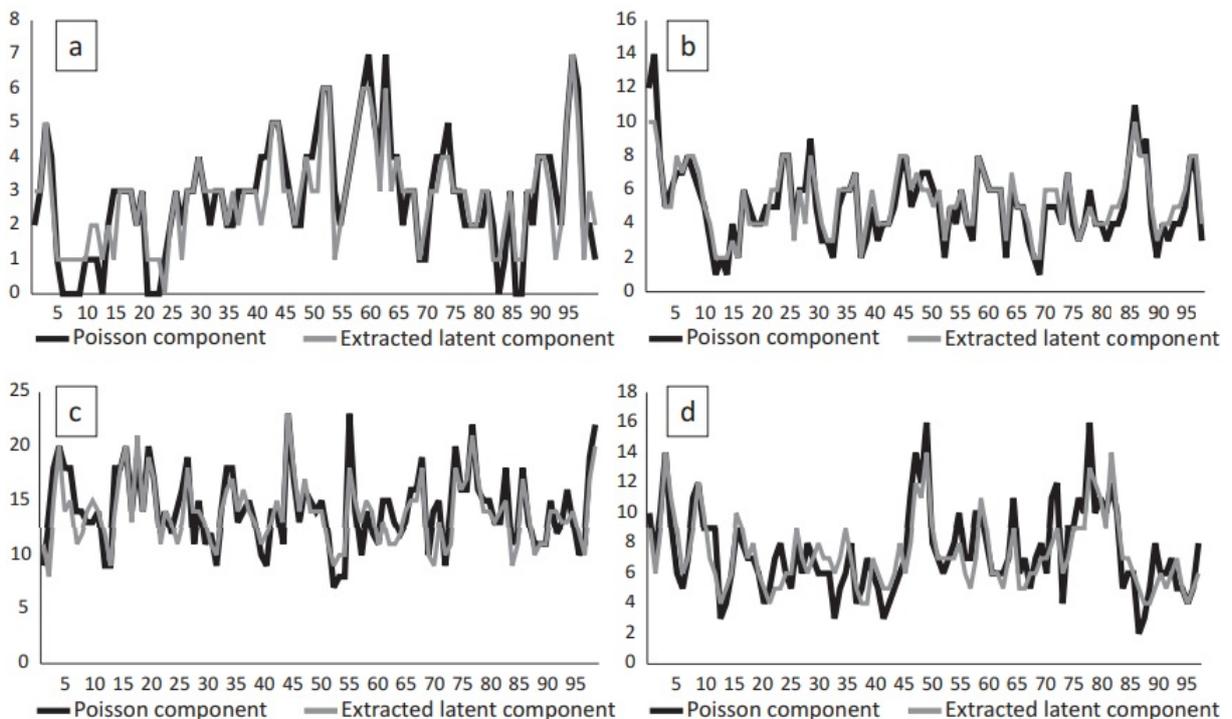


Figure 2.1: Trajectories of the simulated Poisson components $\{X_n\}$ and their extractions from skewed $TINAR(1)$ simulations: (a) $\mu = 0.6$, $\nu = 0.9$, $\alpha = 0.2$, $\beta = 0.7$; (b) $\mu = 2$, $\nu = 0.5$, $\alpha = 0.6$, $\beta = 0.3$; (c) $\mu = 4$, $\nu = 8$, $\alpha = 0.2$, $\beta = 0.4$; (d) $\mu = 3$, $\nu = 3$, $\alpha = 0.6$, $\beta = 0.5$.

As may be noticed, lag-one prediction results are not as good as extraction results. This conclusion is expected, by the way. The prediction RMS values (denoted as $RMS(1)$) are quite larger than their extraction equivalents, but still acceptable keeping in mind the mean values of Poisson components. Figure 2.2 shows trajectories of the originally simulated Poisson $INAR(1)$ components and their lag-one predictions. In comparison with extractions, trajectories of the lag-one predictions demonstrate a bit reduced ability of fitting and a bit decreased flexibility.

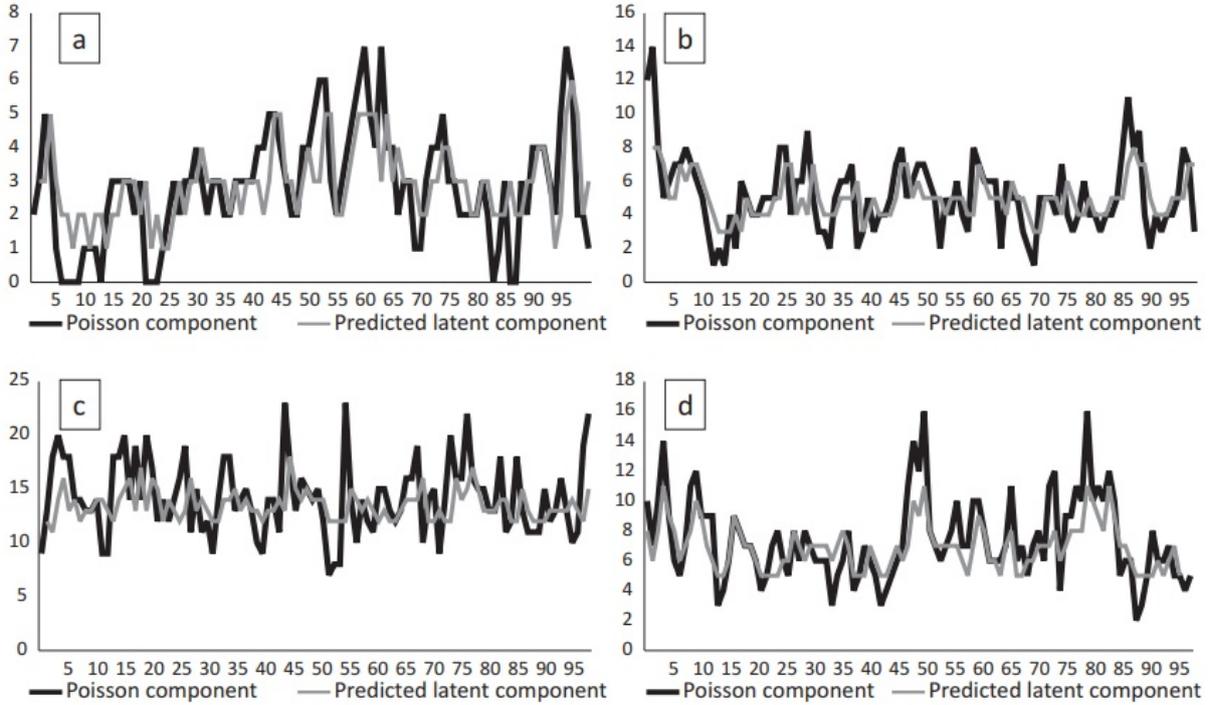


Figure 2.2: Trajectories of the simulated Poisson components $\{X_n\}$ and their lag-one predictions obtained from skewed $TINAR(1)$ simulations: (a) $\mu = 0.6$, $\nu = 0.9$, $\alpha = 0.2$, $\beta = 0.7$; (b) $\mu = 2$, $\nu = 0.5$, $\alpha = 0.6$, $\beta = 0.3$; (c) $\mu = 4$, $\nu = 8$, $\alpha = 0.2$, $\beta = 0.4$; (d) $\mu = 3$, $\nu = 3$, $\alpha = 0.6$, $\beta = 0.5$.

2.5 Application to real-life data

Checking the real power of proposed statistics in extracting and predicting latent components of the skewed $TINAR(1)$ time series, i.e. in extracting and predicting the minuend and the diminutive knowing only the difference, is quite challenging task. Keeping in mind that original latent components are unknown (since they are latent), extractions and predictions have nothing to compare with, which makes the real power of proposed statistics unrevealed. Thus, given task seems even more complicated. To bypass this problem, statistics were applied on "artificial" differences, as in [12]. Their components are not "so latent" and as such eligible for comparison with corresponding extractions and predictions. Goal differences of Southampton FC and BV Borussia Dortmund were used in order to extract and predict number of goals scored by each of these teams or their opponents. Data were collected from the website www.worldfootball.net. For Southampton FC, the data from the season 2010/11 to the season 2014/15 were observed, and for BV Borussia Dortmund, the data from the season 2000/01 to the season 2018/19 were analyzed. Beside these, the differences between number of criminal acts reported to

two police stations were also observed in order to reconstruct and predict the number of criminal acts reported to each police station. The data were collected from the website *www.forecastingprinciples.com*. The differences in number of robberies reported to two police stations, number 8800 and 9602, both in Rochester, New York, USA, during the period from January, 1991 to December, 2001 and the differences in number of aggravated assaults reported to two police stations, number 700 and 1700, also both in Rochester, New York, USA, during the same period were in focus. The distributions of all four data sequences were tested if significantly differ from *Skellam* distribution. The χ^2 -test revealed that all four empirical distributions fit in *Skellam* distribution ($\alpha = 0.05$). The results of the χ^2 -test are given in Table 2.2.

Table 2.2: Results of the χ^2 -test for testing differences between empirical and *Skellam* distributions.

Variable	χ^2	<i>p</i> -value
Southampton FC goal difference	10.708	0.217
BV Borussia Dortmund goas difference	10.462	0.313
Difference in number of robberies	8.181	0.145
Difference in number of aggravated assaults	4.318	0.888

During the model parameters estimation, obtained values for parameter β were negative, and hence replaced by small positive numbers. This could be associated to relatively small sample sizes. Again, the *RMS* is used as a measure of the fitting quality. As may be seen in Table 2.3 (penultimate column), the level of errors in extracting the components is, regarding the observed values, surprisingly good. The quality of extractions is presented in Figure 2.3.

Table 2.3: *RMS* values and *YW* parameter estimates for extractions and predictions from real-life data.

Variable	$\hat{\mu}$	$\hat{\nu}$	$\hat{\alpha}$	$\hat{\beta}$	<i>RMS</i>	<i>RMS</i> (1)
Southampton FC goal difference	1.358	2.199	0.274	0.001	0.798	1.421
BV Borussia Dortmund goal difference	1.544	1.875	0.252	0.001	0.822	1.557
Difference in number of robberies	1.194	0.850	0.280	0.001	0.425	1.152
Difference in number of aggravated assaults	0.630	1.221	0.294	0.001	0.536	1.068

As it was the case with the simulations, predictions have a bit higher values of *RMS*, denoted as *RMS*(1). A lack of ability to reach extremes can be noticed. The higher the extreme, the larger the deviation that prediction model makes from original data. Nevertheless, the prediction trajectories follow in general the trajectories of the real-life latent components. Figure 2.4 confirms such claim.

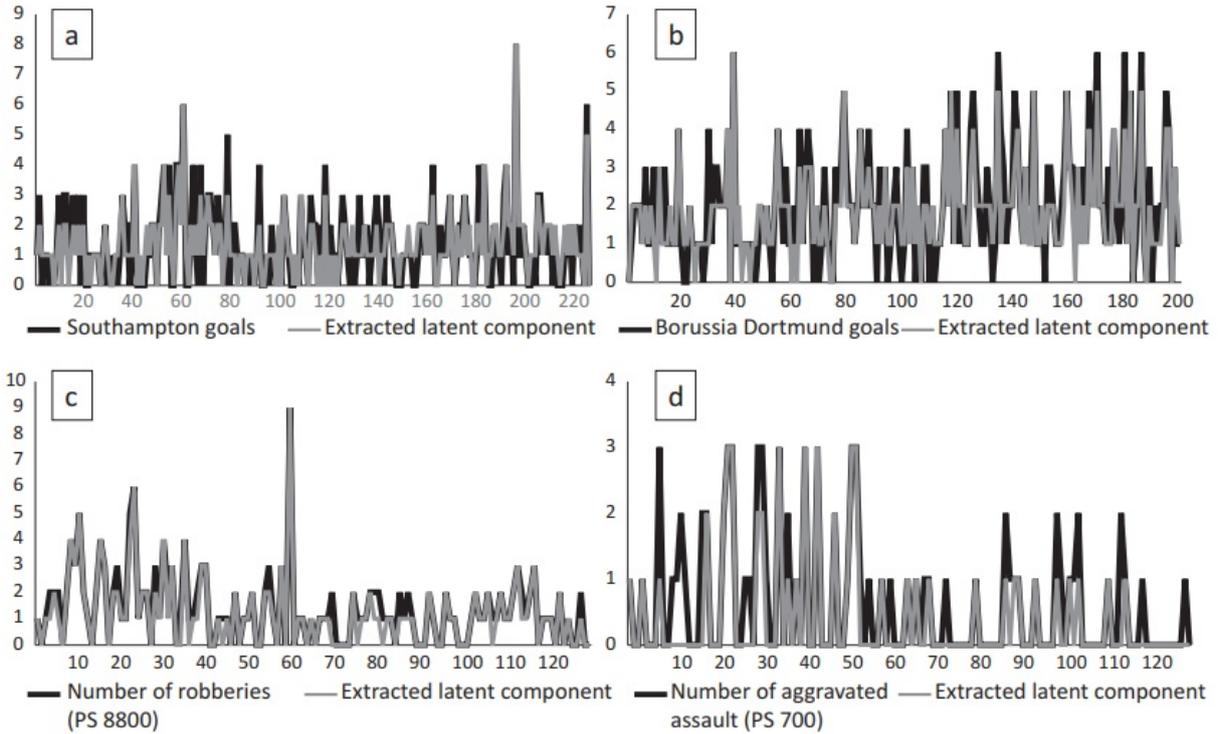


Figure 2.3: Trajectories of the real-life latent components and their extractions obtained from the corresponding skewed $TINAR(1)$ differences: (a) number of goals scored by Southampton; (b) number of goals scored by Borussia Dortmund; (c) number of robberies reported to the Rochester police station No. 8800; (d) number of aggravated assaults reported to the Rochester police station No. 700.

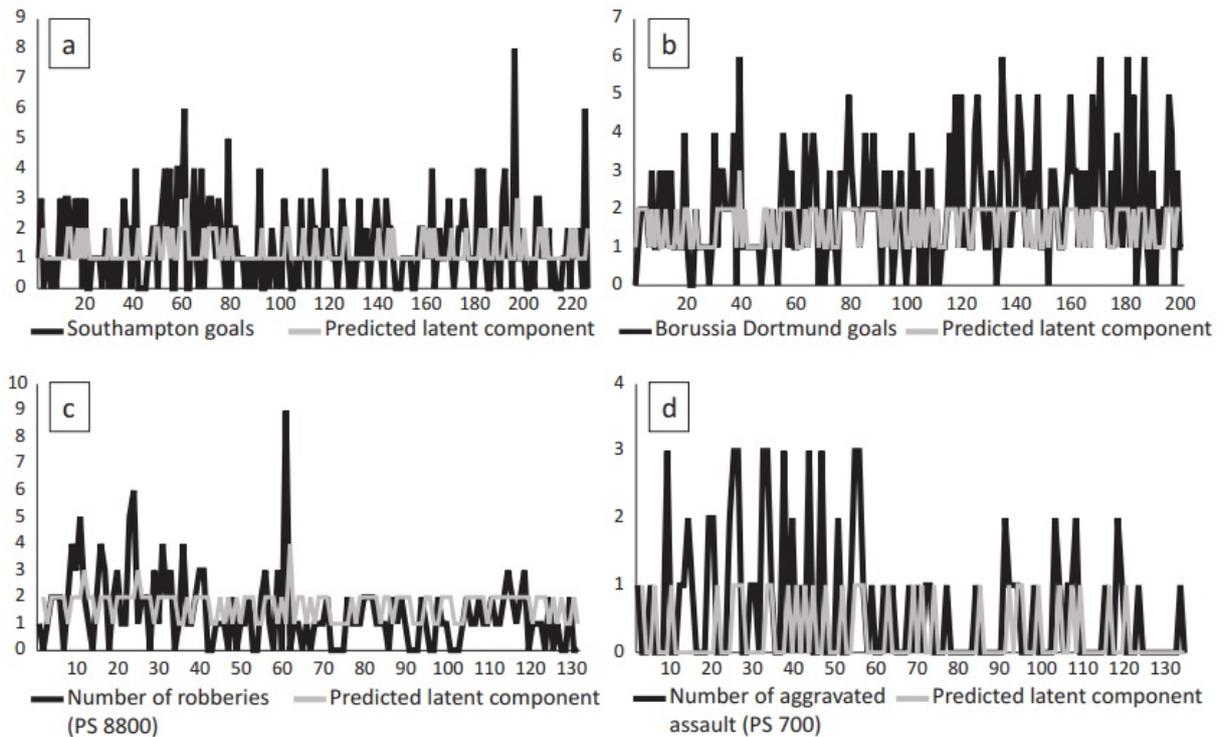


Figure 2.4: Trajectories of the real-life latent components and their lag-one predictions obtained from the corresponding skewed $TINAR(1)$ differences: (a) number of goals scored by Southampton; (b) number of goals scored by Borussia Dortmund; (c) number of robberies reported to the Rochester police station No. 8800; (d) number of aggravated assaults reported to the Rochester police station No. 700.

Chapter 3

Random Environment

Integer-Valued Autoregressive model with discrete Laplace marginal distribution

A $TINAR(1)$ model defined by (2.4) strongly motivated the researches to define new, more advanced models suitable for describing the integer-valued data series with both positive and negative values. Over time, several models with similar structure have appeared. All newly suggested models were stationary, because this property brings some important simplifications. However, the stationary time series are rigid, since some of their properties remain conserved in time. Nevertheless, real-life data sequences are usually not like that. The problem of introducing the non-stationarity into $INAR$ models with values over entire set of integers has not been considered in more detail so far.

A construction of the non-stationary $INAR$ model with positive and negative realizations will be in focus of this chapter. The chapter is heavily relied on results from [42], where the model itself is first mentioned. The non-stationarity is involved using the r states random environment process, given in Definition 1.2.4. The marginal distribution of the model is discrete Laplace, already introduced in Subsection 1.3.2. First of all, a construction of the model is given alongside with its properties, such as the innovation process distribution, the correlation structure and the k -step ahead conditional expectation. After that, the attention is paid to the unknown parameters estimation. For that purpose, two different methods are used: the Yule-Walker method and the conditional least squares method. Also, the quality of estimates is tested and confirmed on simulated data. Realized parameter estimates, based on these simulated data values, converge to true parameter values as the sample size increases. At the very end, the model is applied to real-life data, where it shows better results in regard to other models that might be considered as competitors. Successful application of the newly introduced model significantly increases the flexibility in modeling the data that is not necessarily above the axis denoting the time component in the Cartesian coordinate system, but oscillates around that axis.

3.1 Construction of the model

The results obtained by applying the $DLINAR(1)$ model, defined by (1.7), on real-life data series gave the motivation for new research. Although the model made some success in estimating the data it was tested on, it didn't manage to adjust to the elements that deviate significantly from zero due to its property of stationarity. In particular, the model was struggling to estimate the highest and the lowest peaks, with large differences between true values and corresponding modeled values. So, a large number of peaks and high absolute values of those peaks can make an application of the model more difficult. All the facts mentioned here leave room for model improvement. The first idea was to improve the $DLINAR(1)$ model in the same way as it was done in [38] with $NGINAR(1)$ model (see Subsection 1.2.8).

However, an attempt to apply this idea straightforward brought some difficulties, since the newly acquired model gave the same form of one-step ahead conditional expectation as it was the case with $DLINAR(1)$. This fact disables one to compare newly acquired model with $DLINAR(1)$ via RMS -s. To avoid this problem, the $DLINAR(1)$ model is improved using the concept given in [29], although in a bit simpler form. Namely, it is assumed that information on the environment state is not only carried by the marginal distribution parameter, but it can also be expressed through the thinning parameter value. In other words, it is assumed that the random environment process affects not only the marginal distribution of the model, but the thinning parameter value as well.

In order to improve the $DLINAR(1)$ model, some notations have to be introduced beforehand. To that purpose, let $\{z_n\}$, $n \in \mathbb{N}_0$ be a realization of the r states random environment process $\{Z_n\}$, mentioned in Definition 1.2.4. Further, for fixed $q, s \in E_r$, let $\{e_n(q, s), n \in \mathbb{N}\}$ be the sequences of i.i.d. random variables. To tag an element of the new time series, the notation $Y_n(z_n)$ will be used, where z_n determines the distribution of that element. Bearing in mind everything mentioned above, the following notations are introduced:

$$\begin{aligned}
 Y_n(Z_n) &= \sum_{s=1}^r Y_n(s) I_{\{Z_n=s\}}, \\
 e_n(Z_{n-1}, Z_n) &= \sum_{q=1}^r \sum_{s=1}^r e_n(q, s) I_{\{Z_{n-1}=q, Z_n=s\}}, \\
 \alpha_{Z_n} &= \sum_{s=1}^r \alpha_s I_{\{Z_n=s\}},
 \end{aligned}$$

where $I_{\{Z_n=s\}}$ represents an indicator function of the event $Z_n = s$. Now, a random environment $DLINAR$ time series of order 1 may be defined as it was done by [42]. But before introducing the definition of a new time series that will be in the focus of this chapter, it is necessary to define a random environment $INAR$ time series based on the thinning operator " $\alpha \odot$ ", with variable marginal distribution and inconstant thinning parameter value.

DEFINITION 3.1.1 ([42], Definition 2.1). *Let $\{Z_n\}$ be a random environment process with r possible states from the set $E_r = \{1, 2, \dots, r\}$, $r \in \mathbb{N}$. Further, let $\mathcal{M} = \{\mu_1, \mu_2, \dots, \mu_r\}$,*

$\mu_s > 0$, $s = 1, \dots, r$, be the set which consists of all possible mean values in corresponding states and let $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$, $\alpha_s \in (0, 1)$, $s = 1, \dots, r$, be the set which contains all possible values of thinning parameters corresponding to different states. In this case, the sequence $\{Y_n(Z_n)\}$ is called a Random Environment INAR time series of order 1 based on the thinning operator " $\alpha \odot$ ", with r states, the distribution parameters set \mathcal{M} and the thinning parameters set \mathcal{A} ($RrINAR_1(\mathcal{M}, \mathcal{A})$), if the random variable $Y_n(Z_n)$ is defined for $n \geq 1$ as

$$(3.1) \quad Y_n(Z_n) = \alpha_{Z_n} \odot Y_{n-1}(Z_{n-1}) + e_n(Z_{n-1}, Z_n),$$

where the operator " $\alpha_{Z_n} \odot$ " is defined by (1.6) and for all $q, s \in E_r$, $\{e_n(q, s)\}_{n \in \mathbb{N}}$ are the sequences of i.i.d. random variables satisfying the following conditions:

- (1) sequences of random variables $\{Z_n\}, \{e_n(1, 1)\}, \{e_n(1, 2)\}, \dots, \{e_n(r, r)\}$ are mutually independent;
- (2) random variables Z_m and $e_m(q, s)$ are independent of $Y_n(u)$ for all $n < m$ and all $q, s, u \in E_r$.

The predefined time series is very complex in general, so a simplified version has to be considered. To that purpose, instead of considering a random environment process $\{Z_n\}$, one can suppose knowing a realization $\{z_n\}$ of this process, which is a credible assumption justified in Subsection 1.2.8. Finally, the following definition holds.

DEFINITION 3.1.2 ([42], Definition 2.2). *Let $\{z_n\}$ be a realization of the random environment process $\{Z_n\}$ with r possible states from the set $E_r = \{1, 2, \dots, r\}$, $r \in \mathbb{N}$. Further, let $\mathcal{M} = \{\mu_1, \mu_2, \dots, \mu_r\}$, $\mu_s > 0$, $s = 1, \dots, r$, be the set which consists of all possible mean values in corresponding states and let $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$, $\alpha_s \in (0, 1)$, $s = 1, \dots, r$, be the set which contains all possible values of thinning parameters corresponding to different states. Then, $\{Y_n(z_n)\}$ is called a Random Environment Discrete Laplace INAR time series of order 1 with r states, the distribution parameters set \mathcal{M} and the thinning parameters set \mathcal{A} ($RrDLINAR_1(\mathcal{M}, \mathcal{A})$) if the random variable $Y_n(z_n)$ satisfies*

$$(3.2) \quad Y_n(z_n) = \alpha_{z_n} \odot Y_{n-1}(z_{n-1}) + e_n(z_{n-1}, z_n),$$

for $n \in \mathbb{N}$, where $\{e_n(z_{n-1}, z_n)\}_{n \in \mathbb{N}}$ are given in Definition 3.1.1, conditions (1)-(2) from Definition 3.1.1 are satisfied and the random variable $Y_n(z_n)$ has $DL\left(\frac{\mu_{z_n}}{1+\mu_{z_n}}\right)$ distribution, for all $n \in \mathbb{N}_0$.

To consider the time series introduced here as fully determined, the distributions of random variables $e_n(q, s)$ must be familiar for all $n \geq 1$ and all $q, s \in E_r$. These distributions are given in the theorem that follows.

Theorem 3.1.1 ([42], Theorem 2.1). *Let $\{Y_n(z_n)\}$ be a $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series given in Definition 3.1.2 and let $z_n = s$ and $z_{n-1} = q$ for some q and $s \in E_r$. If $0 < \alpha_s \leq \frac{\mu_s}{1+\max_{q \in E_r} \mu_q}$, then the distribution of the random variable $e_n(q, s)$ can be written as a mixture of discrete Laplace and skewed discrete Laplace distributed random variables in the following form:*

$$(3.3) \quad e_n(q, s) \stackrel{d}{=} \begin{cases} DL\left(\frac{\mu_s}{1+\mu_s}\right), & w.p. \left(1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}\right)^2, \\ SDL\left(\frac{\mu_s}{1+\mu_s}, \frac{\alpha_s}{1+\alpha_s}\right), & w.p. \frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \left(1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}\right), \\ SDL\left(\frac{\alpha_s}{1+\alpha_s}, \frac{\mu_s}{1+\mu_s}\right), & w.p. \frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \left(1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}\right), \\ DL\left(\frac{\alpha_s}{1+\alpha_s}\right), & w.p. \left(\frac{\alpha_s \mu_q}{\mu_s - \alpha_s}\right)^2. \end{cases}$$

Proof. As in [42], the theorem will be proven using the characteristic function $\varphi_{e_n(q,s)}(t)$ of the random variable $e_n(q, s)$. Based on the definition and properties of the time series and the assumption that $z_{n-1} = q$ and $z_n = s$, it holds

$$\varphi_{Y_n(s)}(t) = \varphi_{\alpha_s \odot Y_{n-1}(q)}(t) \cdot \varphi_{e_n(q,s)}(t),$$

whence

$$\varphi_{e_n(q,s)}(t) = \frac{\varphi_{Y_n(s)}(t)}{\varphi_{\alpha_s \odot Y_{n-1}(q)}(t)}.$$

According to [37], the characteristic functions of $Y_n(s)$ and $\alpha_s \odot Y_{n-1}(q)$ are of the following form:

$$\begin{aligned} \varphi_{Y_n(s)}(t) &= \frac{1}{(1 + \mu_s - \mu_s e^{it})(1 + \mu_s - \mu_s e^{-it})}, \\ \varphi_{\alpha_s \odot Y_{n-1}(q)}(t) &= \frac{(1 + \alpha_s - \alpha_s e^{it})(1 + \alpha_s - \alpha_s e^{-it})}{(1 + \alpha_s(1 + \mu_q) - \alpha_s(1 + \mu_q)e^{it})(1 + \alpha_s(1 + \mu_q) - \alpha_s(1 + \mu_q)e^{-it})}. \end{aligned}$$

Using the presented facts, it holds

$$\begin{aligned} \varphi_{e_n(q,s)}(t) &= \frac{[1 + \alpha_s(1 + \mu_q) - \alpha_s(1 + \mu_q)e^{it}][1 + \alpha_s(1 + \mu_q) - \alpha_s(1 + \mu_q)e^{-it}]}{(1 + \alpha_s - \alpha_s e^{it})(1 + \alpha_s - \alpha_s e^{-it})(1 + \mu_s - \mu_s e^{it})(1 + \mu_s - \mu_s e^{-it})} \\ &= \frac{A}{(1 + \alpha_s - \alpha_s e^{it})(1 + \alpha_s - \alpha_s e^{-it})} + \frac{B}{(1 + \alpha_s - \alpha_s e^{it})(1 + \mu_s - \mu_s e^{-it})} \\ &+ \frac{C}{(1 + \mu_s - \mu_s e^{it})(1 + \mu_s - \mu_s e^{-it})} + \frac{D}{(1 + \mu_s - \mu_s e^{it})(1 + \alpha_s - \alpha_s e^{-it})}. \end{aligned}$$

By equalizing the corresponding coefficients, one gets the system

$$\begin{aligned} A(1 + \mu_s)^2 + B(1 + \alpha_s)(1 + \mu_s) + C(1 + \alpha_s)^2 + D(1 + \mu_s)(1 + \alpha_s) &= (1 + \alpha_s(1 + \mu_q))^2 \\ A\mu_s(1 + \mu_s) + B\mu_s(1 + \alpha_s) + C\alpha_s(1 + \alpha_s) + D\alpha_s(1 + \mu_s) &= \alpha_s(1 + \mu_q)(1 + \alpha_s(1 + \mu_q)) \\ A\mu_s(1 + \mu_s) + B\alpha_s(1 + \mu_s) + C\alpha_s(1 + \alpha_s) + D\mu_s(1 + \alpha_s) &= \alpha_s(1 + \mu_q)(1 + \alpha_s(1 + \mu_q)) \\ A\mu_s^2 + B\alpha_s\mu_s + C\alpha_s^2 + D\alpha_s\mu_s &= (\alpha_s(1 + \mu_q))^2. \end{aligned}$$

Now, by solving the system thus obtained, it is derived that

$$A = \left(\frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \right)^2, \quad B = D = \frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \left(1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \right), \quad C = \left(1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \right)^2.$$

Bearing in mind the equality (1.16), characteristic functions of the random variables with $DL\left(\frac{\mu_s}{1+\mu_s}\right)$, $SDL\left(\frac{\mu_s}{1+\mu_s}, \frac{\alpha_s}{1+\alpha_s}\right)$, $SDL\left(\frac{\alpha_s}{1+\alpha_s}, \frac{\mu_s}{1+\mu_s}\right)$ and $DL\left(\frac{\alpha_s}{1+\alpha_s}\right)$ distributions are of the form

$$\begin{aligned} \varphi_1(t) &= \frac{1}{(1 + \mu_s - \mu_s e^{it})(1 + \mu_s - \mu_s e^{-it})}, \quad \varphi_2(t) = \frac{1}{(1 + \mu_s - \mu_s e^{it})(1 + \alpha_s - \alpha_s e^{-it})}, \\ \varphi_3(t) &= \frac{1}{(1 + \alpha_s - \alpha_s e^{it})(1 + \mu_s - \mu_s e^{-it})}, \quad \varphi_4(t) = \frac{1}{(1 + \alpha_s - \alpha_s e^{it})(1 + \alpha_s - \alpha_s e^{-it})} \end{aligned}$$

respectively. Now, it becomes obvious that the random variable $e_n(q, s)$ has got the distribution defined by (3.3).

At the end of the proof, it is left to verify that A, B, C and D are really probabilities. For that purpose, it has to be proven that all of them belong to $[0, 1]$ and that $A + B + C + D = 1$. The second condition is easily confirmable. To verify the first condition, it is enough to provide that $0 \leq \frac{\alpha_s \mu_q}{\mu_s - \alpha_s} \leq 1$. By solving this double inequality, one gets $\alpha_s \leq \frac{\mu_s}{1 + \mu_q}$. As this condition must hold for arbitrary q and s , and $\alpha_s \in (0, 1)$, it holds that $0 < \alpha_s \leq \frac{\mu_s}{1 + \max_{q \in E_r} \mu_q}$. This fact completes the proof. \square

The previous theorem enables an interesting observation. Namely, in accordance with distribution of the random variable $e_n(q, s)$ given in Theorem 3.1.1 and with fact that the discrete Laplace or skewed discrete Laplace distributed random variable can be represented in distribution as a difference between two random variables with geometric distributions, one is able to make an interesting conclusion, given in the following corollary.

Corollary 3.1.1 ([42], Corollary 2.1). *If the condition $0 < \alpha_s \leq \frac{\mu_s}{1 + \max_{q \in E_r} \mu_q}$ is satisfied, then the random variable $e_n(q, s)$ has the same distribution as the difference of two i.i.d. random variables $\varepsilon_n(q, s)$ and $\eta_n(q, s)$ distributed as*

$$(3.4) \quad \begin{cases} \text{Geom}\left(\frac{\mu_s}{1 + \mu_s}\right), & w.p. \left(1 - \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}\right) \\ \text{Geom}\left(\frac{\alpha_s}{1 + \alpha_s}\right), & w.p. \frac{\alpha_s \mu_q}{\mu_s - \alpha_s}. \end{cases}$$

The fact given in previous corollary may help one to calculate in a simplified way many properties of the innovation process, or the properties of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series itself. Some of them are presented in the following corollary.

Corollary 3.1.2 ([42], Corollary 2.2). *On the assumption that $z_n = s$ and $z_{n-1} = q$, $q, s \in E_r$, it holds:*

$$\begin{aligned} E(e_n(q, s)) &= 0, \\ \text{Var}(e_n(q, s)) &= 2(\mu_s(1 + \mu_s) - \alpha_s \mu_q(1 + 2\alpha_s + \alpha_s \mu_q)). \end{aligned}$$

Proof. The proof of the first equality is trivial, given that $\varepsilon_n(q, s)$ and $\eta_n(q, s)$ have the same distributions. Bearing in mind the form of the distribution of $\varepsilon_n(q, s)$ and $\eta_n(q, s)$ and using properties of the probability generating function (PGF), it is easy to prove that

$$\begin{aligned} \text{Var}(\eta_n(q, s)) = \text{Var}(\varepsilon_n(q, s)) &= \Phi''_{\varepsilon_n(q, s)}(1) + \Phi'_{\varepsilon_n(q, s)}(1) - [\Phi'_{\varepsilon_n(q, s)}(1)]^2 \\ &= \mu_s(1 + \mu_s) - \alpha_s \mu_q(1 + 2\alpha_s + \alpha_s \mu_q). \end{aligned}$$

Hence, it holds that

$$\begin{aligned} \text{Var}(e_n(q, s)) &= \text{Var}(\varepsilon_n(q, s)) + \text{Var}(\eta_n(q, s)) \\ &= 2(\mu_s(1 + \mu_s) - \alpha_s \mu_q(1 + 2\alpha_s + \alpha_s \mu_q)). \end{aligned}$$

\square

Remark. Some additional explanations are provided. First of all, for fixed $z_n = s$ and $z_{n-1} = q$, $q, s \in E_r$, random variables $\varepsilon_n(q, s)$ and $\eta_n(q, s)$ are distributed in the same way as the innovation process of the $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ model (see [29]). In addition, when $q = s$, the distribution of $\{e_n(q, s)\}$ matches with the distribution of the innovation process of the $DLINAR(1)$ model, given in [37].

3.2 Properties of the model

This section deals with the most important properties of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model, such as k -step ahead conditional expectation and correlation structure. The results obtaining methodology in the case of $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model is quite similar to the one given in [37] in the case of $DLINAR(1)$ model, which is expected given their common nature.

Some properties can be derived by observing the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series as a difference between two mutually independent $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series, when $\mathcal{P} = \{1\}$. In order to prove this claim, the following has been undertaken. For given sets \mathcal{M}, \mathcal{A} and $\mathcal{P} = \{1\}$, let

$$\begin{aligned} X_n^{(1)}(z_n) &= \alpha_{z_n} * X_{n-1}^{(1)}(z_{n-1}) + \varepsilon_n(z_{n-1}, z_n), \quad n \in \mathbb{N}, \\ X_n^{(2)}(z_n) &= \alpha_{z_n} * X_{n-1}^{(2)}(z_{n-1}) + \eta_n(z_{n-1}, z_n), \quad n \in \mathbb{N} \end{aligned}$$

be two mutually independent $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$, $\mathcal{P} = \{1\}$ time series with the same geometric $Geom\left(\frac{\mu_s}{1+\mu_s}\right)$, $\mu_s \in \mathcal{M}$, marginal distributions for fixed $z_n = s$. Further, let $\{\varepsilon_n(q, s)\}$ and $\{\eta_n(q, s)\}$ be two mutually independent processes with the same marginal distributions given in Corollary 3.1.1, for fixed values $z_n = s$ and $z_{n-1} = q$. By the definition of the $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series given in Subsection 1.2.9, random variables $X_{n-k}^{(1)}(u)$ and $\varepsilon_n(q, s)$, as well as $X_{n-k}^{(2)}(u)$ and $\eta_n(q, s)$, are mutually independent for all $k \geq 1$ and for all $q, s, u \in E_r$.

Let $\{Y_n(z_n)\}$ be a $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series with $DL\left(\frac{\mu_s}{1+\mu_s}\right)$ marginal distribution, given $z_n = s$. Since marginal distributions of $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$ are $Geom\left(\frac{\mu_s}{1+\mu_s}\right)$, Theorem 1.4.1, provides that

$$\alpha_{z_n} \odot Y_{n-1}(z_{n-1}) \stackrel{d}{=} \alpha_{z_n} * X_{n-1}^{(1)}(z_{n-1}) - \alpha_{z_n} * X_{n-1}^{(2)}(z_{n-1}),$$

which implies, alongside with Corollary 3.1.1, that

$$\begin{aligned} X_n^{(1)}(z_n) - X_n^{(2)}(z_n) &= \left(\alpha_{z_n} * X_{n-1}^{(1)}(z_{n-1}) - \alpha_{z_n} * X_{n-1}^{(2)}(z_{n-1}) \right) \\ &\quad + \left(\varepsilon_n(z_{n-1}, z_n) - \eta_n(z_{n-1}, z_n) \right) \\ &\stackrel{d}{=} \alpha_{z_n} \odot Y_{n-1}(z_{n-1}) + e_n(z_{n-1}, z_n) = Y_n(z_n). \end{aligned}$$

Using the results obtained in [29], it becomes trivial to show that $E(Y_n(z_n)) = 0$ and $Var(Y_n(z_n)) = 2Var\left(X_n^{(1)}(z_n)\right) = 2\mu_{z_n}(1 + \mu_{z_n})$.

The following property of " $\alpha \odot$ " provides a very useful decomposition. For that purpose, let Y , $X^{(1)}$, $X^{(2)}$ and D_l , $l \geq 1$, be random variables which satisfy conditions:

- (i) $Y \sim DL\left(\frac{\mu}{1+\mu}\right)$, $X^{(1)} \sim Geom\left(\frac{\mu}{1+\mu}\right)$, $X^{(2)} \sim Geom\left(\frac{\mu}{1+\mu}\right)$;
- (ii) $D_l \sim DL\left(\frac{\alpha}{1+\alpha}\right)$ for all $l \geq 1$;
- (iii) random variables $Y, X^{(1)}, X^{(2)}, D_l$, $l \geq 1$, and random variables involved in " $\alpha *$ " are independent.

In accordance with Theorem 1.4.3,

$$(3.5) \quad \alpha \odot Y \stackrel{d}{=} \text{sgn}(Y)(\alpha * |Y|) + \sum_{l=1}^{\min\{X^{(1)}, X^{(2)}\}} D_l.$$

This result enables one to prove the following claim.

Theorem 3.2.1 ([42], Theorem 3.1). *The RrDLINAR₁(\mathcal{M}, \mathcal{A}) time series $\{Y_n(z_n)\}$ is a Markov process.*

Proof. At the very beginning, events A and B are defined as follows:

$$A = \{Y_j(z_j) = y_j, j = 0, 1, \dots, n-2\}, \quad B = A \cup \{Y_{n-1}(z_{n-1}) = y_{n-1}\}.$$

For the purpose of clearer writing, the following notation is introduced:

$$\delta_{n-1}(z_{n-1}, z_n) = \sum_{l=1}^{\min\{X_{n-1}^{(1)}(z_{n-1}), X_{n-1}^{(2)}(z_{n-1})\}} D_l(z_n).$$

According to (3.5), it holds

$$\alpha_{z_n} \odot Y_{n-1}(z_{n-1}) = \text{sgn}(Y_{n-1}(z_{n-1}))(\alpha_{z_n} * |Y_{n-1}(z_{n-1})|) + \delta_{n-1}(z_{n-1}, z_n),$$

whereby $X_{n-1}^{(1)}(z_{n-1})$ and $X_{n-1}^{(2)}(z_{n-1})$ have the same $Geom\left(\frac{\mu_{z_{n-1}}}{1+\mu_{z_{n-1}}}\right)$ distributions and $D_l(z_n)$ is $DL\left(\frac{\alpha_{z_n}}{1+\alpha_{z_n}}\right)$ distributed for all $l = 1, 2, \dots, \min\{X_{n-1}^{(1)}(z_{n-1}), X_{n-1}^{(2)}(z_{n-1})\}$. Now,

$$\begin{aligned} P(Y_n(z_n) = y_n | B) &= \\ P(\text{sgn}(Y_{n-1}(z_{n-1}))(\alpha_{z_n} * |Y_{n-1}(z_{n-1})|) + \delta_{n-1}(z_{n-1}, z_n) + e_n(z_{n-1}, z_n) = y_n | B). \end{aligned}$$

Bearing in mind the property (iii) mentioned above and the condition (2) from Definition 3.1.1, it becomes obvious that

$$\begin{aligned} P(Y_n(z_n) = y_n | B) &= \sum_{l=-\infty}^{+\infty} P(\text{sgn}(Y_{n-1}(z_{n-1}))(\alpha_{z_n} * |Y_{n-1}(z_{n-1})|) = l | B) \\ &\times P(\delta_{n-1}(z_{n-1}, z_n) + e_n(z_{n-1}, z_n) = y_n - l) \\ &= \sum_{l=-\infty}^{+\infty} \binom{|y_{n-1}| + |l| - 1}{|l|} \frac{\alpha_{z_n}^{|l|}}{(1 + \alpha_{z_n})^{|y_{n-1}| + |l|}} \\ &\times P(\delta_{n-1}(z_{n-1}, z_n) + e_n(z_{n-1}, z_n) = y_n - l). \end{aligned}$$

Now, since the last expression depends only on y_{n-1} , it becomes obvious that the newly introduced RrDLINAR₁(\mathcal{M}, \mathcal{A}) time series is Markov process. \square

3.2.1 The k-step ahead conditional expectation

The possibility of approximating the (unknown) future values of the time series is reflected through the conditional expectation. The following theorem discusses this possibility.

Theorem 3.2.2 ([42], Theorem 3.2). *Let $\{Y_n(z_n)\}$ be a $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series. Then, for $k \geq 1$,*

$$(3.6) \quad E(Y_{n+k}(z_{n+k})|Y_n(z_n)) = \left(\prod_{l=1}^k \alpha_{z_{n+l}} \right) Y_n(z_n).$$

Proof. A mathematical induction will be used to derive the proof. In the initial step, let $k = 1$. Given Theorem 1.4.2, it holds

$$(3.7) \quad \begin{aligned} E(Y_{n+1}(z_{n+1})|Y_n(z_n)) &= E(\alpha_{z_{n+1}} \odot Y_n(z_n)|Y_n(z_n)) + E(e_{n+1}(z_n, z_{n+1})) \\ &= \alpha_{z_{n+1}} Y_n(z_n). \end{aligned}$$

Suppose the equation (3.6) holds for $k < m$. In the inductive step, (3.6) is verified for $k = m$ as well, keeping in mind the Markov property of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series. Namely,

$$\begin{aligned} E(Y_{n+m}(z_{n+m})|Y_n(z_n)) &= E[E(Y_{n+m}(z_{n+m})|Y_{n+m-1}(z_{n+m-1}), \dots, Y_n(z_n))|Y_n(z_n)] \\ &= E[E(Y_{n+m}(z_{n+m})|Y_{n+m-1}(z_{n+m-1}))|Y_n(z_n)] \\ &= E(\alpha_{z_{n+m}} Y_{n+m-1}(z_{n+m-1})|Y_n(z_n)) \\ &= \alpha_{z_{n+m}} \left(\prod_{l=1}^{m-1} \alpha_{z_{n+l}} \right) Y_n(z_n) \\ &= \left(\prod_{l=1}^m \alpha_{z_{n+l}} \right) Y_n(z_n). \end{aligned}$$

This makes the proof of the theorem fully completed. \square

3.2.2 Correlation structure

One of the most important features of any autoregressive time series is the interdependence of its individual elements. This feature is known as a correlation structure, which is discussed in the following theorem.

Theorem 3.2.3 ([42], Theorem 3.3). *The $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series $\{Y_n(z_n)\}$ given by (3.2) is positively correlated with its autocorrelation function given as*

$$(3.8) \quad Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) = \begin{cases} \left(\prod_{l=0}^{k-1} \alpha_{z_{n-l}} \right) \sqrt{\frac{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}{\mu_{z_n}(1+\mu_{z_n})}}, & k \geq 0, \\ \left(\prod_{l=1}^{-k} \alpha_{z_{n+l}} \right) \sqrt{\frac{\mu_{z_n}(1+\mu_{z_n})}{\mu_{z_{n-k}}(1+\mu_{z_{n-k}})}}, & k < 0. \end{cases}$$

Proof. Knowing that $\{Y_n(z_n)\}$ is a time series with k -step ahead conditional expectation of the form $E(Y_{n+k}(z_{n+k})|Y_n(z_n)) = \left(\prod_{l=1}^k \alpha_{z_{n+l}} \right) Y_n(z_n)$, unconditional expectation $E(Y_n(z_n)) = 0$ and finite variance $Var(Y_n(z_n)) = 2\mu_{z_n}(1 + \mu_{z_n})$, for $k \geq 0$, it becomes

easy to obtain that

$$\begin{aligned}
Cov(Y_n(z_n), Y_{n-k}(z_{n-k})) &= Cov(E(Y_n(z_n)|Y_{n-k}(z_{n-k})), Y_{n-k}(z_{n-k})) \\
&= Cov\left(\left(\prod_{l=0}^{k-1} \alpha_{z_{n-l}}\right) Y_{n-k}(z_{n-k}), Y_{n-k}(z_{n-k})\right) \\
&= \left(\prod_{l=0}^{k-1} \alpha_{z_{n-l}}\right) Var(Y_{n-k}(z_{n-k})) \\
&= 2 \left(\prod_{l=0}^{k-1} \alpha_{z_{n-l}}\right) \mu_{z_{n-k}} (1 + \mu_{z_{n-k}}),
\end{aligned}$$

whence it holds that

$$\begin{aligned}
Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) &= \frac{2 \left(\prod_{l=0}^{k-1} \alpha_{z_{n-l}}\right) \mu_{z_{n-k}} (1 + \mu_{z_{n-k}})}{\sqrt{2\mu_{z_n}(1 + \mu_{z_n})2\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}} \\
&= \left(\prod_{l=0}^{k-1} \alpha_{z_{n-l}}\right) \sqrt{\frac{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}{\mu_{z_n}(1 + \mu_{z_n})}}.
\end{aligned}$$

Similarly to this, for $k < 0$,

$$\begin{aligned}
Cov(Y_n(z_n), Y_{n-k}(z_{n-k})) &= E(Y_n(z_n) \cdot Y_{n-k}(z_{n-k})) \\
&= E[E(Y_n(z_n)Y_{n-k}(z_{n-k})|Y_n(z_n))] \\
&= E\left(Y_n(z_n) \left(\prod_{l=1}^{-k} \alpha_{z_{n+l}}\right) Y_n(z_n)\right) \\
&= \left(\prod_{l=1}^{-k} \alpha_{z_{n+l}}\right) Var(Y_n(z_n)) \\
&= \left(\prod_{l=1}^{-k} \alpha_{z_{n+l}}\right) 2\mu_{z_n}(1 + \mu_{z_n}),
\end{aligned}$$

whence one obtains

$$Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) = \left(\prod_{l=1}^{-k} \alpha_{z_{n+l}}\right) \sqrt{\frac{\mu_{z_n}(1 + \mu_{z_n})}{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}}.$$

□

Remark. If $z_n = z_{n-1} = \dots = z_{n-k} = s$, then $Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) = \alpha_s^{|k|}$. According to Theorem 1.4.4, this matches with autocorrelation function of the *DLINAR*(1) time series.

In addition, a validity of the double inequality $0 < Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) < 1$ will be proven. Bearing in mind the equality (3.8) and the facts that $\mu_{z_{n-k}} > 0$, $\mu_{z_n} > 0$ and $\alpha_{z_{n-l}} > 0$ for all $l = 0, 1, \dots, k-1$, it is obvious that $Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) > 0$ when $k \geq 0$. It is left to prove the validity of $Corr(Y_n(z_n), Y_{n-k}(z_{n-k})) < 1$.

For all $l = 0, 1, \dots, k-1$, $\alpha_{z_{n-l}} \leq \frac{\mu_{z_{n-l}}}{1 + \max_{q \in E_r} \mu_q}$, so, obviously

$$\alpha_{z_{n-l}} \leq \frac{\mu_{z_{n-l}}}{1 + \mu_{z_{n-l-1}}} < \frac{\mu_{z_{n-l}}}{\mu_{z_{n-l-1}}} < \frac{1 + \mu_{z_{n-l}}}{\mu_{z_{n-l-1}}}.$$

Then,

$$\alpha_{z_{n-l}}^2 < \frac{\mu_{z_{n-l}}}{1 + \mu_{z_{n-l-1}}} \cdot \frac{1 + \mu_{z_{n-l}}}{\mu_{z_{n-l-1}}},$$

which implies that $\alpha_{z_{n-l}} < \sqrt{\frac{\mu_{z_{n-l}}(1 + \mu_{z_{n-l}})}{\mu_{z_{n-l-1}}(1 + \mu_{z_{n-l-1}})}}$, and further,

$$\begin{aligned} \prod_{l=0}^{k-1} \alpha_{z_{n-l}} &< \sqrt{\frac{\mu_{z_n}(1 + \mu_{z_n})}{\mu_{z_{n-1}}(1 + \mu_{z_{n-1}})}} \sqrt{\frac{\mu_{z_{n-1}}(1 + \mu_{z_{n-1}})}{\mu_{z_{n-2}}(1 + \mu_{z_{n-2}})}} \dots \sqrt{\frac{\mu_{z_{n-k+1}}(1 + \mu_{z_{n-k+1}})}{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}} \\ &= \sqrt{\frac{\mu_{z_n}(1 + \mu_{z_n})}{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}}. \end{aligned}$$

Finally, it holds

$$\text{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) < \sqrt{\frac{\mu_{z_n}(1 + \mu_{z_n})}{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}} \sqrt{\frac{\mu_{z_{n-k}}(1 + \mu_{z_{n-k}})}{\mu_{z_n}(1 + \mu_{z_n})}} = 1.$$

It may be shown in a similar way that $0 < \text{Corr}(Y_n(z_n), Y_{n-k}(z_{n-k})) < 1$ when $k < 0$.

Regarding the form of the autocorrelation function, it may be concluded that it decreases toward 0, when k increases infinitely. This feature places the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ time series in the category of weakly correlated time series.

3.3 Parameter estimation of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model

This section will be dedicated to the estimation of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model parameters. For that purpose, two different kinds of estimates will be provided, Yule-Walker (YW) and conditional least squares (CLS) estimates. Although the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model represents a generalization of the $DLINAR(1)$ model, it is not stationary, and therefore it can not be ergodic, so the approach for proving the strong consistency given in [37] won't be helpful. In order to prove the strong consistency, a procedure similar to one described in [38] will be exploited. The main idea of the procedure might be described in two steps. In the first step, the initial sample $Y_1(z_1), Y_2(z_2), \dots, Y_N(z_N)$ is divided into r subsamples in the way that each subsample contains all the elements corresponding to only one state and doesn't contain elements corresponding to any other state. This partition can be written as:

$$\begin{aligned} I_s &= \{j \in \{1, 2, \dots, N\} | z_j = s\}, \quad s \in \{1, 2, \dots, r\}, \\ \bigcup_{s=1}^r I_s &= \{1, 2, \dots, N\}, \quad |I_s| = n_s, \quad n_1 + n_2 + \dots + n_r = N, \\ K_s &= \{Y_{s_1}(s), Y_{s_2}(s), \dots, Y_{s_{n_s}}(s)\}, \quad s_i \in I_s, \quad s_i < s_{i+1}, \quad i \in \{1, 2, \dots, n_s - 1\}, \end{aligned}$$

where K_s collects all the elements corresponding to the state s , I_s collects proper indexes of those elements and n_s represents cardinal number of the set K_s for all $s \in \{1, 2, \dots, r\}$. In the second step, parameters μ_s and α_s are estimated only using elements from the set K_s .

By observing an arbitrary K_s , $s \in \{1, 2, \dots, r\}$, it becomes clear that this set is consisted of i_s sequences of consecutive elements which are all in the same state s . Each of these sequences may be considered as a subsample of its own, denoted as $K_{s,l}$, $l = 1, 2, \dots, i_s$, and called the ‘maximal’ subsample. It is maximal in the way that it cannot be expanded neither to the left nor right side without violating the property that all of its elements correspond to the state s . To be precise, for arbitrary $K_{s,l}$, $l = 1, 2, \dots, i_s$, one can find natural numbers m_l and n_l , $m_l < n_l$, such that $z_{m_l} \neq s$, $z_{m_l+1} = z_{m_l+2} = \dots = z_{n_l} = s$, $z_{n_l+1} \neq s$. Now, each of these maximal subsamples $K_{s,l}$, $l = 1, 2, \dots, i_s$, may be observed as a sample of the $DLINAR(1)$ time series with marginal distribution parameter μ_s and thinning parameter α_s , and subsample K_s represents a disjoint union of subsamples $K_{s,1}, K_{s,2}, \dots, K_{s,i_s}$.

Now on, one can focus on obtaining the Yule-Walker estimates of unknown parameters of the model.

3.3.1 Yule-Walker estimation

In accordance with previous notations, the index sets $J_{s,l} = \{j \in \{1, 2, \dots, N\} | Y_j(z_j) \in K_{s,l}\}$, $l = 1, 2, \dots, i_s$ are introduced in the first place alongside with their cardinal numbers $n_{s,l} = |J_{s,l}|$, $n_{s,1} + n_{s,2} + \dots + n_{s,i_s} = n_s$. According to [37], the $DLINAR(1)$ time series is stationary and ergodic, whence the corresponding sample variance and the first-order sample covariance are strongly consistent estimates of the variance and the first-order covariance of the time series. Finally, by observing only the subsample $K_{s,l}$, these estimators are of the form

$$\hat{\gamma}_{0,l}^{(s)} = \frac{1}{n_{s,l}} \sum_{i \in J_{s,l}} Y_i^2(s) \quad \text{and} \quad \hat{\gamma}_{1,l}^{(s)} = \frac{1}{n_{s,l}} \sum_{i, i+1 \in J_{s,l}} Y_i(s) Y_{i+1}(s).$$

The following definition describes the estimators based on entire subsamples K_s , $s = 1, 2, \dots, r$, without taking into account that they are consisted of unconnected parts, i.e. of maximal subsamples.

DEFINITION 3.3.1 ([42], Definition 4.1). *For all $s = 1, 2, \dots, r$, estimators obtained from the subsample K_s with realizations corresponding to the state s are defined as*

$$(3.9) \quad \hat{\gamma}_0^{(s)} = \frac{1}{n_s} \sum_{i \in I_s} Y_i^2(s), \quad \hat{\gamma}_1^{(s)} = \frac{1}{n_s} \sum_{i, i+1 \in I_s} Y_i(s) Y_{i+1}(s).$$

The following theorem deals with strong constancy of mentioned estimators.

Theorem 3.3.1 ([42], Theorem 4.1). *Estimators $\hat{\gamma}_0^{(s)}$ and $\hat{\gamma}_1^{(s)}$ from Definition 3.3.1 are strongly consistent.*

Proof. This proof follows the idea given in [38]. First of all, the strong consistency of $\widehat{\gamma}_0^{(s)}$ will be proven, i.e. it will be shown that the equality $P(\widehat{\gamma}_0^{(s)} \rightarrow \gamma_0^{(s)}, n_s \rightarrow \infty) = 1$ holds. As mentioned earlier, $\widehat{\gamma}_{0,l}^{(s)}$ is strongly consistent for all $l \in \{1, 2, \dots, i_s\}$ and thus, it holds that $\widehat{\gamma}_{0,l}^{(s)} \rightarrow \gamma_0^{(s)}$, $n_{s,l} \rightarrow \infty$, everywhere except on the set $\Omega_{s,l}$, where $P(\Omega_{s,l}) = 0$. On the other hand, it holds

$$\widehat{\gamma}_0^{(s)} = \frac{1}{n_s} \sum_{i \in I_s} Y_i^2(s) = \frac{1}{n_s} \sum_{l=1}^{i_s} \sum_{i \in J_{s,l}} Y_i^2(s) = \sum_{l=1}^{i_s} \frac{n_{s,l}}{n_s} \frac{1}{n_{s,l}} \sum_{i \in J_{s,l}} Y_i^2(s) = \sum_{l=1}^{i_s} \frac{n_{s,l}}{n_s} \widehat{\gamma}_{0,l}^{(s)}.$$

Let $n_s = n_{s,1} + n_{s,2} + \dots + n_{s,i_s}$. On the assumption that $n_s \rightarrow \infty$, there is at least one $n_{s,l}$ that approaches infinity. In this stage, one can change the numeration of subsamples to get the following notations:

$$(3.10) \quad \begin{aligned} n_{s,l} &\rightarrow \infty, \quad \text{for all } l \in \{1, 2, \dots, b\}, \\ n_{s,l} &\rightarrow c_l \quad \text{for all } l \in \{b+1, b+2, \dots, i_s\}. \end{aligned}$$

This fact provides two important conclusions. First of all,

$$\lim_{n_s \rightarrow \infty} \frac{n_{s,l}}{n_s} = 0, \quad \text{for } l \in \{b+1, b+2, \dots, i_s\}.$$

The second,

$$n_s = n_{s,1} + n_{s,2} + \dots + n_{s,i_s} = \sum_{l=1}^b n_{s,l} + \sum_{l=b+1}^{i_s} n_{s,l} \longrightarrow \sum_{l=1}^b n_{s,l} + \sum_{l=b+1}^{i_s} c_l,$$

when $n_s \rightarrow \infty$, whereby the sum $\sum_{l=b+1}^{i_s} c_l$ is finite and thus incomparable with $n_{s,l}$ for all $l \in \{1, 2, \dots, b\}$. Hence, it is allowed to write $n_s = n_{s,1} + n_{s,2} + \dots + n_{s,b}$ if a limit value $n_s \rightarrow \infty$ is considered. Regarding everything mentioned above,

$$\lim_{n_s \rightarrow \infty} \widehat{\gamma}_0^{(s)} = \lim_{n_s \rightarrow \infty} \sum_{l=1}^b \frac{n_{s,l}}{n_s} \widehat{\gamma}_{0,l}^{(s)},$$

and consequently,

$$(3.11) \quad \begin{aligned} \lim_{n_s \rightarrow \infty} \widehat{\gamma}_0^{(s)} &= \lim_{n_{s,l} \rightarrow \infty, \forall l \in \{1, 2, \dots, b\}} \widehat{\gamma}_0^{(s)} \\ &= \lim_{n_{s,l} \rightarrow \infty, \forall l \in \{1, 2, \dots, b\}} \sum_{l=1}^b \frac{n_{s,l}}{n_s} \widehat{\gamma}_{0,l}^{(s)} \\ &= \gamma_0^{(s)} \lim_{n_{s,l} \rightarrow \infty, \forall l \in \{1, 2, \dots, b\}} \sum_{l=1}^b \frac{n_{s,l}}{n_s} \\ &= \gamma_0^{(s)}. \end{aligned}$$

At the beginning of the proof, it was mentioned that $\lim_{n_{s,l} \rightarrow \infty} \widehat{\gamma}_{0,l}^{(s)} = \gamma_0^{(s)}$ holds everywhere except on the set $\Omega_{s,l}$, whereby $P(\Omega_{s,l}) = 0$, and hence, the equality (3.11) holds everywhere except on the set $\Omega_s = \bigcup_{l=1}^b \Omega_{s,l}$, where

$$P(\Omega_s) = P\left(\bigcup_{l=1}^b \Omega_{s,l}\right) \leq \sum_{l=1}^b P(\Omega_{s,l}) = 0.$$

From the non-negativity of probability, it follows that $P(\Omega_s) = 0$. Thus, $\widehat{\gamma}_0^{(s)}$ is a strongly consistent estimator of the variance $\gamma_0^{(s)}$.

The proof for $\widehat{\gamma}_1^{(s)}$ is analogous. It holds that

$$\begin{aligned}
\widehat{\gamma}_1^{(s)} &= \frac{1}{n_s} \sum_{i,i+1 \in I_s} (Y_i(s)Y_{i+1}(s)) \\
&= \frac{1}{n_s} \sum_{l=1}^{i_s} \sum_{i,i+1 \in J_{s,l}} (Y_i(s)Y_{i+1}(s)) \\
&= \sum_{l=1}^{i_s} \frac{n_{s,l}}{n_s} \frac{1}{n_{s,l}} \sum_{i,i+1 \in J_{s,l}} (Y_i(s)Y_{i+1}(s)) \\
&= \sum_{l=1}^{i_s} \frac{n_{s,l}}{n_s} \widehat{\gamma}_{1,l}^{(s)} \\
&= \sum_{l=1}^b \frac{n_{s,l}}{n_s} \widehat{\gamma}_{1,l}^{(s)} + \sum_{l=b+1}^{i_s} \frac{n_{s,l}}{n_s} \widehat{\gamma}_{1,l}^{(s)}.
\end{aligned}$$

In the same way as it was done in the first part of the proof, one derives that

$$\begin{aligned}
\lim_{n_s \rightarrow \infty} \widehat{\gamma}_1^{(s)} &= \lim_{n_s \rightarrow \infty} \sum_{l=1}^b \frac{n_{s,l}}{n_s} \widehat{\gamma}_{1,l}^{(s)} \\
&= \lim_{n_{s,l} \rightarrow \infty, \forall l \in \{1,2,\dots,b\}} \sum_{l=1}^b \frac{n_{s,l}}{n_s} \widehat{\gamma}_{1,l}^{(s)} \\
&= \gamma_1^{(s)} \lim_{n_{s,l} \rightarrow \infty, \forall l \in \{1,2,\dots,b\}} \sum_{l=1}^b \frac{n_{s,l}}{n_s} \\
(3.12) \qquad &= \gamma_1^{(s)}.
\end{aligned}$$

Finally, the equality (3.12) holds everywhere except on the set $\Omega_s^1 = \bigcup_{l=1}^b \Omega_{s,l}^1$. In the same way as before, it is easy to prove that $P(\Omega_s^1) = 0$. Hence, one may claim that $P(\widehat{\gamma}_1^{(s)} \rightarrow \gamma_1^{(s)}, n_s \rightarrow \infty) = 1$, which actually represents the strong consistency of the estimator $\widehat{\gamma}_1^{(s)}$. \square

In final stage, parameters μ_s and α_s are estimated on the subsample K_s . According to Theorem 1.4.4, it holds

$$\gamma_0^{(s)} = 2\mu_s(1 + \mu_s), \quad \gamma_1^{(s)} = 2\alpha_s\mu_s(1 + \mu_s).$$

Thus,

$$\widehat{\mu}_s^{YW} = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 2\widehat{\gamma}_0^{(s)}}, \quad \widehat{\alpha}_s^{YW} = \frac{\widehat{\gamma}_1^{(s)}}{\widehat{\gamma}_0^{(s)}}.$$

Now, the estimator $\widehat{\mu}_s^{YW}$ is strongly consistent, according to Theorem 1.4.5 and the fact that $f(x) = -\frac{1}{2} + \frac{1}{2}\sqrt{1 + 2x}$, $x \geq 0$, is a continuous function. The strong consistency of $\widehat{\alpha}_s^{YW}$ follows from Theorem 1.4.6, when $p = 1$.

3.3.2 Conditional least squares estimation

In order to obtain the *CLS* estimates of unknown parameters of the model, one supposes that, for fixed $s = 1, \dots, r$, the entire subsample K_s represents the sample of the *DLINAR*(1) time series with marginal distribution parameter μ_s and thinning parameter α_s . This assumption is taken to enable *CLS* estimates providing and does not play any role in ensuring the validity of estimates such obtained. In particular, the assumption will not be used in confirming the property of strong consistency.

A *CLS* estimator of α_s is provided by minimizing the function

$$Q_N^{(s)}(\mu_s, \alpha_s) = \sum_{i, i+1 \in I_s} (Y_{i+1}(s) - E(Y_{i+1}(s)|Y_i(s)))^2 = \sum_{i, i+1 \in I_s} (Y_{i+1}(s) - \alpha_s Y_i(s))^2,$$

where $s = 1, 2, \dots, r$. Unfortunately, regarding the fact that μ_s does not figure within $Q_N^{(s)}$, the one-step *CLS* method cannot provide the estimate of this parameter. The two-step conditional least squares procedure given in [26] may be an alternative. Yet, the equations obtained using this alternative method contain parameter μ_s as an argument of the polynomial functions of degree 4 or higher, which makes the procedure too complicated. This is the reason why the mentioned method isn't taken into account. To summarize, only the one-step *CLS* estimate of α_s , $s = 1, \dots, r$, is provided.

The partial derivative of the function $Q_N^{(s)}$ with respect to α_s is of the form

$$\frac{\partial Q_N^{(s)}}{\partial \alpha_s} = -2 \sum_{i, i+1 \in I_s} Y_i(s)(Y_{i+1}(s) - \alpha_s Y_i(s)).$$

By equalizing the previous expression with zero and solving the equation thus obtained for parameter α_s , one gets the *CLS* estimator of α_s as

$$\hat{\alpha}_s^{CLS} = \frac{\sum_{i, i+1 \in I_s} Y_i(s)Y_{i+1}(s)}{\sum_{i, i+1 \in I_s} Y_i^2(s)}.$$

It has remained to prove the strong consistency of $\hat{\alpha}_s^{CLS}$. It is obvious that

$$\hat{\alpha}_s^{CLS} = \frac{\sum_{i, i+1 \in I_s} Y_i(s)Y_{i+1}(s)}{\sum_{i, i+1 \in I_s} Y_i^2(s)} = \frac{\hat{\gamma}_1^{(s)}}{\frac{1}{n_s} \sum_{i, i+1 \in I_s} Y_i^2(s)} = \frac{\hat{\gamma}_1^{(s)}}{\hat{\gamma}_0^{(s)} - \frac{1}{n_s} \sum_{l=1}^{i_s} Y_{n_l}^2(s)},$$

where $Y_{n_l}(s)$ represents the last element inside the maximal subsample $K_{s,l}$. Now,

$$\lim_{n_s \rightarrow \infty} \frac{1}{n_s} \sum_{i, i+1 \in I_s} Y_i^2(s) = \lim_{n_s \rightarrow \infty} \left(\hat{\gamma}_0^{(s)} - \frac{1}{n_s} \sum_{l=1}^{i_s} Y_{n_l}^2(s) \right) = \gamma_0^{(s)}.$$

Theorem 3.3.1 proved that $\lim_{n_s \rightarrow \infty} \hat{\gamma}_0^{(s)} = \gamma_0^{(s)}$ everywhere except on the set Ω_s , $P(\Omega_s) = 0$.

Beside this, $\lim_{n_s \rightarrow \infty} \frac{1}{n_s} \sum_{l=1}^{i_s} Y_{n_l}^2(s) = 0$. Hence,

$$\lim_{n_s \rightarrow \infty} \frac{1}{n_s} \sum_{i, i+1 \in I_s} Y_i^2(s) = \gamma_0^{(s)}$$

everywhere except on the set Ω_s . To be precise, $\frac{1}{n_s} \sum_{i, i+1 \in I_s} Y_i^2(s)$ is the strongly consistent estimator of $\gamma_0^{(s)}$. Bearing in mind that $\widehat{\gamma}_1^{(s)}$ is the strongly consistent estimator of $\gamma_1^{(s)}$, a strong consistency of $\widehat{\alpha}_s^{CLS}$ follows again from the Theorem 1.4.6 when $p = 1$.

3.4 Simulations study-Estimates quality analysis

As presented in [42], the quality of *YW* and *CLS* estimates defined in previous section will be examined. The aim is to show that the estimates of unknown parameters in practice really converge towards their true values, as the sample size increases. This may be achieved by simulating samples of desired sizes from the observed time series and calculating the estimates over such samples. Thus, hundred *RrDLINAR*₁(\mathcal{M}, \mathcal{A}) replicates are simulated, each of size 10000. Of course, it was necessary at first to simulate random environment processes, and then, using simulations thus obtained, to simulate corresponding *RrDLINAR*₁(\mathcal{M}, \mathcal{A}) time series. Sequences $\{Y_n(z_n)\}$ are simulated using the fact that the model itself is distributed as a difference between two independent *RrNGINAR*($\mathcal{M}, \mathcal{A}, \mathcal{P}$) time series $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$ when $\mathcal{P} = \{1\}$. Hence, hundred pairs of mutually independent *RrNGINAR*($\mathcal{M}, \mathcal{A}, \mathcal{P}$) time series $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$ are simulated first, and *RrDLINAR*₁(\mathcal{M}, \mathcal{A}) time series $\{Y_n(z_n)\}$ represent their differences.

In practice, it is often sufficient to observe time series with two or three different states, so these cases are considered separately below. The following parameters should be set:

- number of states r ,
- distribution parameter set $\mathcal{M} = \{\mu_1, \mu_2, \dots, \mu_r\}$,
- thinning parameter set $\mathcal{A} = \{\alpha_1, \alpha_2, \dots, \alpha_r\}$,
- initial states probability vector p_{vec} of the length r ,
- transition probability matrix p_{mat} of dimension $r \times r$, contained of transition probabilities $[p_{mat}]_{qs}$ from the state q to the state s .

3.4.1 The case of two states

It has been assumed here that the random environment process takes its values from the set $E_2 = \{1, 2\}$. To gain an insight into how parameter values affect the estimates behavior, the following combinations of parameters will be distinguished.

- 1.1. The distribution parameter set is $\mathcal{M} = \{1, 3\}$. Bearing in mind the condition $\alpha_s \in \left(0, \frac{\mu_s}{1 + \max_{q \in E_r} \mu_q}\right]$, the case when parameters α_s , $s = 1, 2$, were close to their upper limits has been analyzed. So, the thinning parameter set is $\mathcal{A} = \{0.25, 0.7\}$. A fair initial state has been set, i.e. $p_{vec} = (0.5, 0.5)$. As for the transition probability matrix, its significance is reflected through the fact that it determines the appearance of the random environment process, and therefore the appearance of the *RrDLINAR*₁(\mathcal{M}, \mathcal{A}) time series. In order to preserve the *RrDLINAR*₁(\mathcal{M}, \mathcal{A})

simulations in one state as long as possible, favoring the current state of the random environment process has been chosen, i.e. probabilities of the environment state change are significantly smaller. Thus, $p_{mat} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$.

- 1.2. In this case, parameters α_s , $s = 1, 2$, have been chosen to be smaller than the midpoints of intervals $\left(0, \frac{\mu_s}{1 + \max_{q \in E_r} \mu_q}\right]$, $s = 1, 2$. In order not to shrink these intervals too much, parameters μ_s , $s = 1, 2$, must have relatively close values. Hence, $\mathcal{M} = \{2, 3\}$ and $\mathcal{A} = \{0.2, 0.3\}$. The initial state has a slight tendency towards higher value due to the form of its distribution $p_{vec} = (0.45, 0.55)$, while transition probabilities of random states have been given by $p_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$. Again, the current state of the random environment process has been preferred, according to values on the transition matrix main diagonal.

3.4.2 The case of three states

Unlike the previous case, three possible random states will be supposed here, i.e. $E_3 = \{1, 2, 3\}$. Again, two different parameter combinations will be distinguished.

- 2.1 First of all, $\mathcal{M} = \{1, 2, 5\}$ has been picked. And yet again, the case when parameters α_s , $s = 1, 2, 3$, are close to their upper limits has been analyzed, so $\mathcal{A} = \{0.1, 0.25, 0.7\}$. An initial state has a slight tendency towards middle value, due to the value of its distribution $p_{vec} = (0.3, 0.4, 0.3)$. A transition probability

matrix is of the form $p_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$. The current state of the random environment process has been preferred, according to values on the transition matrix main diagonal.

- 2.2 The last parameter combination has been chosen again in the way to assign to α_s , $s = 1, 2, 3$, the values smaller than the midpoints of corresponding intervals $\left(0, \frac{\mu_s}{1 + \max_{q \in E_r} \mu_q}\right]$, $s = 1, 2, 3$. Thus, $\mathcal{M} = \{2, 3, 5\}$ and $\mathcal{A} = \{0.1, 0.2, 0.4\}$. An initial state is approximately fair, $p_{vec} = (0.33, 0.34, 0.33)$ and a transition probability matrix significantly prefers the current state of the random environment process,

$$p_{mat} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

3.4.3 Estimation results

In both of previously mentioned cases, with two or three different states, unknown parameters α_s and μ_s , $s = 1, 2, \dots, r$, $r = 2, 3$, of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model have been estimated. The task has been fulfilled using *YW* and *CLS* methods. Following the idea given in [38], the estimation of transition probabilities has been avoided, since those are not parameters of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model. Nevertheless, the transition

Table 3.1: The case of 2 states.

n_1	$\hat{\mu}_1^{YW}$	$\hat{\mu}_2^{YW}$	$\hat{\alpha}_1^{YW}$	$\hat{\alpha}_2^{YW}$	$\hat{\alpha}_1^{CLS}$	$\hat{\alpha}_2^{CLS}$
a) True values $\mathcal{M} = \{1, 3\}$, $\mathcal{A} = \{0.25, 0.7\}$,						
$p_{vec} = (0.5, 0.5)$, $p_{mat} = \begin{bmatrix} 0.6 & 0.4 \\ 0.2 & 0.8 \end{bmatrix}$.						
200	0.966	2.864	0.228	0.661	0.230	0.671
St. errors	0.163	0.497	0.135	0.113	0.130	0.095
500	0.979	2.942	0.235	0.685	0.240	0.690
St. errors	0.112	0.349	0.091	0.067	0.089	0.061
1000	0.983	3.002	0.240	0.699	0.242	0.699
St. errors	0.081	0.237	0.065	0.051	0.062	0.040
5000	0.987	3.001	0.248	0.699	0.248	0.699
St. errors	0.037	0.116	0.032	0.022	0.030	0.016
10000	0.989	2.999	0.249	0.699	0.249	0.700
St. errors	0.024	0.086	0.023	0.014	0.022	0.009
b) True values $\mathcal{M} = \{2, 3\}$, $\mathcal{A} = \{0.2, 0.3\}$,						
$p_{vec} = (0.45, 0.55)$, $p_{mat} = \begin{bmatrix} 0.7 & 0.3 \\ 0.3 & 0.7 \end{bmatrix}$.						
200	2.051	3.032	0.209	0.290	0.207	0.308
St. errors	0.273	0.367	0.113	0.120	0.110	0.116
500	2.025	2.991	0.196	0.293	0.196	0.293
St. errors	0.165	0.255	0.079	0.072	0.075	0.069
1000	2.023	2.991	0.202	0.297	0.201	0.297
St. errors	0.107	0.208	0.056	0.055	0.056	0.052
5000	1.999	3.001	0.198	0.299	0.199	0.297
St. errors	0.054	0.081	0.026	0.022	0.025	0.022
10000	2.000	3.001	0.199	0.299	0.199	0.299
St. errors	0.040	0.055	0.016	0.016	0.016	0.015

Table 3.2: The case of 3 states.

n_1	$\hat{\mu}_1^{YW}$	$\hat{\mu}_2^{YW}$	$\hat{\mu}_3^{YW}$	$\hat{\alpha}_1^{YW}$	$\hat{\alpha}_2^{YW}$	$\hat{\alpha}_3^{YW}$	$\hat{\alpha}_1^{CLS}$	$\hat{\alpha}_2^{CLS}$	$\hat{\alpha}_3^{CLS}$
a) True values $\mathcal{M} = \{1, 2, 5\}$, $\mathcal{A} = \{0.1, 0.25, 0.7\}$, $p_{vec} = (0.3, 0.4, 0.3)$, $p_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.2 & 0.1 & 0.7 \end{bmatrix}$.									
200	1.029	2.102	4.956	0.114	0.262	0.694	0.114	0.262	0.668
St. errors	0.201	0.406	0.944	0.129	0.129	0.140	0.124	0.129	0.112
500	1.008	2.037	5.050	0.106	0.258	0.704	0.106	0.258	0.698
St. errors	0.132	0.249	0.695	0.084	0.091	0.103	0.085	0.088	0.082
1000	0.992	2.036	5.017	0.103	0.258	0.704	0.102	0.257	0.698
St. errors	0.089	0.173	0.520	0.061	0.060	0.077	0.060	0.058	0.056
5000	0.999	2.009	5.008	0.102	0.249	0.702	0.102	0.252	0.702
St. errors	0.034	0.072	0.215	0.029	0.029	0.032	0.029	0.028	0.025
10000	1.000	2.003	4.994	0.101	0.250	0.701	0.101	0.249	0.701
St. errors	0.024	0.052	0.163	0.019	0.019	0.025	0.019	0.019	0.017
b) True values $\mathcal{M} = \{2, 3, 5\}$, $\mathcal{A} = \{0.1, 0.2, 0.4\}$, $p_{vec} = (0.33, 0.34, 0.33)$, $p_{mat} = \begin{bmatrix} 0.8 & 0.1 & 0.1 \\ 0.1 & 0.7 & 0.2 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}$.									
200	1.977	3.015	4.936	0.109	0.180	0.389	0.105	0.175	0.387
St. errors	0.358	0.651	0.736	0.121	0.140	0.118	0.113	0.0131	0.115
500	2.015	3.015	4.945	0.102	0.206	0.389	0.103	0.205	0.388
St. errors	0.215	0.365	0.470	0.082	0.120	0.082	0.079	0.115	0.079
1000	1.990	2.993	4.953	0.102	0.202	0.392	0.103	0.199	0.392
St. errors	0.156	0.247	0.311	0.059	0.079	0.053	0.060	0.078	0.049
5000	1.992	3.005	4.988	0.098	0.198	0.395	0.099	0.201	0.395
St. errors	0.077	0.119	0.145	0.027	0.035	0.026	0.027	0.034	0.026
10000	1.992	3.005	4.999	0.099	0.200	0.398	0.099	0.201	0.399
St. errors	0.047	0.086	0.111	0.017	0.022	0.017	0.018	0.023	0.018

probabilities could be estimated as in Subsection 1.2.8. The newly defined model, compared to $DLINAR(1)$, has greater number of unknown parameters, which leads to better data fitting, since the model is more flexible. On the other hand, the need for a larger sample is evident in order to obtain the same estimation accuracy. This is because each state has its own parameters that can be estimated only based on the part of the sample corresponding to that state. Thus, parameter estimates have been derived for subsamples of sizes 200, 500, 1000, 5000 and 10000. In each of these cases, hundred simulated samples have been used, and corresponding standard errors have been calculated. All estimates converge to true parameter values when sample sizes increase, while standard errors decrease towards 0. The differences between the YW and CLS estimates are negligible. It is noticeable as well that the smaller standard errors of $\hat{\mu}_j s^{YW}$ have appeared for smaller values of μ_s . This sounds reasonable because, with lower values of the parameter μ_s , lower dispersions of the time series values appear as well. Estimation results, corresponding to the cases with two and three random states, are given in Table 3.1 and Table 3.2 respectively.

3.5 Application to real-life data

So far, many non-stationary time series with both positive and negative values have been modeled using stationary models, in the absence of better solutions. One such example is the data collected by the City of Pittsburgh Bureau of Police, which represents the differences in number of motor vehicle thefts reported on a monthly basis to police stations number 1608 and 2811, in Pittsburgh, Pennsylvania, USA, between January, 1990 and December, 2001. The data sequence is given in Table 3.3. In [37], the $DLINAR(1)$ model was proved to be convincingly the best for this data. However, only stationary models were considered as competitors, although the data indicates that sharp ups and downs occasionally occur. As given in [42], it is reasonable to assume that the non-stationary $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model will be even more appropriate.

Table 3.3: Differences between motor vehicle thefts reported on a monthly basis to police stations number 1608 and 2811.

12	-1	2	3	8	-2	-3	4	4	6	5	5	5	4	4
5	4	5	4	0	1	0	1	2	3	-6	0	-1	-1	1
0	2	-1	0	1	-4	-5	-13	-4	-4	-5	-4	-6	-5	-8
-5	-5	-4	-4	-6	-5	0	1	-3	3	0	1	-2	0	0
-3	-1	-3	-3	-1	3	1	-1	0	0	-1	-1	2	1	1
1	3	0	2	1	0	0	2	1	1	-2	-2	-1	0	1
0	0	-3	0	1	-2	0	-2	2	-2	-3	2	2	2	3
2	1	-2	0	0	2	3	-3	0	-2	3	3	1	0	0
2	3	1	0	-3	-2	1	-3	-3	-3	2	3	-2	-2	1
3	1	2	0	3	2	3	2	-3						

Based on the sample of size $N = 144$, the mean difference is $\bar{y} = 0.048$. This proves the fact that mean values of the motor vehicle theft counts, denoted in stations number 1608 and 2811, are approximately the same. This condition is crucial here, since it had

been claimed earlier that both time series $\{X_n^{(1)}(z_n)\}$ and $\{X_n^{(2)}(z_n)\}$ must have the same distributions.

In order to get information about the character of the observed data sequence, one can take a look at the plots of autocorrelation and partial autocorrelation functions, given in Figure 3.1. The figure successfully justifies the usage of $INAR(1)$ modeling. Among all $INAR$ models of order 1, there is only one more model familiar so far beside $DLINAR(1)$ which may consist of positive and negative integer values and which, at the same time, assumes the same distribution for $\{X_n^{(1)}\}$ and $\{X_n^{(2)}\}$. It's the $TINAR(1)$ model with symmetric Skellam marginal distribution, introduced in Section 2.1. This model will be used as a competitor to $DLINAR(1)$ and $RrDLINAR_1(\mathcal{M}, \mathcal{A})$. To reinforce competition, the $STINAR(1)$ model given in [6] will also be considered as competitor. Brief notes on this model can be found within Section 1.2.7.

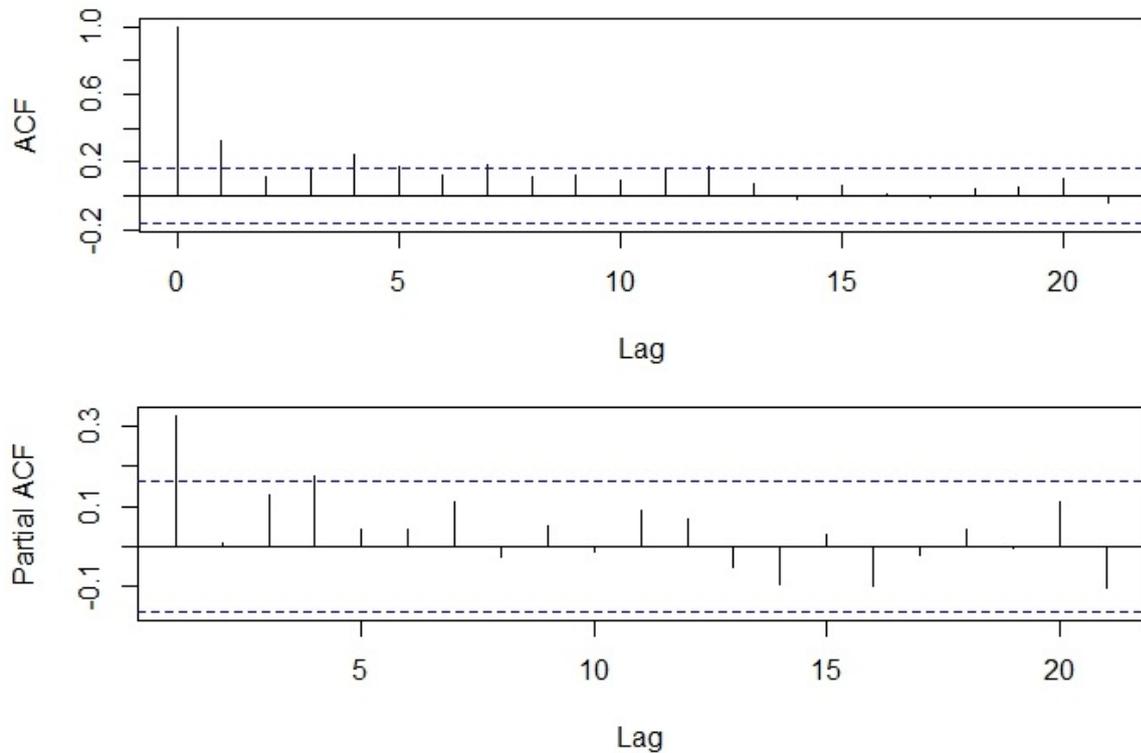


Figure 3.1: Autocorrelations and partial autocorrelations.

The next step is data clustering. This is how one actually obtains estimates of the corresponding random environment process. As explained in Subsection 1.2.8, it is assumed that the realized value of the random environment process in n -th month is equal s , i.e. $z_n = s$, $s \in E_r$, if the theft difference in that month is in the s -th cluster. In that way, $\{z_n\}$ is fully determined. For this particular case, it is decided to divide the theft difference realizations into two clusters. For that purpose, the K-means clustering technique is performed. Figure 3.2 provides such obtained clustering results. According to the figure, a decision to divide the theft differences into two clusters is proved to be reasonable. In the first cluster (triangles), one may find the differences that don't deviate significantly from zero. However, in two time intervals, a different behavior is noticed. Precisely,

from August, 1990 to July, 1991 and from December, 1993 to March, 1995 the deviations from zero are quite large. This suggests that environment state changes could have happened. Hence, the theft difference realizations from these two periods are located in the second cluster (circles). With each increase in number of clusters, at least one cluster with very few realizations in it is obtained. This leads to frequent state changes, which ruins any chance to successfully apply models in random environment, including $RrDLINAR_1(\mathcal{M}, \mathcal{A})$. Thus, $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ models with more than two different environment states should not be discussed for this data.

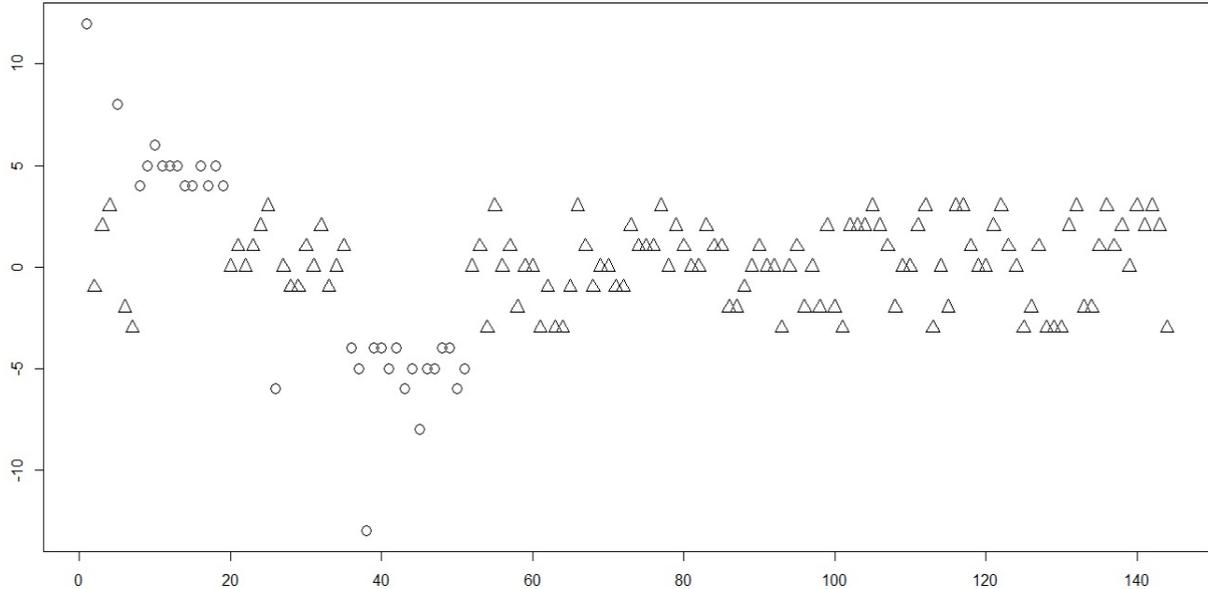


Figure 3.2: K-means clustering results.

The validity of the $RrDLINAR_1(\mathcal{M}, \mathcal{A})$ model will be examined in two ways, by testing its fitting quality and forecasting accuracy. For that purpose, the data sample is divided into two sets, the training set and the prediction set. The training set contains the first 120 sample elements, and will be used in estimating model parameters and evaluating the fitting quality. On the other hand, the prediction set contains the last 24 elements of the sample and will be utilized to assess the forecasting accuracy.

3.5.1 Fitting quality evaluation

To evaluate the fitting quality, unknown model parameters μ_1, μ_2, α_1 and α_2 must be estimated first. This was done using the corresponding YW estimators. Now, the construction of the $R2DLINAR_1(\mathcal{M}, \mathcal{A})$ time series $\{Y_n(z_n)\}$ can be performed. Finally, the RMS of differences between real-life data and modeled data is calculated and compared to corresponding counterparts of other models. These RMS values, together with corresponding YW estimates, are presented in Table 3.4. Providing the smallest RMS value, the $R2DLINAR_1(\mathcal{M}, \mathcal{A})$ model based on the two states random environment process proved to be the best for fitting given real-life data.

The realized time series of theft differences is given in Figure 3.3, alongside with one-step ahead predictions of $R2DLINAR_1(\mathcal{M}, \mathcal{A})$ and $STINAR(1)$ models. Predicted values of

Table 3.4: *RMS* values and *YW* parameter estimates for *INAR*(1) modeling of the theft differences.

Model	<i>YW</i>	<i>RMS</i>
<i>TINAR</i> (1)	$\hat{\alpha} = 0.333$ $\hat{\lambda} = 1.817$	2.690
<i>DLINAR</i> (1)	$\hat{\alpha} = 0.331$ $\hat{\mu} = 1.881$	2.689
<i>STINAR</i> (1)	$\hat{\alpha} = 0.337$ $\hat{\mu}_1 = 2.007$ $\hat{\mu}_2 = 1.991$	2.680
<i>R2DLINAR</i> ₁ (\mathcal{M}, \mathcal{A})	$\hat{\alpha}_1 = 0.189$ $\hat{\alpha}_2 = 0.808$ $\hat{\mu}_1 = 0.815$ $\hat{\mu}_2 = 3.648$	2.187

TINAR(1) and *DLINAR*(1) models are omitted, since they do not differ much from the predictions provided by *STINAR*(1). Obviously, both models shown in the figure approximate well the values close to zero. A big difference in fitting quality is noticed in realizations that deviate significantly from zero. In this case, the *R2DLINAR*₁(\mathcal{M}, \mathcal{A}) model shows much better ability to adjust to the real-life data. The higher the deviations, the larger the benefit that *R2DLINAR*₁(\mathcal{M}, \mathcal{A}) model offers in regard to *STINAR*(1). This flexibility is unquestionably a repercussion of the non-stationary nature of the newly proposed *RrDLINAR*₁(\mathcal{M}, \mathcal{A}) model. At the very end, the *R2DLINAR*₁(\mathcal{M}, \mathcal{A}) trajectory perfectly follows the trajectory of the realized time series.

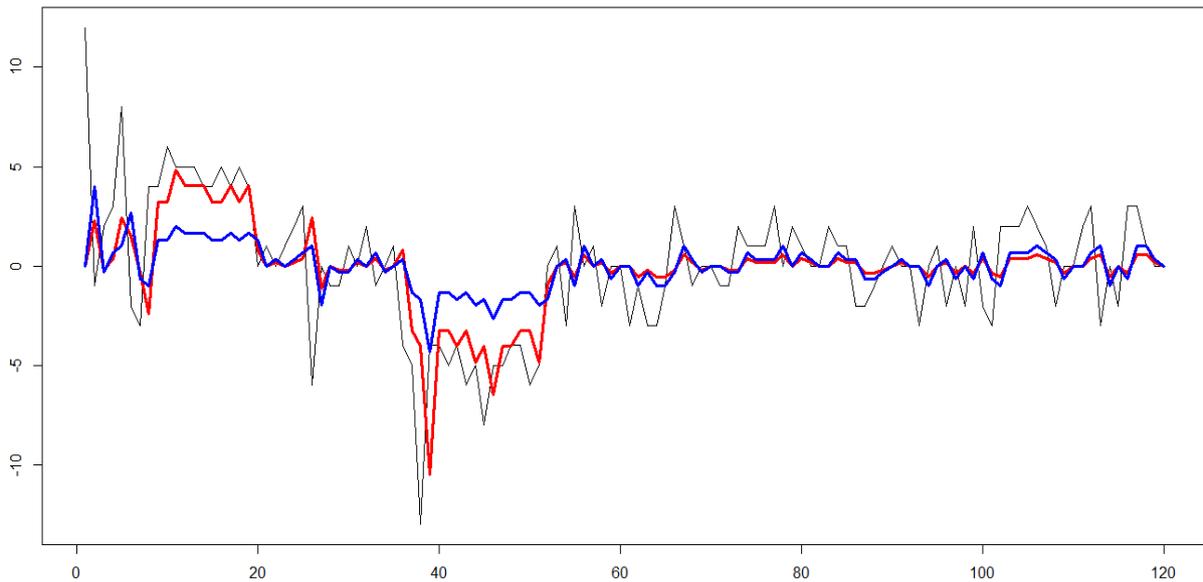


Figure 3.3: Results of fitting the theft difference realizations: black–real-life theft differences; red–*R2DLINAR*₁(\mathcal{M}, \mathcal{A}) predictions; blue–*STINAR*(1) predictions.

3.5.2 Forecasting accuracy assessment

An assessment of the forecasting accuracy follows the same path as in [42]. First of all, 10000 different sequences of length 24 are generated for each model discussed in this section. In addition, all model parameters required for generating are estimated based on the training set. These generated sequences of predictions will be compared to the prediction set. Further, it is necessary to choose the criterion on basis of which the predictive capacity of proposed models will be analyzed. For that purpose, the forecasting log-score criterion (*FLSC*) is chosen. This criterion has already been introduced in [32], and represents an adaptation of the more famous log-score criterion (*LSC*) described in [17]. The criterion itself is given by the equality

$$FLSC = \sum_{k=1}^{n_2} \log \hat{p}_{n_1+k}(x_{n_1+k}),$$

where $\hat{p}_{n_1+k}(x_{n_1+k})$ denotes the estimated probability of correctly predicting the value x_{n_1+k} taken from the prediction set, i.e.,

$$\hat{p}_{n_1+k}(x_{n_1+k}) = \frac{\text{number of correct predictions of } x_{n_1+k}}{10000}.$$

Calculated values of the *FLSC* criterion for all considered models are provided in Table 3.5, whereby better forecasting provides the model with higher *FLSC*. According to the table, the $R2DLINAR_1(\mathcal{M}, \mathcal{A})$ model provides the most accurate forecasting, since it has the largest *FLSC* value among all considered models.

Table 3.5: *FLSC* criterion value for the prediction set of theft differences.

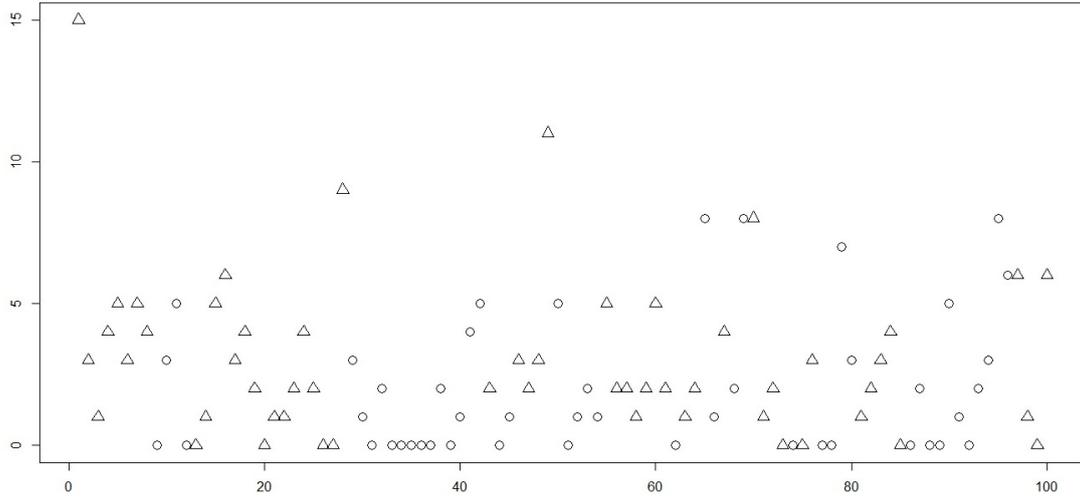
	<i>TINAR</i> (1)	<i>DLINAR</i> (1)	<i>STINAR</i> (1)	$R2DLINAR_1(\mathcal{M}, \mathcal{A})$
<i>FLSC</i>	-65.711	-63.981	-64.150	-63.135

Chapter 4

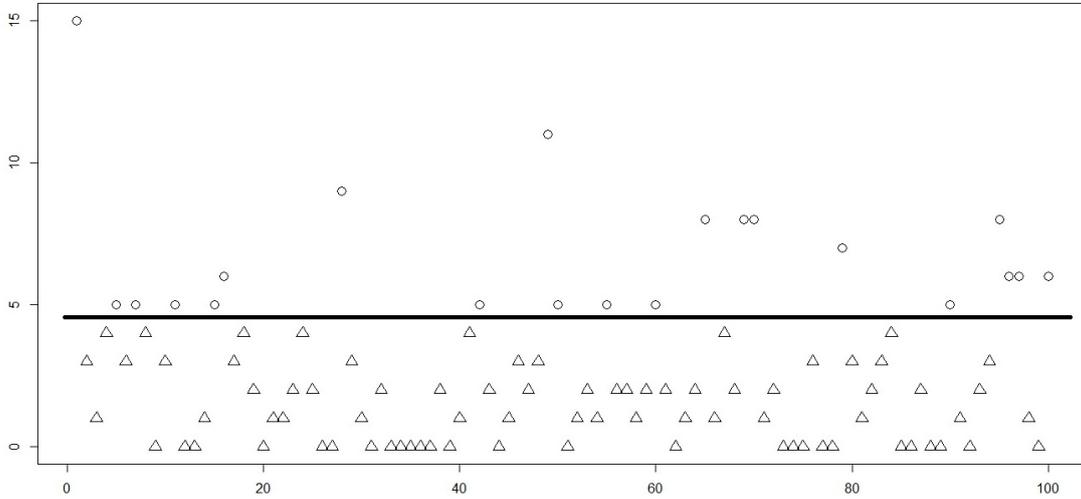
Random Environment Estimation (*RENES*) Method for Generalized Random Environment *INAR* Models of Higher Order

As mentioned in the thesis title, the focus of this dissertation is on *INAR* models based on the random environment process. Some of the random environment *INAR* models, as well as the random environment process itself, have already been presented in Chapter 1. In order to model given real-life data using this kind of models and to measure the goodness of fit of such obtained predictions, corresponding environment states $\{z_n\}$ for all real-life observations must be estimated. As noticed in Section 1.2.8, this is where the clustering methods take place. Clustering methods are indispensable when it's necessary to classify the data points into several disjoint sets, so that sufficiently similar data points belong to the same set. Each set (cluster) is observed as a specific state. Among several clustering techniques available so far, the K-means was the one that authors preferred to use.

The K-means clustering method was first introduced in [18]. After the *RrNGINAR*(1) time series was defined in [38], the method usually gave satisfactory results in estimating environment states of the *RrNGINAR*(1) model. This was especially the case when realizations within the same environment state were sufficiently similar. However, some shortcomings appear in situations when realizations within the same environment state deviate significantly from each other. This happens due to the fact that K-means takes into account only realization values. Namely, once the K-means is performed, one can divide graphed representation of the database into strips by several horizontal lines, as given in the lower panel of Figure 4.1. The number of strips corresponds to the number of clusters one wants to create. All realization values in the same strip are similar to each other and are considered to be in the same cluster, i.e. in the same environment state. This particularity entails that all high values in the database belong to the same cluster. Similar to this, all low values in the database belong to the same cluster as well. However, it doesn't have to be necessarily the case in reality. In the upper panel of Figure 4.1, the data simulated from the *R2NGINAR*(1) model is presented. As may be seen, it is possible for high realizations to appear also in environment conditions different than those assumed for high data values, so the database can no longer be divided by a horizontal



Exact environment states of the simulated $R2NGINAR(1)$ time series



States estimated by the application of standard K-means clustering method

Figure 4.1: Environment states of the $R2NGINAR(1)$ simulation: different states presented with different symbols—a circle or a triangle. The estimated states are obviously divided by a horizontal strip, unlike the exact states.

line. The K-means method totally rules out this possibility. To clarify additionally the mentioned disadvantage, a vivid example follows.

COVID-19 pandemic led to the greatest global crisis of the modern age. Despite the strict and harsh measures conducted by many nations to restrict the spread of the virus, virus continued spreading all over the globe. Based on experiences from all continents, weather conditions significantly affected the spread rate. In summertime, the presence of UV light and high temperatures created an environment unsuitable for virus spreading, and the low number of newly registered cases was appearing day after day. In autumn, weather conditions have worsened. Hence, the increment in number of newly registered cases on daily basis showed up. The situation changed dramatically during winter. Low temperatures and the lack of sunlight induced the perfect environment for virus spreading, and the number of newly registered cases kept increasing. Nevertheless, some deviations were detected. For instance, in many regions, short time intervals with unusually high

number of newly registered cases on daily basis appeared during the summer months. This might be so confusing, actually. The discrepancies can be explained by the fact that not only weather conditions affect the daily number of newly registered COVID-19 cases, but some other undetected circumstances as well, or circumstances which are detected but not measurable. In situations like these, the proper clustering method is expected to show additional flexibility. It has to take into account undetected circumstances and keep high values (noticed in summer months) in 'summer' cluster. This is a difficult task for K-means. By taking into account only numerical values of the data, K-means will recognize high summertime values as autumn (or even winter) occasions and locate them into the wrong cluster. None of the K-means adaptations familiar so far is capable to help either. An improved estimation method is obviously needed in this case.

After more complex random environment *INAR* models had emerged, the list of problems that K-means encountered became even longer. The following difficulty is particularly related to generalized random environment *INAR* models of higher order, defined by (1.9). To explain this difficulty, the simplest case with $r = 2$ states is observed and the similarity between mean values within states is supposed (i.e. $\mu_1 \approx \mu_2$), while other model parameters differ significantly. In that case, observations within states are not that much different and are accumulated around parallel horizontal lines that are close to each other. In situation like this, one might expect the existence of a strip in which points from both environment states will be represented. In other words, the border line between states won't be straight, but a jagged and wavy. Keeping in mind the fact that K-means method separates clusters by straight horizontal lines, it becomes obvious that the method itself is pretty much incapable to perform a proper separation between states. However, by including all parameters of the model that carry information about the environment state in the clustering process, this problem can be overcome.

A new random environment estimation (abbrev. *RENES*) method, based on transformation (before applying clustering) of the data sample that corresponds to the generalized random environment *INAR* model of higher order, will be presented in this chapter. The chapter contains results given in [41]. Although the new approach is heavily relied on K-means method, suggested transformation is going to eliminate disadvantages given above. It should be emphasized here that the *RENES* method can be applied to the data corresponding to any generalized random environment *INAR* model of higher order, with an arbitrary marginal distribution and a thinning operator that is not necessarily a negative binomial. Nevertheless, this chapter provides an application of the newly defined *RENES* method only to the data that correspond to $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models, defined by (1.9). For more information about $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models, see Section 1.2.9. As one may notice from the definition of $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$, each of μ_{z_n} , α_{z_n} and P_n carries the information about environment state in which the realization x_n took place, since these parameters directly participate in model construction. To eliminate the information loss, the main goal is to create a three-dimensional sequence based on real-life data realizations, that mimics the behavior of $\{(\mu_{z_n}, \alpha_{z_n}, P_n)\}_{n=1}^{\infty}$. Finally, the K-means algorithm will be applied on such obtained three-dimensional data sequence. To avoid the confusion, it is important to highlight here that the *RENES* method doesn't represent a tool for estimating the parameters of $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models, but only a new tool for estimating $\{z_n\}$. However, estimators of the $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ model parameters are defined under the assumption that $\{z_n\}$ is known in advance, meaning that a differ-

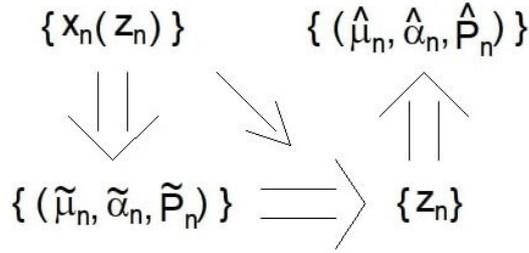


Figure 4.2: Illustration of the role the 'pre-estimates' have in estimating model parameters. Namely, for a given data $\{x_n(z_n)\}$ the first step is to create $\{(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)\}$. After that, an estimate of $\{z_n\}$ is provided by K-means clustering of the three-dimensional sequence $\{(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)\}$. Having estimated $\{z_n\}$, one obtains the estimates $\{(\hat{\mu}_n, \hat{\alpha}_n, \hat{P}_n)\}$ of model parameters. In the previous approach used in [29], K-means was applied directly on $\{x_n(z_n)\}$ (shown by a diagonal arrow in diagram) instead of creating pre-estimates.

ent approach for estimating $\{z_n\}$ will for sure imply the difference in parameter estimates.

The chapter starts with construction of the new *RENES* method, which overcomes disadvantages mentioned above. After that, suitable simulated data series are created. Observing the simulations, one may examine whether the changes in number of states and parameter values affect the efficiency of the newly proposed method. The *RENES* method is applied on simulated data and the results such obtained are compared with those obtained by usage of the standard K-means method. Finally, the *RENES* method is applied to the popular real-life data, alongside with K-means. In that way, two different clustering results are obtained. To confirm the supremacy of the *RENES* method, a fitting quality of corresponding $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models is examined for each clustering result.

4.1 Construction of the new *RENES* method

In order to construct the *RENES* method, a sample $\{X_n\} = \{X_n(z_n)\}$ of size $N \in \mathbb{N}$ that corresponds to the $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series is considered. As noticed in [41], the main idea is to provide some kind of 'pre-estimators' $\{\tilde{\mu}_{z_n}\}$, $\{\tilde{\alpha}_{z_n}\}$ and $\{\tilde{P}_n\}$ of parameter sequences $\{\mu_{z_n}\}$, $\{\alpha_{z_n}\}$ and $\{P_n\}$, by taking into account only on the realized sample, without knowing the realizations $\{z_n\}$ of the random environment process. The three-dimensional sequence $\{(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)\}$ such obtained is supposed to mimic the behavior of the model parameters over time. Then, clustering the three-dimensional data $\{(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)\}$ would prevent the information loss and produce better estimation of $\{z_n\}$, compared to clustering the starting sequence $\{x_n\}$.

As mentioned before, the research goal is not to introduce new estimators of model parameters, but to upgrade the estimation of $\{z_n\}$. Given sequence of so-called 'pre-estimators' $\{(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)\}$ is just a helpful tool to estimate $\{z_n\}$, and it does not represent an alternative estimate of model parameters. Model parameters have already been successfully estimated in [29], and those results are used here as well in evaluating the *RENES* method. Figure 4.2 provides an illustration of the role the 'pre-estimates' $\{(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)\}$ have in estimating model parameters.

Proposed method can be additionally improved by considering trimmed (truncated) means. Hence, for a given vector $\mathbf{v} = (v_0, v_1, \dots, v_k)'$ and sequence a_1, a_2, \dots, a_N , let the function $T(a_n, \mathbf{v})$ be defined as

$$(4.1) \quad T(a_n, \mathbf{v}) = \begin{cases} a_n, & n \leq k \text{ or } n > N - k, \\ \sum_{l=n-k}^{n+k} v_{|l-n|} a_l, & k < n \leq N - k, \end{cases}$$

whereby $N > 2k$. Coordinates of the vector \mathbf{v} are supposed to be non-increasing positive real numbers, i.e. $v_0 \geq v_1 \geq \dots \geq v_k$, $v_j > 0$, $j = 0, 1, \dots, k$. Beside this, the condition $v_0 + 2 \sum_{j=1}^k v_j = 1$ must be satisfied. The trimmed mean $T(a_n, \mathbf{v})$ thus obtained is affected the most by the current value a_n . Regarding the effect of the k neighboring elements of a_n on both sides, it decreases when moving away from a_n . As mentioned in [29], *RrNGINAR*($\mathcal{M}, \mathcal{A}, \mathcal{P}$) models demonstrate poor performances when environment states are changing frequently. Their application make sense only if the probability of remaining in the same state is high enough. In particular, for each $q, s \in E_r = \{1, 2, \dots, r\}$, the probability $P(Z_{n+1} = q | Z_n = q)$ is supposed to be greater than the probability $P(Z_{n+1} = s | Z_n = q)$. It is even recommended that $P(Z_{n+1} = q | Z_n = q) > \sum_{s \neq q} P(Z_{n+1} = s | Z_n = q)$. The higher values of $P(Z_{n+1} = q | Z_n = q) - \sum_{s \neq q} P(Z_{n+1} = s | Z_n = q)$ imply better model application. If the last inequality is satisfied, then for k small enough one may assume that all $2k + 1$ neighboring elements of z_n are equal with high probability in most situations, i.e. $z_{n-k} = \dots = z_n = \dots = z_{n+k}$. Hence, all elements $X_{n-k}, \dots, X_n, \dots, X_{n+k}$ correspond to the same state z_n and thus, all of them carry information about z_n . As a consequence, all pre-estimates $\{(\tilde{\mu}_l, \tilde{\alpha}_l, \tilde{P}_l)\}_{l=n-k}^{n+k}$ are supposed to carry information about z_n as well. By combining them (using the trimmed mean function T), it is expected to get even more accurate information. In other words, it sounds reasonable to replace $\{\tilde{\mu}_n\}$, $\{\tilde{\alpha}_n\}$ and $\{\tilde{P}_n\}$ with $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$, $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ and $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for some vectors \mathbf{v}_m , \mathbf{v}_a and \mathbf{v}_p . Theoretically speaking, the mentioned vectors do not have to be the same length. The upper limit of the vector's length k might also be the subject of discussion. Certainly, higher values of k give better pre-estimates, provided all of observations $X_{n-k}, \dots, X_n, \dots, X_{n+k}$ correspond to the same state. Otherwise, pre-estimates might be even worsened. Namely, although higher values of k provide some benefit when a particular observation is surrounded by sufficient number of elements in the same state, they will do a lot of damage when observations are located near the state change. In order to reconcile these two opposing claims, vectors \mathbf{v}_m , \mathbf{v}_a , \mathbf{v}_p of length higher than 4 won't be discussed at all.

In order to equalize the impact of each particular coordinate of the three-dimensional vector $(T(\tilde{\mu}_n, \mathbf{v}_m), T(\tilde{\alpha}_n, \mathbf{v}_a), T(\tilde{P}_n, \mathbf{v}_p))$ on clustering procedure, it is necessary to scale the coordinates. For that purpose, for a given element a_n of the sequence $\{a_n\}$, a function $S(a_n, \mathbf{v})$ defined as

$$S(a_n, \mathbf{v}) = \frac{T(a_n, \mathbf{v}) \cdot N}{\sum_{l=1}^N T(a_l, \mathbf{v})}$$

assigns to a_n the properly scaled (normed) value of $T(a_n, \mathbf{v})$. Furthermore, to control the level of impact each coordinate has on the clustering procedure, three more parameters $C_m, C_a, C_p \in \mathbb{R}$ have to be introduced. The *RENES* procedure ends with clustering the three-dimensional data vector

$$(4.2) \quad (C_m S(\tilde{\mu}_n, \mathbf{v}_m), C_a S(\tilde{\alpha}_n, \mathbf{v}_a), C_p S(\tilde{P}_n, \mathbf{v}_p))$$

using standard K-means. At this point, defining the starting estimators $(\tilde{\mu}_n, \tilde{\alpha}_n, \tilde{P}_n)$, $n = 1, 2, \dots, N$, in a reasonable way is the only left to do, taking into account the construction of $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models. To determine values of the sequence $\{\tilde{\mu}_n\}$, the idea already exploited in [30] is allowed to be used. Namely, keeping in mind that μ_s , $s = 1, 2, \dots, r$, represent means within clusters, one is available to set for any $n = 1, 2, \dots, N$ that

$$(4.3) \quad \tilde{\mu}_n = X_n.$$

A fact that the partial autocorrelation function represents a useful tool for determining the order of the time series gives a unique opportunity to predict the sequence $\{P_n\}$. Let p_{z_n} be a maximal order allowed for particular element in the state z_n . In this case, one may set

$$(4.4) \quad \tilde{P}_n = \begin{cases} \max_{K=1, \dots, p_{z_n}} pacf_K(X_1, \dots, X_{2d_p+1}), & n \leq d_p, \\ \max_{K=1, \dots, p_{z_n}} pacf_K(X_{n-d_p}, \dots, X_{n+d_p}), & d_p < n \leq N - d_p, \\ \max_{K=1, \dots, p_{z_n}} pacf_K(X_{N-2d_p}, \dots, X_N), & n > N - d_p, \end{cases}$$

where $d_p \in \mathbb{N}$ and $pacf_K$ represents the partial autocorrelation function at lag K . To maximize the accuracy of the pre-estimate \tilde{P}_n , it would be ideal if all $2d_p + 1$ elements of the sequence $\{X_n\}$ involved in \tilde{P}_n correspond to the same state z_n . However, this requirement is not demanding, since an application of the $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ model is reasonable only when the state change is an infrequent occasion.

A prediction of the thinning parameter value in moment n is heavily relied on the well known property of the negative binomial thinning operator $E(\alpha * X|X) = \alpha X$. By looking at (1.9), one may define α_n^* as

$$\alpha_n^* = \begin{cases} Q_n/R_n, & R_n \neq 0, \quad n > 1, \\ 1, & Q_n = R_n = 0, \quad n > 1, \\ \max \left\{ \left(\frac{Q_l}{R_l} : l \in \{2, \dots, N\}, R_l > 0 \right) \right\}, & otherwise, \end{cases}$$

for all $n \in \mathbb{N}$, where $Q_n = (x_n - T(\tilde{\mu}_n, \mathbf{v}_m))_+$ and $R_n = \frac{1}{b} \sum_{l=1}^b Q_l$ for $b = \min\{n-1, \tilde{P}_n\}$. Notation $(x)_+ = \max\{x, 0\}$ represents the positive part of $x \in \mathbb{R}$. In order to avoid the thinning parameter value greater than 1, one more step must be performed as follows:

$$(4.5) \quad \tilde{\alpha}_n = \frac{\alpha_n^*}{\max_{l=1, \dots, N} \alpha_l^*}, \quad n \in N.$$

An application of the *RENES* method implies an optimal choice of parameters d_p , \mathbf{v}_m , \mathbf{v}_a , \mathbf{v}_p , C_m , C_a and C_p (called in sequel the *RENES* method parameters) based on $RrNGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$ and $RrNGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations. It is crucial to distinguish the *RENES* method parameters from the model parameters. Section 4.2 gives details about the choice of model parameters. On the other hand, Section 4.3 provides results of the simulation study with such chosen model parameters and discuss how to choose *RENES* method parameters.

4.2 Simulation study—the choice of model parameters

In order to test the new *RENES* method for estimating the random environment states, simulated $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series of length $N = 500$ are provided. Simulation properties are such that they disable smooth application of the standard K-means method. Precisely, the time series with similar means within states are created, while some other model parameters differ significantly. The case with $r = 2$ different environment states can be found within this section. In addition, the case with $r = 3$ environment states is given in Appendix A. Further, two different combinations of model parameters are analyzed for each of the cases. Further more, each parameter combination is supposed to generate two different replications of the corresponding $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series. One of those will be used to provide optimal values of the *RENES* method parameters. With the help of such obtained $d_p, \mathbf{v}_m, \mathbf{v}_a, \mathbf{v}_p, C_m, C_a$ and C_p , the other replication will be reconstructed in order to evaluate the efficiency of the proposed *RENES* method. Finally, both versions of the model analyzed in Subsection 1.2.9, $RrNGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$ and $RrNGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$, will be discussed simultaneously.

Here are a few notes regarding the notation used. Being a Markov chain, the random environment process is basically depended on two factors. One of them is a vector containing initial probabilities of being in certain state, denoted as p_{vec} . Another is a transition probability matrix, denoted as p_{mat} , which contains the probability $P(Z_n = s | Z_{n-1} = q)$ in the intersection of its q -th row and s -th column, for all $q, s \in \{1, \dots, r\}$. One more remark regarding the notation is needed here. Although introduced as sets, \mathcal{M}, \mathcal{A} and \mathcal{P} are all written down as vectors in the following text. This has been done to eliminate the ambiguity and preserve the order of the states.

To create appropriate $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations, the following combinations of parameters are used.

1. First of all, a similarity of means within states is assumed, while other model parameters are significantly different. Surrounding like this would make K-means pretty much useless. Thus, $\mathcal{M} = (1, 1.5)$ is chosen. On the other hand, thinning parameters and maximal orders within states should differ significantly. Hence, $\mathcal{A} = (0.05, 0.6)$ and $\mathcal{P} = (2, 4)$ are selected. Bearing in mind the upper limits existence for parameters $\alpha_s, s = 1, 2$, the first one is chosen to be small, while the second one is selected to be close to its upper limit. Probabilities $\phi_{j,l}^s$ that correspond to the $R2NGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulation are contained in the following matrices:

$$\phi_1 = \begin{bmatrix} 1 & 0 \\ 0.9 & 0.1 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.1 & 0.9 & 0 & 0 \\ 0.1 & 0.45 & 0.45 & 0 \\ 0.1 & 0.1 & 0.4 & 0.4 \end{bmatrix}.$$

In addition, probabilities that correspond to the $R2NGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulation are located in last rows of aforementioned matrices ϕ_1 and ϕ_2 . Finally, an initial state is nearly fair, due to the value of its distribution $p_{vec} = (0.6, 0.4)$. In order to preserve the simulated $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series in one state as long as possible, the current state of the random environment process is preferred. In other

words, transition probabilities out of the main diagonal are significantly smaller than those located on the main diagonal. Thus,

$$p_{mat} = \begin{bmatrix} 0.9 & 0.1 \\ 0.2 & 0.8 \end{bmatrix}.$$

2. A great similarity between thinning parameters characterizes the second parameter combination. Therewithal, the means are similar enough to make an application of the standard K-means method difficult. A valid environment states estimation will only be possible based on the values of model orders, which represents a great test for introduced *RENES* method. Hence, the following parameter values are chosen: $\mathcal{M} = (3, 5)$, $\mathcal{A} = (0.4, 0.5)$ and $\mathcal{P} = (2, 5)$. Further, in the case of $R2NGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$,

$$\phi_1 = \begin{bmatrix} 1 & 0 \\ 0.4 & 0.6 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 & 0 \\ 0.4 & 0.4 & 0.2 & 0 & 0 \\ 0.3 & 0.3 & 0.3 & 0.1 & 0 \\ 0.4 & 0.2 & 0.2 & 0.1 & 0.1 \end{bmatrix}.$$

Corresponding probabilities for the $R2NGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulation are contained in last rows of these matrices. Since $p_{vec} = (0.5, 0.5)$, one may claim the initial state is fair. The transition probability matrix provides long arrays of elements corresponding to the same state, since $P(Z_{n+1} = q | Z_n = q) > P(Z_{n+1} = s | Z_n = q)$ when $q, s = 1, 2, q \neq s$. Precisely,

$$p_{mat} = \begin{bmatrix} 0.8 & 0.2 \\ 0.25 & 0.75 \end{bmatrix}.$$

4.3 Simulation study — simulation results and selection of the *RENES* method parameters

The section offers a procedure for obtaining optimal values of the *RENES* method parameters, based on $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations with $r = 2$ environment states. The procedure follows the steps given in [41] and is relied on corresponding model parameters given in previous section. To improve the readability of the following consideration, a procedure of obtaining $d_p, \mathbf{v}_m, \mathbf{v}_a, \mathbf{v}_p, C_m, C_a$ and C_p will be fully exposed only for the first parameter combination. Regarding the second combination, the procedure will be omitted and only final results will be provided. As for the $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations with $r = 3$ environment states, corresponding discussion regarding the optimal *RENES* method parameters is given in Appendix B.

To start the procedure, two replications of each of the corresponding $R2NGINAR_{max}(2, 4)$ and $R2NGINAR_1(2, 4)$ simulations were created using the first combination of model parameters. The first replication of each pair is used to obtain the unknown *RENES* method parameters. For that purpose, one can determine the sequence $\{\tilde{\mu}_n\}$ as suggested in (4.3). To improve pre-estimates such obtained, a proper selection of the vector \mathbf{v}_m is required. According to previous assumption that all $X_{n-k}, \dots, X_n, \dots, X_{n+k}$ correspond to the same state for k small enough, all pre-estimates $\tilde{\mu}_{n-k}, \dots, \tilde{\mu}_n, \dots, \tilde{\mu}_{n+k}$ can have a

similar contribution to $T(\tilde{\mu}_n, \mathbf{v}_m)$. Hence, coordinates of the vector \mathbf{v}_m are chosen to be as equal as possible. Eventually, v_0 might be a bit higher, since it multiplies the middle realization x_n . Sequences of pre-estimates $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ obtained for various selections of \mathbf{v}_m are shown in Figure 4.3 and Figure 4.4. The sequence of exact means $\{\mu_n\}$ was used as a benchmark. To increase the readability of the plot, only the first 200 elements are displayed.

As figures show, the usage of the vector \mathbf{v}_m managed to improve the accuracy of pre-estimates of the sequence $\{\mu_n\}$ in both cases. With help of this technique, peaks that deviate significantly from exact mean values have been trimmed. Foremost, in the case of $R2NGINAR_{max}(2, 4)$ simulation, the best result is obtained for $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$. Speaking of $R2NGINAR_1(2, 4)$ simulation, equally good results are obtained for $\mathbf{v}_m = (0.2, 0.2, 0.2)$ and $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$, and both vectors are available to use. In particular, the second option is chosen for this research.

A determination of $\{\tilde{P}_n\}$ is contained of two steps. The procedure starts with determination of d_p , mentioned in (4.4). In order to obtain the optimal value of d_p , let Δ_p denotes the root mean square of differences between exact orders P_n , $n = 1, 2, \dots, 500$, and corresponding estimated order values \tilde{P}_n , $1, 2, \dots, 500$, provided by (4.4). The error Δ_p is calculated for various choices of d_p and results such obtained are given in Table 4.1. The smallest value of Δ_p implies the optimal value of parameter d_p . As one may conclude, in the case of $R2NGINAR_{max}(2, 4)$ simulation optimal value of d_p is 8 ($\Delta_p = 1.438$). Similarly, in the case of $R2NGINAR_1(2, 4)$ simulation optimal d_p value is 15 ($\Delta_p = 1.371$).

Table 4.1: Values of Δ_p for various selections of d_p .

$R2NGINAR_{max}(2, 4)$		$R2NGINAR_1(2, 4)$		$R2NGINAR_{max}(2, 5)$		$R2NGINAR_1(2, 5)$	
d_p	Δ_p	d_p	Δ_p	d_p	Δ_p	d_p	Δ_p
5	1.478	5	1.560	5	2.033	5	2.126
6	1.456	6	1.581	6	2.138	6	2.051
7	1.450	7	1.588	7	2.219	7	2.053
8	1.438	8	1.612	8	2.165	8	2.071
9	1.480	9	1.620	9	2.171	9	2.009
10	1.518	10	1.521	10	2.111	10	2.068
11	1.503	11	1.492	11	2.123	11	2.098
12	1.456	12	1.510	12	2.084	12	2.100
13	1.475	13	1.495	13	2.082	13	2.088
14	1.482	14	1.430	14	2.034	14	2.074
15	1.474	15	1.371	15	2.077	15	2.103
16	1.495	16	1.424	16	2.041	16	2.101
17	1.512	17	1.390	17	2.015	17	2.137
18	1.441	18	1.372	18	2.051	18	2.120
19	1.457	19	1.380	19	2.057	19	2.088
20	1.440	20	1.446	20	2.064	20	2.089

It is left in the second step to determine the corresponding vector \mathbf{v}_p , i.e. to determine its coordinates and its length k for fixed optimal value of d_p . Similarly as for

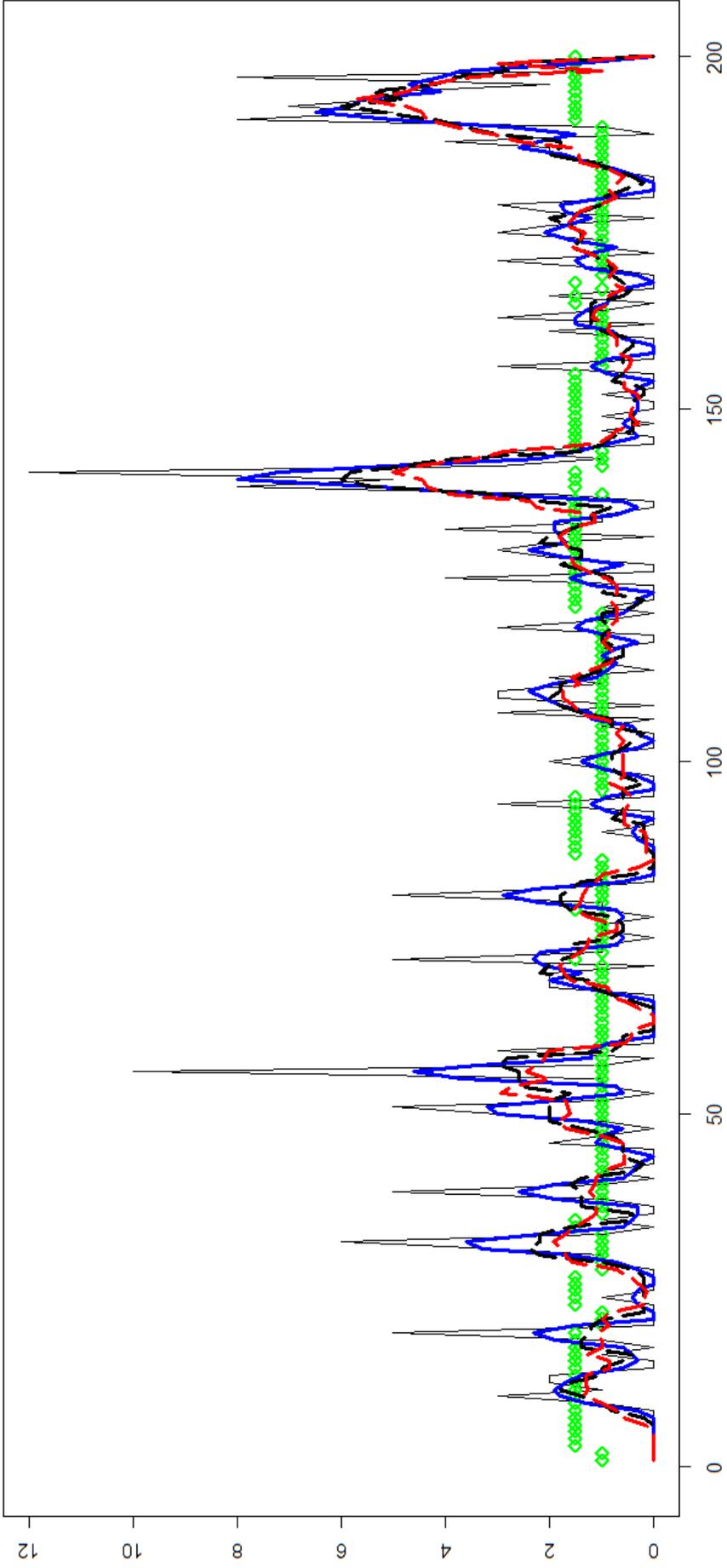


Figure 4.3: Corresponding pre-estimates of $\{\mu_n\}$ obtained for various selections of the vector \mathbf{v}_m in the case of simulated $R2NGINAR_{max}(2, 4)$ models: green – sequence of exact means $\{\mu_n\}$; black (regular) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = 1$; blue (thick) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$.

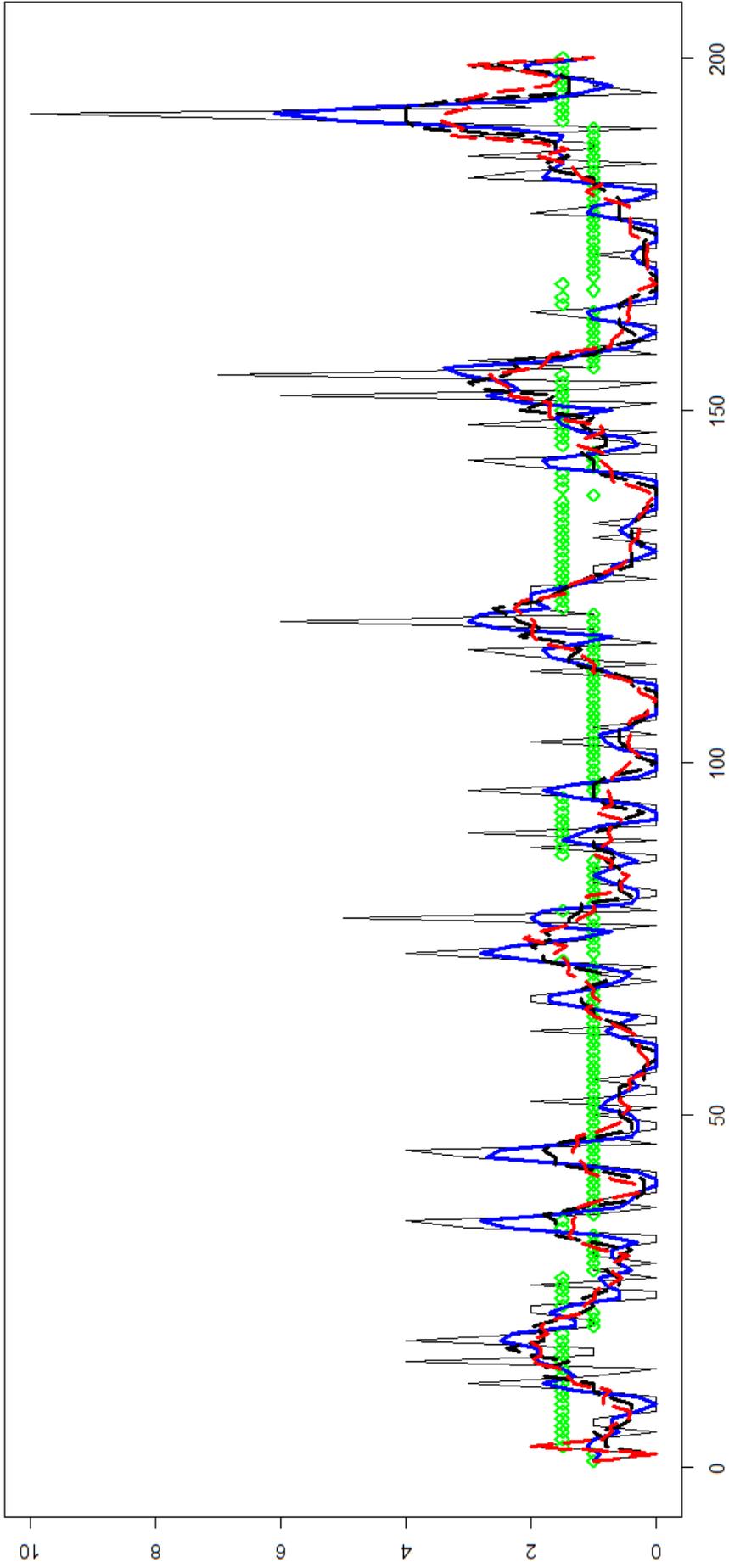


Figure 4.4: Corresponding pre-estimates of $\{\mu_n\}$ obtained for various selections of the vector \mathbf{v}_m in the case of simulated $R2NGINAR_1(2, 4)$ models – sequence of exact means $\{\mu_n\}$; black (regular) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = 1$; blue (thick) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$.

Table 4.2: Optimal values of the constant d_p and vectors \mathbf{v}_m , \mathbf{v}_a , \mathbf{v}_p , related to the simulated $R2NGINAR(2, 4)$ time series.

$R2NGINAR_{max}(2, 4)$			
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p
8	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)
$R2NGINAR_1(2, 4)$			
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p
15	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)

\mathbf{v}_m , $X_{n-k}, \dots, X_n, \dots, X_{n+k}$ are all assumed to correspond to the same state. Hence, $\tilde{P}_{n-k}, \dots, \tilde{P}_n, \dots, \tilde{P}_{n+k}$ are all assumed to have almost equal contribution to $T(\tilde{P}_n, \mathbf{v}_p)$. Thus, coordinates of \mathbf{v}_p are selected to be almost equal. Sequences of pre-estimates $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ obtained for various selections of \mathbf{v}_p are given in Figure 4.5 and Figure 4.6. The exact order sequence $\{P_n\}$ is observed as a benchmark.

Before an interpretation of these figures happens, one must be aware of the goal needs to be achieved. Namely, the primary goal is to obtain order pre-estimates which provide the highest probability of placing corresponding observations in correct clusters. More precisely, the best pre-estimate of $\{P_n\}$ is not necessarily the one that most often matches the exact order value, but the one that is close enough in most of the cases. Keeping this in mind, the following conclusion holds. Although the pre-estimates obtained for $k = 4$ (red dashed line) struggle to reach maximal orders, they stay close enough to exact order values in most of the cases and do not make large deviations. Hence, $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14)$ is taken in both cases, $R2NGINAR_{max}(2, 4)$ and $R2NGINAR_1(2, 4)$.

Having calculated $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ and $\{\tilde{P}_n\}$, one is able to calculate $\tilde{\alpha}_n$, $n = 1, 2, \dots, N$, as shown in (4.5). Such obtained pre-estimates are enhanced by appropriate selection of the vector \mathbf{v}_a . For the same reasons as before, coordinates of \mathbf{v}_a are chosen to be as similar as possible. Regarding the length of \mathbf{v}_a , various options are selected. Sequences $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ obtained for such chosen \mathbf{v}_a are given in Figure 4.7 and Figure 4.8. The sequence $\{\alpha_n\}$ of exact thinning parameters is used as a benchmark.

According to these figures, pre-estimates of $\{\alpha_n\}$ obtained for $\mathbf{v}_a = (0.2, 0.2, 0.2)$ and $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$ (dashed black and dashed red line) seem more accurate than those obtained for $\mathbf{v}_a = 1$ or $\mathbf{v}_a = (0.4, 0.3)$. More precisely, the sequences of pre-estimates obtained when $k = 3$ and $k = 4$ don't show sudden and sharp ups and downs frequently, while the most of their values stay in a strip between α_1 and α_2 . Behavior like this is actually expected for a fine sequence of pre-estimates. It is hard to choose the better one, but it seems that the plot line obtained for $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$ stays a bit closer to the real parameter values. The same conclusion holds for both, $R2NGINAR_{max}(2, 4)$ and $R2NGINAR_1(2, 4)$ simulation and thus, the same $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$ is selected in both cases.

To summarize all in one place, Table 4.2 provides values of d_p , \mathbf{v}_m , \mathbf{v}_a and \mathbf{v}_p involved in *RENES* method. One may find interesting that all three vectors \mathbf{v}_m , \mathbf{v}_a and \mathbf{v}_p are of the

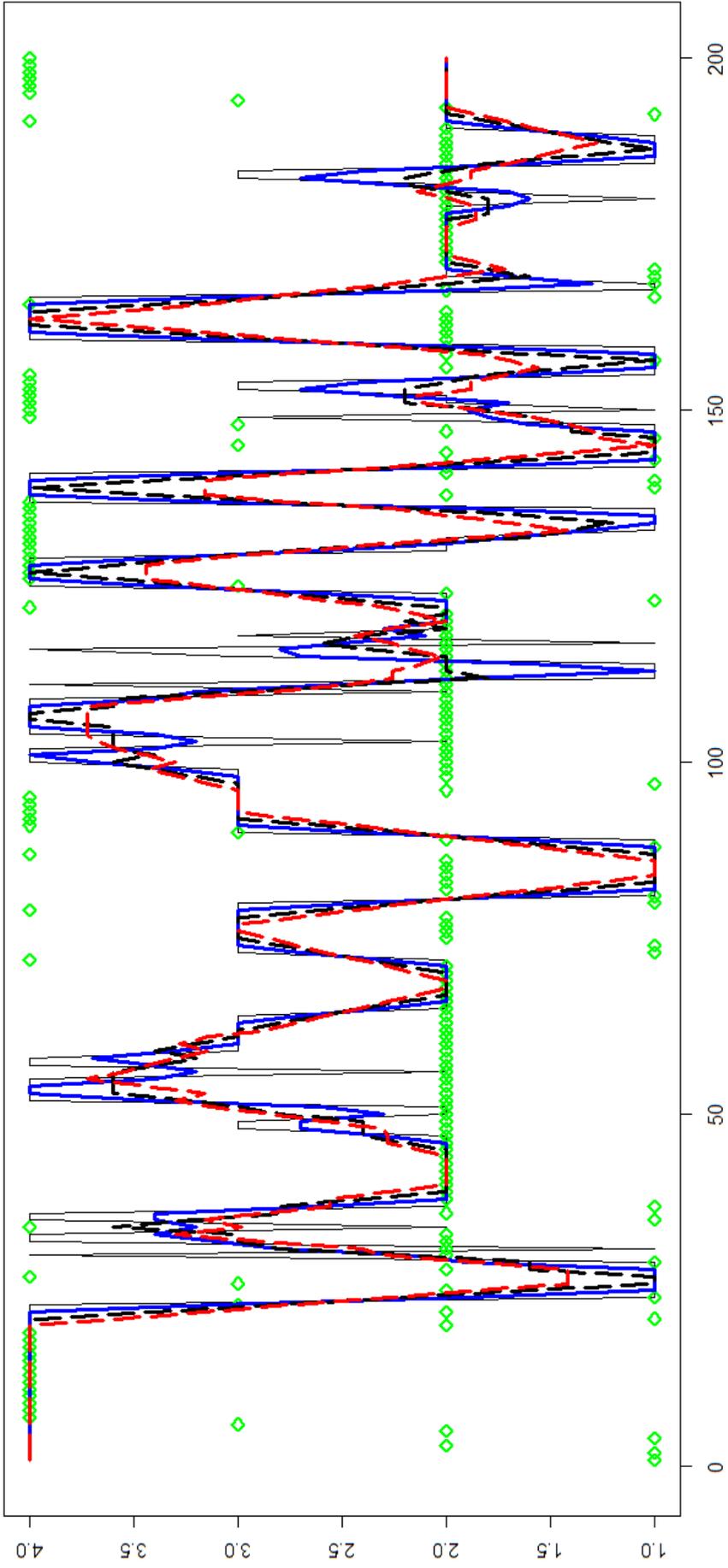


Figure 4.5: Corresponding pre-estimates of $\{P_n\}$ obtained for various selections of the vector \mathbf{v}_p in the case of simulated $R2NGINAR_{max}(2, 4)$ models: green – exact order sequence $\{P_n\}$; black (regular) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = 1$; blue (thick) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.4, 0.3)$; black (dashed) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14)$.

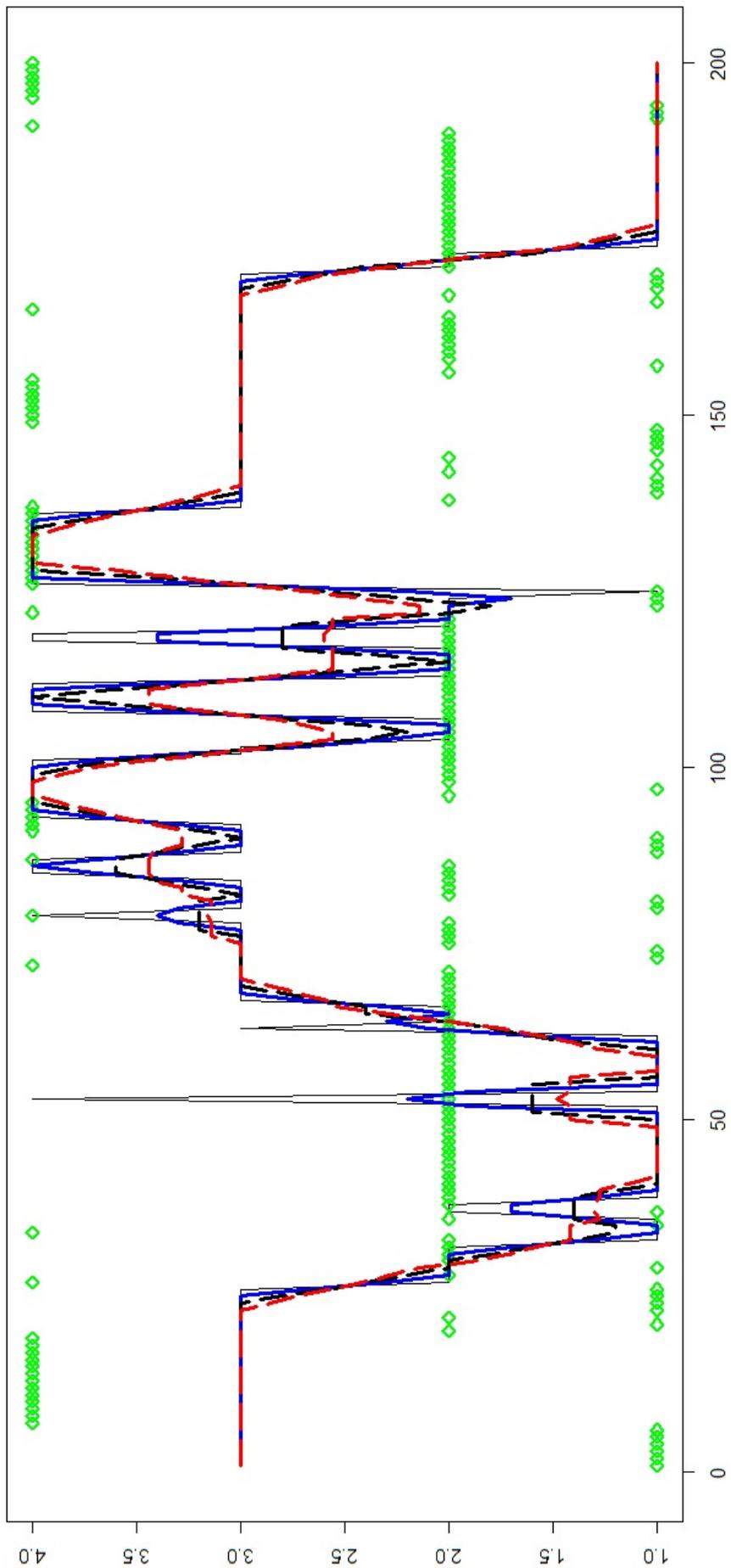


Figure 4.6: Corresponding pre-estimates of $\{P_n\}$ obtained for various selections of the vector \mathbf{v}_p in the case of simulated $R2NGINAR_1(2,4)$ models: green – exact order sequence $\{P_n\}$; black (regular) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = 1$; blue (thick) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14, 0.14)$.

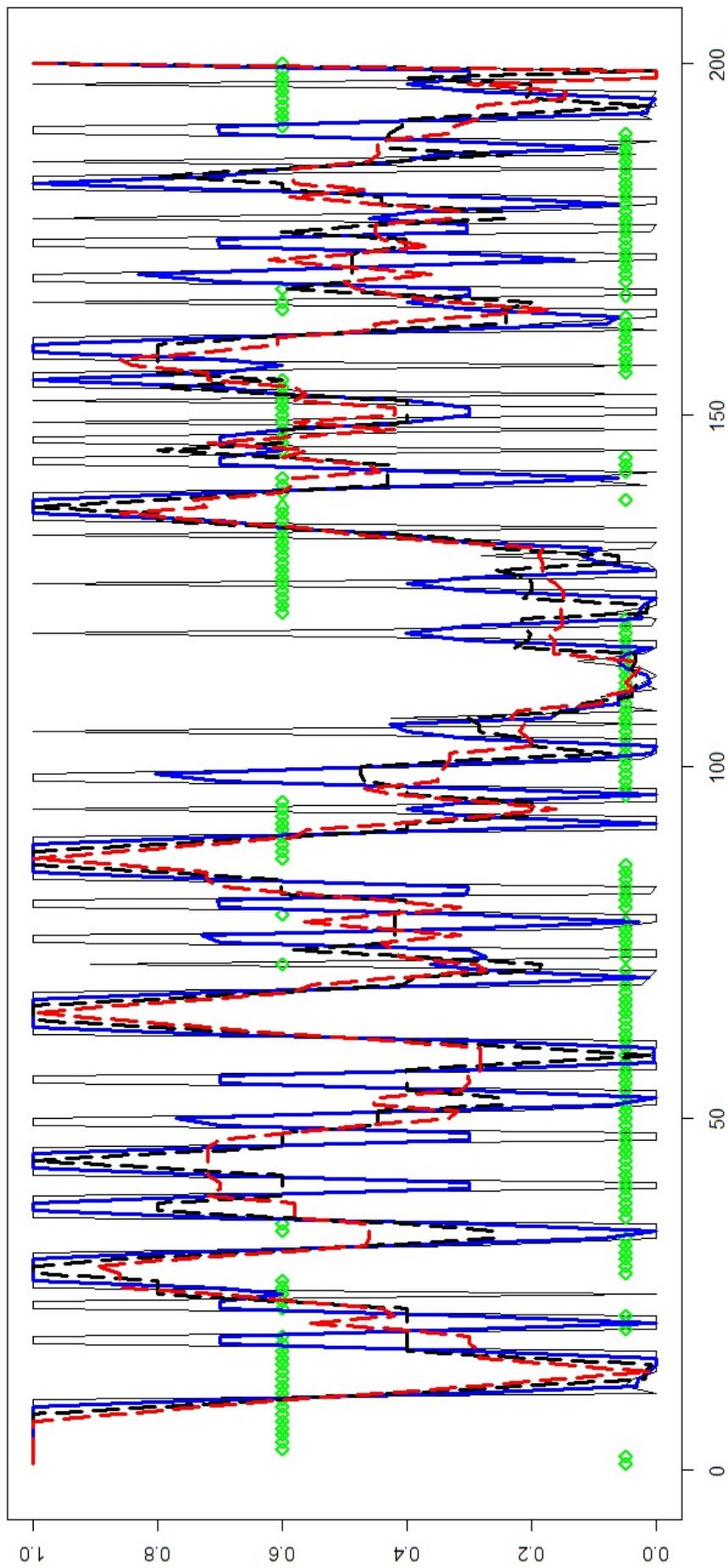


Figure 4.7: Corresponding pre-estimates of $\{\alpha_n\}$ obtained for various selections of the vector \mathbf{v}_a in the case of simulated $R2NGINAR_{max}(2, 4)$ models: green – exact thinning parameters sequence $\{\alpha_n\}$; black (regular) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = 1$; blue (thick) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$.

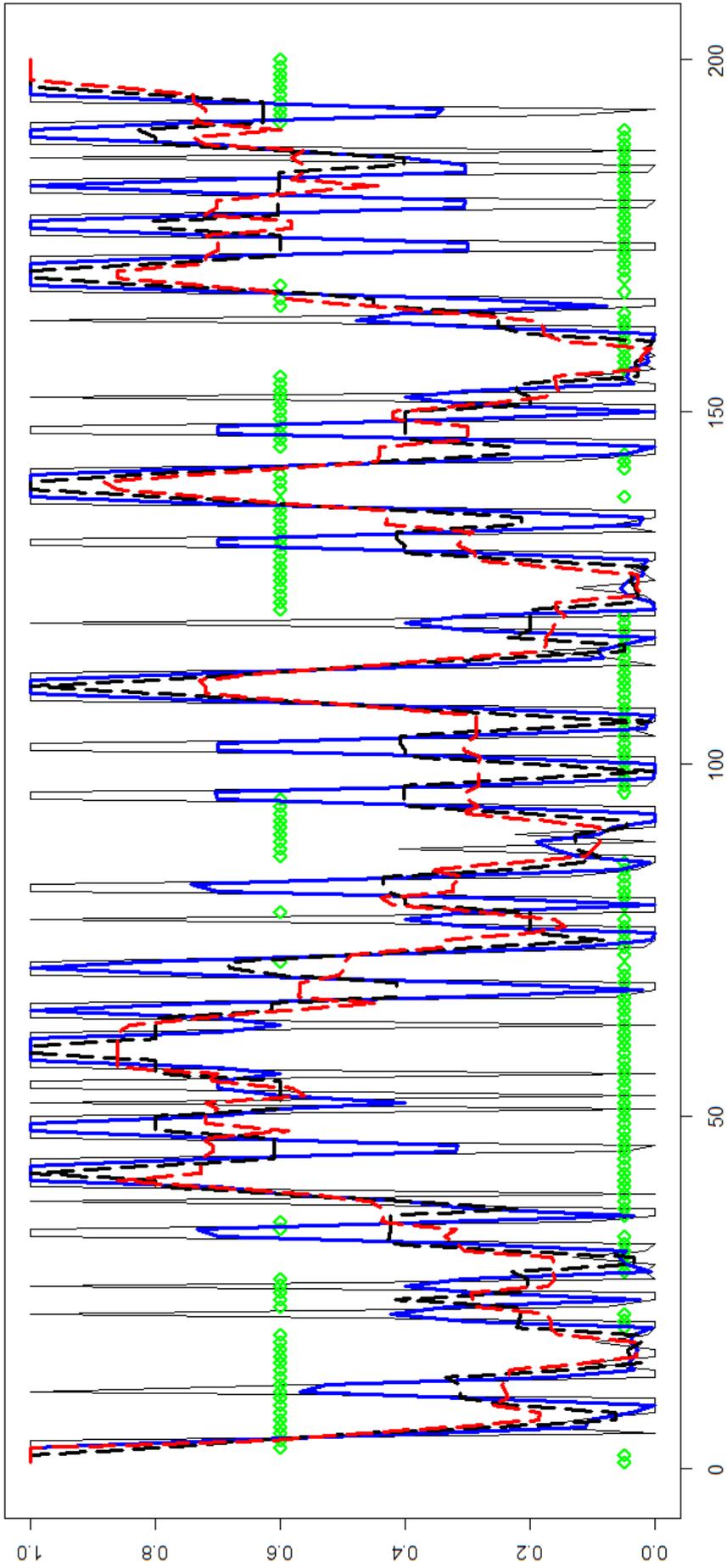


Figure 4.8: Corresponding pre-estimates of $\{\alpha_n\}$ obtained for various selections of the vector \mathbf{v}_a in the case of simulated $R2NGINAR_1(2, 4)$ models: green – exact thinning parameters sequence $\{\alpha_n\}$; black (regular) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = 1$; blue (thick) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$.

maximal length $k = 4$ in both cases. This is not so surprising, by the way. Recall that the transition probability matrix was set to have huge values on the main diagonal, much higher than those out of the main diagonal. Thus, long arrays of consecutive elements corresponding to the same state appeared, as anticipated. So, for most of pre-estimates, it was possible to enlarge the length of corresponding vectors without including elements from a different state. If the elements on the main diagonal had smaller values, shorten vectors would appear.

Now, one may provide 3-dimensional sequences $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}, \{T(\tilde{\alpha}_n, \mathbf{v}_a)\}, \{T(\tilde{P}_n, \mathbf{v}_p)\}$, and after that $\{S(\tilde{\mu}_n, \mathbf{v}_m)\}, \{S(\tilde{\alpha}_n, \mathbf{v}_a)\}, \{S(\tilde{P}_n, \mathbf{v}_p)\}$. Finally, it is left to determine parameters C_m, C_a and C_p mentioned in (4.2), by which one controls the level of impact each $\{S(\tilde{\mu}_n, \mathbf{v}_m)\}, \{S(\tilde{\alpha}_n, \mathbf{v}_a)\}, \{S(\tilde{P}_n, \mathbf{v}_p)\}$ has on the clustering procedure. For that cause, a modified procedure already used to determine d_p is applied. As given in [41], for each $C_m = i, C_a = j, C_p = l, i, j, l = 1, 2, \dots, 10$, a clustering of the three-dimensional data sequence

$$\{(C_m S(\tilde{\mu}_n, \mathbf{v}_m), C_a S(\tilde{\alpha}_n, \mathbf{v}_a), C_p S(\tilde{P}_n, \mathbf{v}_p))\}$$

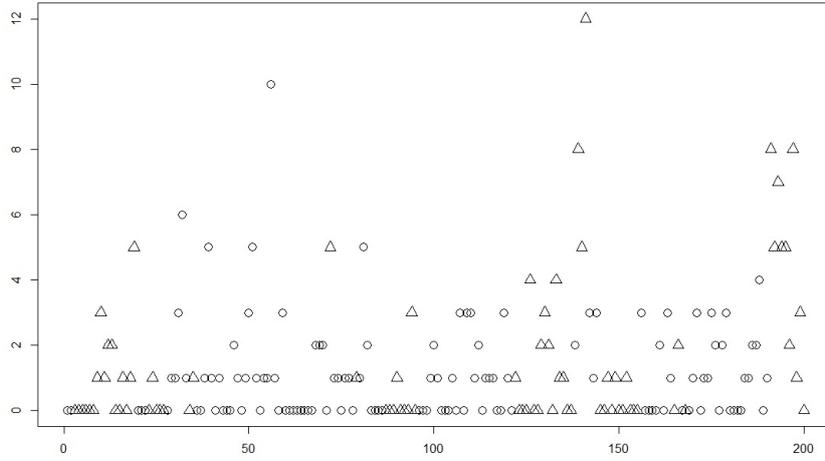
is performed. In this way, thousand different estimates of the environment state sequence $\{z_n\}$ are provided. To select the best one, estimates such obtained are compared to the sequence of exact states. The highest number of exactly estimated states will reveal the best combination of parameters C_m, C_a, C_p . In the case of $R2NGINAR_{max}(2, 4)$ simulation, the best result in random environment estimation is obtained for $C_m = 6, C_a = 2, C_p = 9$, having 328 estimated states equal to corresponding exact states. On the other hand, the result obtained by standard K-means managed to have 301 exactly estimated states. A comparative overview of exact states, states obtained by standard K-means and states obtained by usage of the *RENES* method is provided by Figure 4.9. Again, only the first 200 states are given in each graph.

Obviously, the *RENES* method gives better results in data clustering. Except the higher number of correctly estimated states, two more benefits are important to highlight. First of all, the newly proposed *RENES* method produces much longer data series that correspond to the same state. Keeping in mind that the random environment *INAR* models show poor performances when environment states are changing frequently, mentioned improvement seems convenient. The second, the newly proposed *RENES* method doesn't make a crisp data division by horizontal lines, which was the case with standard K-means method. The *RENES* method allows the data elements with low values to belong the cluster with predominantly high values, and vice versa. This property makes the method appropriate for clustering the data where, beside a detected predominant environment condition, some hidden circumstances also have an impact on time series realizations.

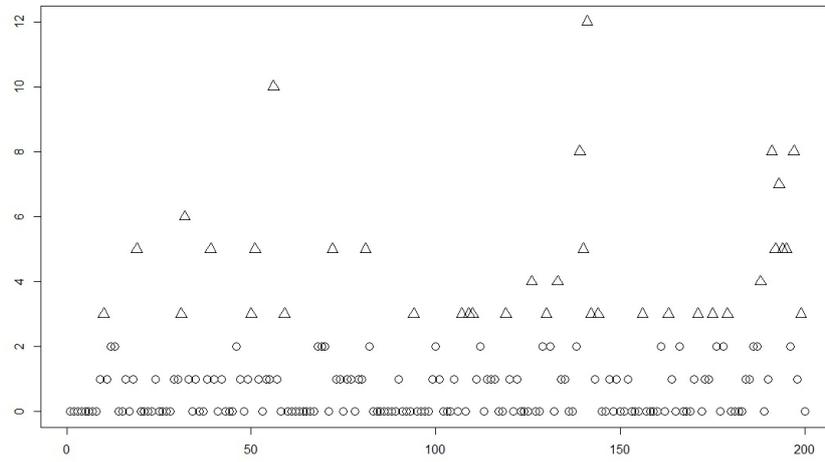
Similar conclusions hold in the case of $R2NGINAR_1(2, 4)$ simulation. After performing the clustering of three-dimensional data

$$\{(C_m S(\tilde{\mu}_n, \mathbf{v}_m), C_a S(\tilde{\alpha}_n, \mathbf{v}_a), C_p S(\tilde{P}_n, \mathbf{v}_p))\}$$

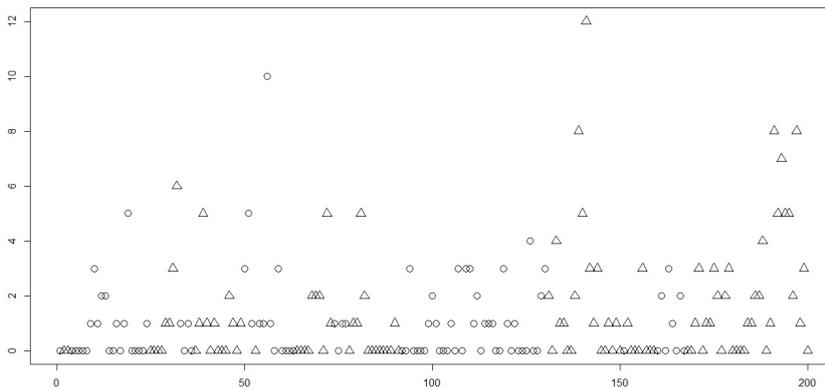
for each $C_m = i, C_a = j, C_p = l, i, j, l = 1, 2, \dots, 10$, the best result is obtained for $C_m = 8, C_a = 2, C_p = 3$. The *RENES* method with such chosen C_m, C_a and C_p provides 326 estimated states equal to corresponding exact states. Compared to this, the standard K-means managed to reach 309 correctly estimated states. Figure 4.10 provides a comparative overview of the exact states, states obtained by standard K-means method and



Exact states of the $R2NGINAR_{max}(2,4)$ simulation



States obtained by standard K-means clustering method



States obtained by *RENES* method for $d_p = 8$, $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14)$, $C_m = 6$, $C_a = 2$, $C_p = 9$.

Figure 4.9: Environment states of the $R2NGINAR_{max}(2,4)$ simulation.

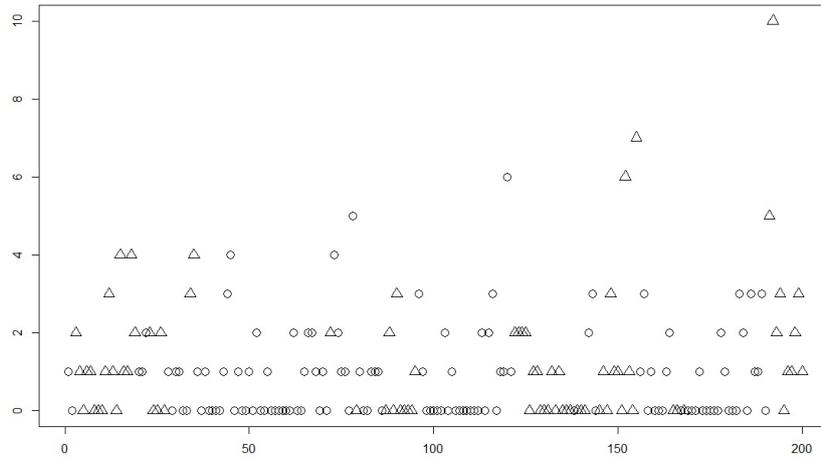
states obtained by newly introduced *RENES* method in the case of $R2NGINAR_1(2, 4)$ simulation. Figure undoubtedly confirms the dominance of the *RENES* method in regard to the standard K-means method. All benefits achieved with $R2NGINAR_{max}(2, 4)$ simulation are achieved here as well.

As shown in [41], the next goal is to confirm that the *RENES* clustering method is more appropriate for $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ model application than the standard K-means. Unused replications of $R2NGINAR_{max}(2, 4)$ and $R2NGINAR_1(2, 4)$ time series will be helpful for this purpose. These data are particularly convenient because they had not been exploited earlier to estimate the *RENES* method parameters. First of all, the standard K-means clustering method and the newly proposed *RENES* method are both applied on given unused replications. Further, the data reconstruction using the corresponding $R2NGINAR_{max}(2, 4)$ or $R2NGINAR_1(2, 4)$ model may happen for each clustering result, whereby corresponding parameters of the model are obtained by conditional maximum likelihood (*CML*) estimation procedure. The fitting quality is measured by calculating *RMS* of differences between simulated data and their reconstructions. Table 4.3 shows modeling results obtained after applying the standard K-means and the *RENES* clustering technique. Results confirm that the advantage of the *RENES* method is unquestionable. Namely, the usage of the standard K-means method leads to unexpectedly high *RMS* values ($RMS = 1.988$ for $R2NGINAR_{max}(2, 4)$ model and $RMS = 1.835$ for $R2NGINAR_1(2, 4)$ model). High *RMS* values certify an assumption presented in the introduction of this chapter which claims that K-means is not a useful tool for clustering the data corresponding to the $RrINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series with similar means within states. On contrary to that, the usage of the newly proposed *RENES* method leads to much more acceptable results ($RMS = 1.528$ for $R2NGINAR_{max}(2, 4)$ model and $RMS = 1.477$ for $R2NGINAR_1(2, 4)$ model).

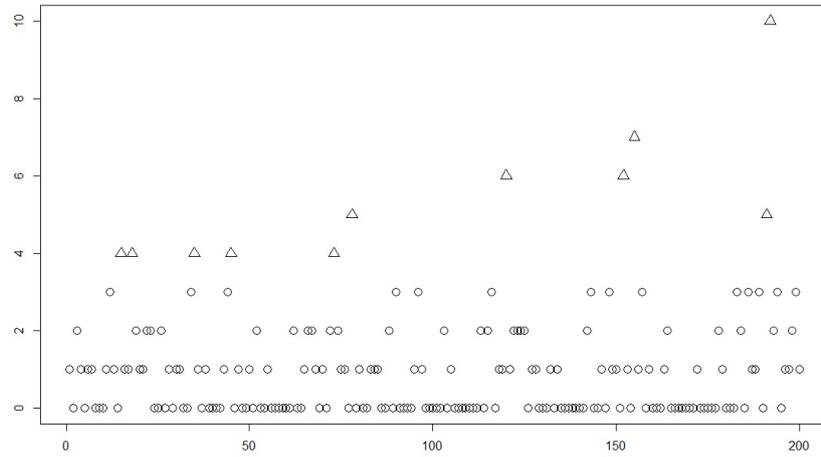
Table 4.3: *RMS* values and *CML* parameter estimates obtained after reconstruction of unused data sequences corresponding to the $R2NGINAR_{max}(2, 4)$ and $R2NGINAR_1(2, 4)$ time series.

	$R2NGINAR_{max}(2, 4)$		$R2NGINAR_1(2, 4)$	
Clustering	<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
Regular K-means	$\widehat{\mathcal{M}} = (0.543, 4.167)$ $\widehat{\mathcal{A}} = (0.001, 0.402)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.999 & 0.001 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 \\ 0.327 & 0.331 & 0.342 & 0 \\ 0.251 & 0.201 & 0.243 & 0.305 \end{bmatrix}$	1.988	$\widehat{\mathcal{M}} = (0.712, 5.434)$ $\widehat{\mathcal{A}} = (0.253, 0.385)$ $\widehat{\phi}_1 = (0.938, 0.062)$ $\widehat{\phi}_2 = (0.251, 0.200, 0.247, 0.302)$	1.835
<i>RENES</i>	$\widehat{\mathcal{M}} = (0.902, 1.588)$ $\widehat{\mathcal{A}} = (0.002, 0.308)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.999 & 0.001 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 \\ 0.328 & 0.330 & 0.342 & 0 \\ 0.246 & 0.200 & 0.241 & 0.313 \end{bmatrix}$	1.528	$\widehat{\mathcal{M}} = (0.930, 1.411)$ $\widehat{\mathcal{A}} = (0.172, 0.584)$ $\widehat{\phi}_1 = (0.951, 0.049)$ $\widehat{\phi}_2 = (0.243, 0.205, 0.230, 0.322)$	1.477

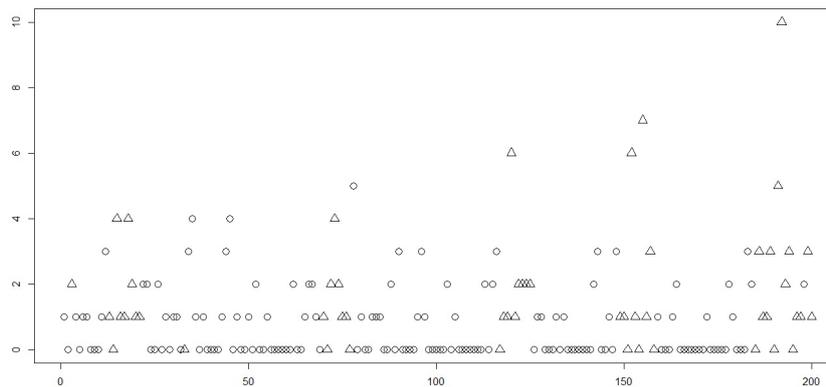
Regarding the second combination of model parameters, exactly the same steps are taken. First of all, corresponding $R2NGINAR_{max}(2, 5)$ and $R2NGINAR_1(2, 5)$ simulations are



Exact states of the $R2NGINAR_1(2, 4)$ simulation



States obtained by standard K-means clustering method



States obtained by *RENES* method for $d_p = 15$, $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14)$, $C_m = 8$, $C_a = 2$, $C_p = 3$.

Figure 4.10: Environment states of the $R2NGINAR_1(2, 4)$ simulation.

created, two replications of each. By applying the procedure equivalent to the one presented for $R2NGINAR(2, 4)$ simulations, optimal values of the $RENES$ method parameters are obtained from the first replication. Corresponding results can be found in Table 4.4. After the environment states have been estimated using both K-means method and $RENES$ method, unused replications are reconstructed for each clustering result using $R2NGINAR_{max}(2, 5)$ or $R2NGINAR_1(2, 5)$ model. The RMS -s of differences between simulated data and their reconstructions are provided in Table 4.5.

Table 4.4: Optimal values of the $RENES$ method parameters related to simulated $R2NGINAR(2, 5)$ time series.

$R2NGINAR_{max}(2, 5)$						
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p	C_m	C_a	C_p
17	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.4,0.3)	4	2	3
$R2NGINAR_1(2, 5)$						
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p	C_m	C_a	C_p
9	(0.2,0.2,0.2)	(0.16,0.14,0.14,0.14)	(0.4,0.3)	9	6	7

Table 4.5: RMS values and CML parameter estimates obtained after reconstruction of unused data sequences corresponding to the $R2NGINAR_{max}(2, 5)$ and $R2NGINAR_1(2, 5)$ time series.

Clustering	$R2NGINAR_{max}(2, 5)$	RMS	$R2NGINAR_1(2, 5)$	RMS
	CML		CML	
Regular K-means	$\widehat{\mathcal{M}} = (2.481, 14.169)$ $\widehat{\mathcal{A}} = (0.079, 0.132)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.007 & 0.993 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.002 & 0.998 & 0 & 0 & 0 \\ 0.398 & 0.401 & 0.201 & 0 & 0 \\ 0.298 & 0.301 & 0.301 & 0.100 & 0 \\ 0.197 & 0.198 & 0.200 & 0.200 & 0.205 \end{bmatrix}$	4.029	$\widehat{\mathcal{M}} = (2.421, 13.472)$ $\widehat{\mathcal{A}} = (0.166, 0.201)$ $\widehat{\phi}_1 = (0.024, 0.976)$ $\widehat{\phi}_2 = (0.203, 0.202, 0.202, 0.200, 0.193)$	4.140
$RENES$	$\widehat{\mathcal{M}} = (3.551, 5.488)$ $\widehat{\mathcal{A}} = (0.011, 0.434)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.005 & 0.995 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 & 0 \\ 0.388 & 0.398 & 0.214 & 0 & 0 \\ 0.299 & 0.298 & 0.298 & 0.105 & 0 \\ 0.174 & 0.202 & 0.202 & 0.200 & 0.222 \end{bmatrix}$	3.421	$\widehat{\mathcal{M}} = (2.852, 5.006)$ $\widehat{\mathcal{A}} = (0.175, 0.291)$ $\widehat{\phi}_1 = (0.364, 0.636)$ $\widehat{\phi}_2 = (0.175, 0.200, 0.203, 0.220, 0.202)$	3.528

In general, the higher RMS values are noticed when simulations are dictated by the second parameter combination. Such a conclusion makes sense due to the slightly higher simulated values in this case. The aforesaid implies a higher benefit in RMS values when $RENES$ clustering method is applied. A comparison of the corresponding parameter estimates also leads to some interesting conclusions. Namely, an application of the $RENES$ method provides much more accurate estimates of means. A bit more accurate estimates of thinning parameters are also noticed in this case. More accurate estimates

of mentioned parameters eventually lead to significant benefits in RMS values when the *RENES* clustering method is used.

4.4 Application to real-life data

Although the COVID-19 pandemic has shaken the world from its foundations, pandemic monitoring has provided numerous data sets suitable for testing statistical models and techniques. One of such data set has been used to confirm the efficiency of the newly introduced *RENES* clustering method. Same as in [41], a data sequence that represents the number of newly detected COVID-19 cases on daily basis in Mauritius between March 18, 2020 and April 25, 2021 has been chosen. This data can be found on the website Data Europa (<http://www.data.europa.eu>). Figure 4.11 provides the plot of newly detected COVID-19 cases in given period. As figure shows, the number of newly detected cases was kept within acceptable limits in most of the time. Moreover, many days have passed without a single newly infected inhabitant. Sudden jumps were occasional and isolated. Nevertheless, unexpected results showed up between March 22, 2020 and April 9, 2020, as well as between March 6, 2021 and April 9, 2021. As can be seen in Figure 4.11, the number of newly detected cases of the virus oscillated dramatically during these two periods. Sharp and frequent ups and downs started to occur. In other words, very high values began to appear followed by sudden decrements, and vice versa. These observations lead to an assumption that changes in the environment state might have happened.

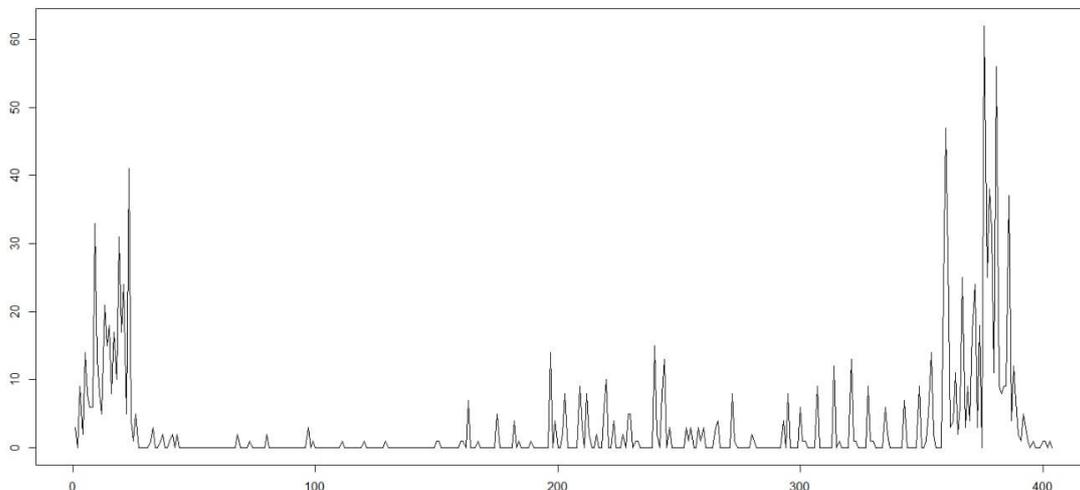


Figure 4.11: Number of newly detected COVID-19 cases on the island of Mauritius on daily basis.

Figure 4.12 provides the plot of the partial autocorrelation function. As figure shows, all orders up to order 5 are significant. To remind, Section 4.3 discusses the impact of the *RENES* method on data modeling by $R2NGINAR(2, 4)$ and $R2NGINAR(2, 5)$ models. Since maximal orders of these models don't exceed 5, they are suitable to be applied on given real-life data. Before models application, the standard K-means and the *RENES* method are both applied in order to provide estimates of the sequence $\{z_n\}$. Optimal values of the *RENES* method parameters are taken from Section 4.3. Clustering results thus obtained are given in Figure 4.13. As one may notice, the *RENES* method managed to recognize the uncommon behavior of the time series during two mentioned time

intervals and showed a great success in placing values that were realized during these periods in a separate cluster. On the other hand, K-means also recognized the uncommon behavior, but the recognition quality is far lower. To be precise, the periods of uncommon behavior are significantly shortened, and many realizations belonging in reality to these periods are recognized by K-means as regular. This implies that the usage of the *RENES* method could make selected $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models even more effective.

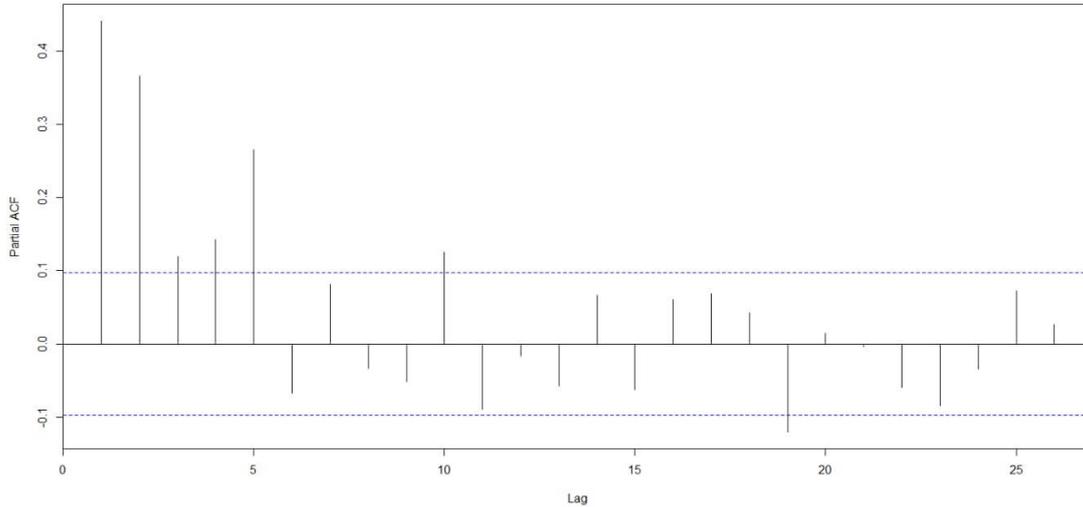
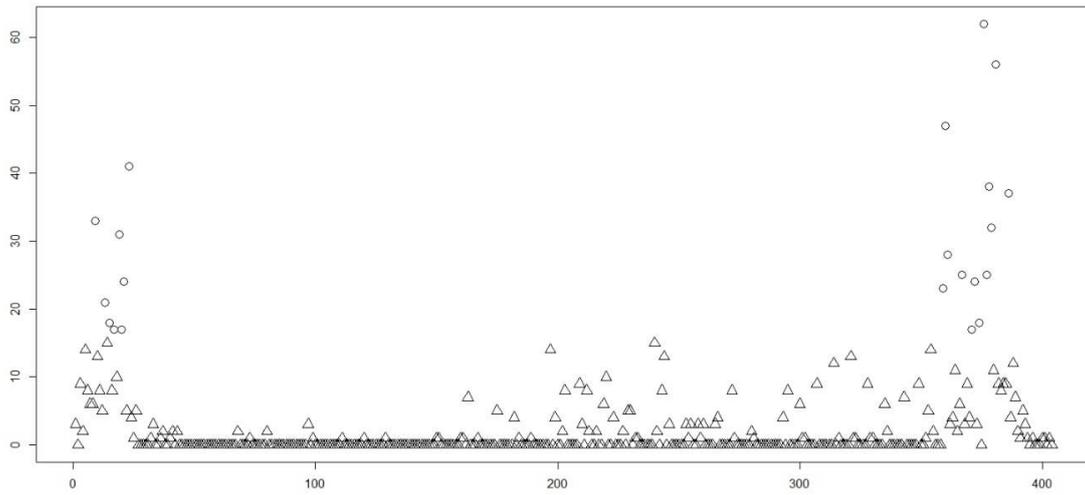


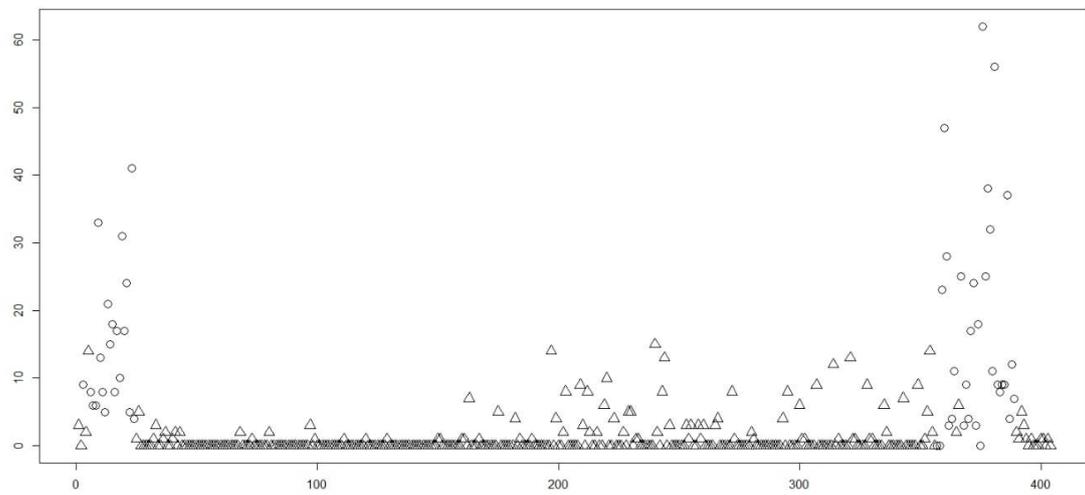
Figure 4.12: PACF for the data that represent a daily number of new COVID-19 cases on the island of Mauritius.

After $\{z_n\}$ has been estimated, the fitting quality of given models ($R2NGINAR(2, 4)$ and $R2NGINAR(2, 5)$) can be examined for each clustering result. The *RMS* of differences between observations and their predicted values is used as a measure of goodness of fit. *RMS*-s obtained after application of $R2NGINAR_{max}(2, 4)$, $R2NGINAR_1(2, 4)$, $R2NGINAR_{max}(2, 5)$ and $R2NGINAR_1(2, 5)$ model are provided in Table 4.6. As can be seen, a choice of the clustering method significantly affects the fitting quality. Namely, all selected models produce much lower *RMS* values after applying the *RENES* method. All of the above proves the supremacy of the *RENES* method and confirms benefits of its use.

Finally, one more fact needs to be emphasized. To make sure that $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models are the most appropriate for modeling given real-life data, many other *INAR* models of stationary or non-stationary nature have also been considered. However, a confirmation of the fact that $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models are the best for given data is not the aim of this chapter. Similar has already been done in [29]. The aim is to confirm that the *RENES* method really contributes to a more efficient application of the $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models. For this reason, modeling results obtained using other *INAR* models are omitted here and can be found in Appendix C.



Standard K-means



RENES method

Figure 4.13: Clustering results for the real-life data.

Table 4.6: *RMS* values and *CML* parameter estimates obtained after reconstruction of chosen real-life data using two different $R2NGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$ and $R2NGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models (both clustering methods are considered).

	$R2NGINAR_{max}(2, 4)$		$R2NGINAR_1(2, 4)$	
Clustering	<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
Regular K-means	$\widehat{\mathcal{M}} = (0.493, 30.191)$ $\widehat{\mathcal{A}} = (0.001, 0.480)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.001 & 0.999 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 \\ 0.328 & 0.328 & 0.344 & 0 \\ 0.248 & 0.197 & 0.241 & 0.314 \end{bmatrix}$	4.259	$\widehat{\mathcal{M}} = (0.489, 30.191)$ $\widehat{\mathcal{A}} = (0.259, 0.473)$ $\widehat{\phi}_1 = (0.018, 0.982)$ $\widehat{\phi}_2 = (0.262, 0.260, 0.200, 0.278)$	4.216
<i>RENES</i>	$\widehat{\mathcal{M}} = (1.102, 14.790)$ $\widehat{\mathcal{A}} = (0.001, 0.512)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.999 & 0.001 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 \\ 0.328 & 0.327 & 0.345 & 0 \\ 0.190 & 0.206 & 0.249 & 0.355 \end{bmatrix}$	3.870	$\widehat{\mathcal{M}} = (1.521, 14.792)$ $\widehat{\mathcal{A}} = (0.248, 0.936)$ $\widehat{\phi}_1 = (0.048, 0.952)$ $\widehat{\phi}_2 = (0.247, 0.244, 0.239, 0.270)$	3.827
	$R2NGINAR_{max}(2, 5)$		$R2NGINAR_1(2, 5)$	
Clustering	<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
Regular K-means	$\widehat{\mathcal{M}} = (0.493, 30.192)$ $\widehat{\mathcal{A}} = (0.001, 0.481)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.001 & 0.999 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 & 0 \\ 0.398 & 0.398 & 0.204 & 0 & 0 \\ 0.298 & 0.300 & 0.301 & 0.101 & 0 \\ 0.198 & 0.201 & 0.198 & 0.200 & 0.203 \end{bmatrix}$	4.150	$\widehat{\mathcal{M}} = (0.492, 30.192)$ $\widehat{\mathcal{A}} = (0.201, 0.472)$ $\widehat{\phi}_1 = (0.021, 0.979)$ $\widehat{\phi}_2 = (0.201, 0.201, 0.200, 0.200, 0.198)$	4.152
<i>RENES</i>	$\widehat{\mathcal{M}} = (1.051, 13.998)$ $\widehat{\mathcal{A}} = (0.002, 0.534)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.999 & 0.001 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 & 0 \\ 0.399 & 0.397 & 0.204 & 0 & 0 \\ 0.298 & 0.300 & 0.298 & 0.104 & 0 \\ 0.123 & 0.201 & 0.201 & 0.202 & 0.273 \end{bmatrix}$	3.768	$\widehat{\mathcal{M}} = (1.102, 14.098)$ $\widehat{\mathcal{A}} = (0.198, 0.493)$ $\widehat{\phi}_1 = (0.016, 0.984)$ $\widehat{\phi}_2 = (0.198, 0.200, 0.203, 0.200, 0.199)$	3.797

Conclusion

This dissertation provides the results of research in the field of *INAR* models, with special reference to *INAR* models with both positive and negative values and random environment *INAR* models. Proposed improvements have enabled the introduction of non-stationarity into *INAR* models with values from the entire set of integers. For this purpose, a concept of random environment, given in [38], served as a useful tool. In that way, new possibilities in modeling the data that are not necessarily above the axis denoting the time component in the Cartesian coordinate system, but oscillate around that axis, have been provided. Beside this, innovations have been also introduced in the environment state estimation procedure. Namely, it's well known that the environment state estimation of each individual realization is one of the crucial steps in modeling real-life processes by usage of models in random environment. For that purpose, the K-means clustering method is most commonly used. However, K-means doesn't show acceptable performances in clustering realizations corresponding to the random environment *INAR* time series of higher order. The main disadvantage of this method is the fact that the value of the time series realization is the only parameter involved in clustering. In order to make the environment state estimates as accurate as possible, the K-means adaptation took place.

At the beginning, an overview of the *INAR* models development was provided, starting with models based on the binomial thinning operator. Further, special attention was paid to the models that are of great importance for the dissertation itself. This primarily refers to *INAR* models that can take both positive and negative values, as well as to nonnegative *INAR* models in random environment. The random environment process itself was also presented. Furthermore, some useful theorems and distributions were given, in order of better understanding the content that has been followed.

After that, a detailed analysis of some particular integer-valued autoregressive time series with values over entire set \mathbb{Z} was presented and some of their properties were examined. In particular, the possibility of extracting and predicting latent components of the true *INAR* time series with skewed *Skellam* marginal distribution was approached. First, the theoretical foundations on which the time series itself is based were given. Appropriate formulas for extracting and predicting latent (hidden) components were derived, provided that realizations of the time series had been known. Formulas such obtained were tested on some real-life data sequences. Results showed a satisfactory goodness of fit for both, latent components extractions and their one-step ahead predictions.

Further, the dissertation was dealing with possibilities of involving non-stationarity into *INAR* models with positive and negative values. The main idea in accomplishing this goal was to combine two familiar kinds of models: stationary *INAR* models with values

over entire \mathbb{Z} and nonnegative *INAR* models in random environment. In particular, a new non-stationary *INAR* model with both positive and negative values was defined by combining models described in [37] and [39], when $\mathcal{P} = \{1\}$. Mentioned model is the first of this kind introduced so far. A thinning operator needed to define the model was borrowed from [37]. In order to estimate unknown parameters of the model, an adaptation of the estimation technique given in [38] was used. Adaptation efficiency was tested on simulated data sequences. Such obtained model showed a remarkable success in application to real-life data, compared to other models that made sense to apply to the given data.

Finally, author discussed the problem of improving the K-means clustering method, in order to make it more suitable for clustering the data that correspond to the random environment *INAR* time series of higher order. Although the similar idea can be applied to an arbitrary random environment *INAR* model of higher order, with an arbitrary marginal distribution and thinning operator, the dissertation provides the K-means improvement particularly suitable for the data that correspond to the generalized *RrNGINAR*($\mathcal{M}, \mathcal{A}, \mathcal{P}$) time series. The improved method, named *RENES*, follows the behavior of all parameters of the *RrNGINAR*($\mathcal{M}, \mathcal{A}, \mathcal{P}$) model that carry information about belonging to the particular environment state. Efficiency of such improved clustering method was tested on simulated data and compared to clustering results obtained using the standard K-means method. The progress in number of correctly estimated states and longer arrays of consecutive realizations corresponding to the same state were detected. At the very end, a supremacy of the newly proposed *RENES* method over standard K-means is confirmed on popular real-life data.

Further research might be performed into several directions. Regarding the non-stationary *INAR* models with positive and negative values, generalized models of order higher than 1 might be created. Further, improvements might be achieved by setting some other marginal distribution, or a distribution of the counting sequence involved in thinning operator. Furthermore, random environment process might be defined in some other way. On the other hand, K-means improvement might be constructed to suit the data corresponding to some other random environment *INAR* time series of higher order. Also, improving other clustering methods might be useful.

Bibliography

- [1] Al-Osh, M. A., Aly, E. E. A. A. (1992) First order autoregressive time series with negative binomial and geometric marginals, *Communications in Statistics-Theory and Methods* 21(9), 2483-2492.
- [2] Al-Osh, M. A., Alzaid, A. A. (1987) First-order integer-valued autoregressive (INAR(1)) process, *Journal of Time Series Analysis* 8(3), 261-275.
- [3] Alzaid, A. A., Al-Osh, M. A. (1988) First-order integer-valued autoregressive (INAR(1)) process: distributional and regression properties, *Statistica Neerlandica* 42(1), 53-61.
- [4] Alzaid, A. A., Al-Osh, M. A. (1993) Some autoregressive moving average processes with generalized Poisson marginal distributions, *Annals of the Institute of Statistical Mathematics* 45(2), 223-232.
- [5] Alzaid, A. A., Omair, M. A. (2014) Poisson difference integer valued autoregressive model of order one, *Bulletin of the Malaysian Mathematical Sciences Society* 37(2), 465-485.
- [6] Barreto-Souza, W., Bourguignon, M. (2015) A skew INAR(1) process on \mathbb{Z} , *AStA Advances in Statistical Analysis* 99(2), 189-208.
- [7] Brockwell, P., Davis, R. (2002) Introduction to time series and forecasting, Second edition, *Springer-Verlag*, New York.
- [8] Chesneau, C., Kachour, M. (2012) A parametric study for the first-order signed integer-valued autoregressive process, *Journal of Statistical Theory and Practice* 6(4), 760-782.
- [9] Cox, D. R., Miller, H. D. (1965) The theory of stochastic processes, *Methuen & Co*, London.
- [10] Djordjević, M. S. (2016) Doprinos analizi vremenskih nizova sa celobrojnim vrednostima (In Serbian), Doctoral dissertation, *University of Niš, Faculty of science and Mathematics*.
- [11] Djordjević, M. S. (2017) An extension on INAR models with discrete Laplace marginal distributions, *Communication in Statistics - Theory and Methods* 46(12), 5896-5913.
- [12] Djordjević, M. S., Ristić, M. M., **Pirković, B. A.** (2021) Identifying latent components of the TINAR(1) model, *Filomat* 35(13), 4469-4482.

- [13] Du, J. G., Li, Y. (1991) The integer-valued autoregressive (INAR(p)) model, *Journal of Time Series Analysis* 12(2), 129-142.
- [14] Franke, J., Subba Rao T. (1993) Multivariate first-order integer-valued autoregressions, *Technical Report, Mathematics Department, UMIST*.
- [15] Freeland, R. K. (2010) True integer value time series, *AStA Advances in Statistical Analysis* 94(3), 217-229.
- [16] Freeland, R. K., McCabe, B. (2004) Analysis of low count time series by Poisson autoregression, *Journal of Time Series Analysis* 25(5), 701-722.
- [17] Gelman, A., Hwang, J., Vehtari, A. (2014) Understanding predictive information criteria for Bayesian models, *Statistics and Computing* 24(6), 997-1016.
- [18] Hartigan, J. A., Wong, M. A. (1979) Algorithm AS 136: a K-Means clustering algorithm, *Journal of the Royal Statistical Society, Series C* 28(1), 100-108.
- [19] Inusah, S., Kozubowski, T. J. (2006) A discrete analogue of the Laplace distribution, *Journal of Statistical Planning and Inference* 136(3), 1090-1102.
- [20] Irwin, J. O. (1937) The frequency distribution of the difference between two independent variates following the same Poisson distribution, *Journal of the Royal Statistical Society* 100(3), 415-416.
- [21] Jacobs, P. A., Lewis, P. A. W. (1978) Discrete time series generated by mixtures I: correlational and runs properties, *Journal of the Royal Statistical Society, Series B* 40(1), 94-105.
- [22] Jacobs, P. A., Lewis, P. A. W. (1978) Discrete time series generated by mixtures II: asymptotic properties, *Journal of the Royal Statistical Society, Series B* 40(2), 222-228.
- [23] Jacobs, P. A., Lewis, P. A. W. (1978) Discrete time series generated by mixtures III: autoregressive processes (DAR(p)), *Naval Postgraduate School-Technical Report*, Accession Number: ADA062208.
- [24] Joe, H. (1996) Time series models with univariate margins in the convolution-closed infinitely divisible class, *Journal of Applied Probability* 33(3), 664-677.
- [25] Kachour, M., Truquet, L. (2011) A p-Order signed integer-valued autoregressive (SINAR(p)) model, *Journal of Time Series Analysis* 32(3), 223-236.
- [26] Karlsen, H., Tjøstheim, D. (1988), Consistent estimates for the NEAR(2) and NLAR(2) time series models, *Journal of the Royal Statistical Society, Series B* 50(2), 313-320.
- [27] Kim, H. Y., Park, Y. (2008) A non-stationary integer-valued autoregressive model, *Statistical Papers* 49(3), 485-502.
- [28] Kozubowski, T. J., Inusah, S. (2006) A skew Laplace distribution on integers, *Annals of the Institute of Statistical Mathematics* 58(3), 555-571.

- [29] Laketa, P. N., Nastić, A. S., Ristić, M. M. (2018) Generalized random environment INAR models of higher order, *Mediterranean Journal of Mathematics* 15(1):9.
- [30] Laketa, P. N. Crossed bivariate integer-valued autoregressive process based on bivariate random environment process, *Communications in Statistics-Theory and Methods*, (to appear).
- [31] Latour, A., Truquet, L. (2009) An integer-valued bilinear type model, Research report available on <http://hal.archives-ouvertes.fr/hal-00373409>.
- [32] Liu, Z., Li, Q., Zhu, F. (2021) Semiparametric integer-valued autoregressive models on \mathbb{Z} , *The Canadian Journal of Statistics* 49(4), 1317-1337.
- [33] Mann, H. B., Wald, A. (1943) On stochastic limit and order relationships, *The Annals of Mathematical Statistics* 14(3), 217-226.
- [34] McKenzie, E. (1985) Some simple models for discrete variate time series, *Journal of the American Water Resources Association* 21(4), 645-650.
- [35] Nastić, A. S. (2008) Autoregresivni procesi sa nenegativnim celobrojnim vrednostima (in Serbian), Magisterial thesis, *University of Niš, Faculty of science and Mathematics*.
- [36] Nastić, A. S., Ristić, M. M. (2012) Some geometric mixed integer-valued autoregressive (INAR) models, *Statistics & Probability Letters* 82(4), 805-811.
- [37] Nastić, A. S., Ristić, M. M., Djordjević, M. S. (2016) An INAR model with discrete Laplace marginal distributions, *Brazilian Journal of Probability and Statistics* 30(1), 107-126.
- [38] Nastić, A. S., Laketa, P. N., Ristić, M. M. (2016) Random environment integer-valued autoregressive process, *Journal of Time Series Analysis* 37(2), 267-287.
- [39] Nastić, A. S., Laketa, P. N., Ristić, M. M. (2019) Random environment INAR models of higher order, *REVSTAT-Statistical Journal* 17(1), 35-65.
- [40] Nastić, A. S., Ristić, M. M., Bakouch, H. S. (2012) A combined geometric INAR(p) model based on negative binomial thinning, *Mathematical and Computer Modeling* 55(5-6), 1665-1672.
- [41] **Pirković, B. A.**, Laketa, P. N., Nastić, A. S. (2021) On generalized random environment INAR models of higher order: Estimation of random environment states, *Filomat* 35(13), 4545-4576.
- [42] **Pirković, B. A.**, Ristić, M. M., Nastić, A. S. (2022) Random environment integer-valued autoregressive process with discrete Laplace marginal distributions, *REVSTAT-Statistical Journal*, (to appear).
- [43] Ristić, M. M., Bakouch, H. S., Nastić, A. S. (2009) A new geometric first-order integer-valued autoregressive (NGINAR(1)) process, *Journal of Statistical Planning and Inference* 139(7), 2218-2226.
- [44] Ristić, M. M., Nastić, A. S. (2012) A mixed INAR(p) model, *Journal of Time Series Analysis* 33(6), 903-915.

- [45] Ristić, M. M., Nastić, A. S., Bakouch, H. S. (2012) Estimation in an integer-valued autoregressive process with negative binomial marginals (NBINAR(1)), *Communications in Statistics-Theory and Methods* 41(4), 606-618.
- [46] Silva, M. E., Oliveira, V. L. (2005) Difference equations for the higher order moments and cumulants of the INAR(1) model, *Journal of Time Series Analysis* 26(1), 17-36.
- [47] Skellam, J. G., (1946) The frequency distribution of the difference between two Poisson variates belonging to different populations, *Journal of the Royal Statistical Society* 109(3), 296.
- [48] Steutel, F. W., van Harn, K. (1979) Discrete analogues of self-decomposability and stability, *The Annals of Probability* 7(5), 893-899.
- [49] Weiß, C. H. (2008) Thinning operations for modeling time series of counts-a survey, *ASTA Advances in Statistical Analysis* 92(3), 319-341.
- [50] Zheng, H., Basawa, I. V., Datta, S. (2007) First-order random coefficient integer-valued autoregressive processes, *Journal of Statistical Planning and Inference* 137(1), 212-229.

Appendix

Appendix A. The choice of model parameters in case the of $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations with 3 environment states

To create simulated $R3NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ time series, the following parameter combinations are used.

1. Same as before, the first combination assumes similar means within states, i.e. $\mathcal{M} = (0.5, 1, 1.5)$. Besides, thinning parameters α_s , $s = 1, 2, 3$, differ significantly, having the values $\mathcal{A} = (0.1, 0.35, 0.6)$. The same holds for vector of maximal orders $\mathcal{P} = (2, 4, 2)$, while corresponding probability matrices are of the form

$$\phi_1 = \begin{bmatrix} 1 & 0 \\ 0.9 & 0.1 \end{bmatrix}, \quad \phi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.2 & 0.8 & 0 & 0 \\ 0.2 & 0.4 & 0.4 & 0 \\ 0.2 & 0.2 & 0.3 & 0.3 \end{bmatrix}, \quad \phi_3 = \begin{bmatrix} 1 & 0 \\ 0.1 & 0.9 \end{bmatrix}.$$

Probability matrices given above are used to create $R3NGINAR_{max}(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations. Corresponding probabilities related to $R3NGINAR_1(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations are located in last rows of given matrices. One short clarification is needed here. Bearing in mind that the first and third state have the same maximal orders, one may seem that these two states do not 'differ enough'. However, the analysis of other parameters concludes the opposite. Corresponding means, as well as corresponding thinning parameters, do differ significantly. On the other hand, the first and second state have much more similar mean values, but corresponding thinning parameters and maximal orders differ significantly. The same holds for the second and third state. This surrounding, where various properties of elements must be taken into account to obtain the proper clustering results, seems perfect for *RENES* method to show its potential.

Further, an initial state is nearly fair, due to the value of its distribution $p_{vec} = (0.3, 0.4, 0.3)$, while the transition probability matrix favors simulations to remain in the same state, having the values on the main diagonal significantly higher than

those out of the main diagonal. Hence, $p_{mat} = \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.1 & 0.8 & 0.1 \\ 0.2 & 0.2 & 0.6 \end{bmatrix}$.

2. An interesting challenge for *RENES* method is also created with the second combination of model parameters. Namely, parameters are set as follows: $\mathcal{M} = (2, 4, 6)$,

$\mathcal{A} = (0.2, 0.3, 0.6)$ and $\mathcal{P} = (2, 4, 5)$. Therewithal,

$$\phi_1 = \begin{bmatrix} 1 & 0 \\ 0.7 & 0.3 \end{bmatrix}, \phi_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 \\ 0.3 & 0.3 & 0.4 & 0 \\ 0.3 & 0.2 & 0.2 & 0.3 \end{bmatrix}, \phi_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 & 0 \\ 0.2 & 0.5 & 0.3 & 0 & 0 \\ 0.25 & 0.3 & 0.2 & 0.25 & 0 \\ 0.2 & 0.2 & 0.3 & 0.1 & 0.2 \end{bmatrix}.$$

It can be noticed that means within states grow progressively, even though the jumps are not that much high. The similarity of thinning parameters characterizes the first and second state, while corresponding orders differ significantly. On the other hand, the second and third state have similar orders, while corresponding thinning parameters differ significantly. Finally, all parameters of the first and third state differ significantly. To place the realization at moment n in the appropriate cluster, it is crucial for clustering method to possess information about the behavior of all parameters of the model at the same moment.

The distribution of an initial state is $p_{vec} = (0.35, 0.35, 0.3)$ and a transition probability matrix is of the form

$$p_{mat} = \begin{bmatrix} 0.9 & 0.05 & 0.05 \\ 0.2 & 0.7 & 0.1 \\ 0.1 & 0.1 & 0.8 \end{bmatrix}.$$

Appendix B. Simulation results and selection of the *RENES* method parameters in the case of $RrNGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ simulations with 3 environment states

A testing of the newly defined *RENES* method on simulated data with 3 environment states follows the same path as in Section 4.3. Namely, as it was the case with the 2 environment state simulations, $R3NGINAR_{max}(2, 4, 2)$ and $R3NGINAR_1(2, 4, 2)$ time series with 3 environment states are simulated first (two replications of each). After that, one may start with determination of the *RENES* method parameters. The sequence $\{\tilde{\mu}_n\}$ is once again obtained using equality (4.3). In order to improve the sequence of pre-estimates obtained in this manner, an optimal selection of \mathbf{v}_m is of importance. Figure B.1 and Figure B.2 provide sequences of pre-estimates $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ obtained for various selections of \mathbf{v}_m . Again, the sequence of exact means $\{\mu_n\}$ is taken as a benchmark.

According to figures, the best pre-estimate sequence is obtained by usage of the vector $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$. The same conclusion holds for both simulations with 3 environment states, $R3NGINAR_{max}(2, 4, 2)$ and $R3NGINAR_1(2, 4, 2)$. Beside the capability of trimming high peaks, a significance of such obtained pre-estimates is especially reflected through its ability of adequately assessing means within the second (middle) state.

In order to carry out an estimation of the sequence $\{P_n\}$, optimal value of the parameter d_p needs to be determined first. Again, various d_p values are selected and corresponding values of the error Δ_p are calculated. All corresponding results are given in Table B.1. In the case of $R3NGINAR_{max}(2, 4, 2)$ simulation, the best result is obtained for $d_p = 17$

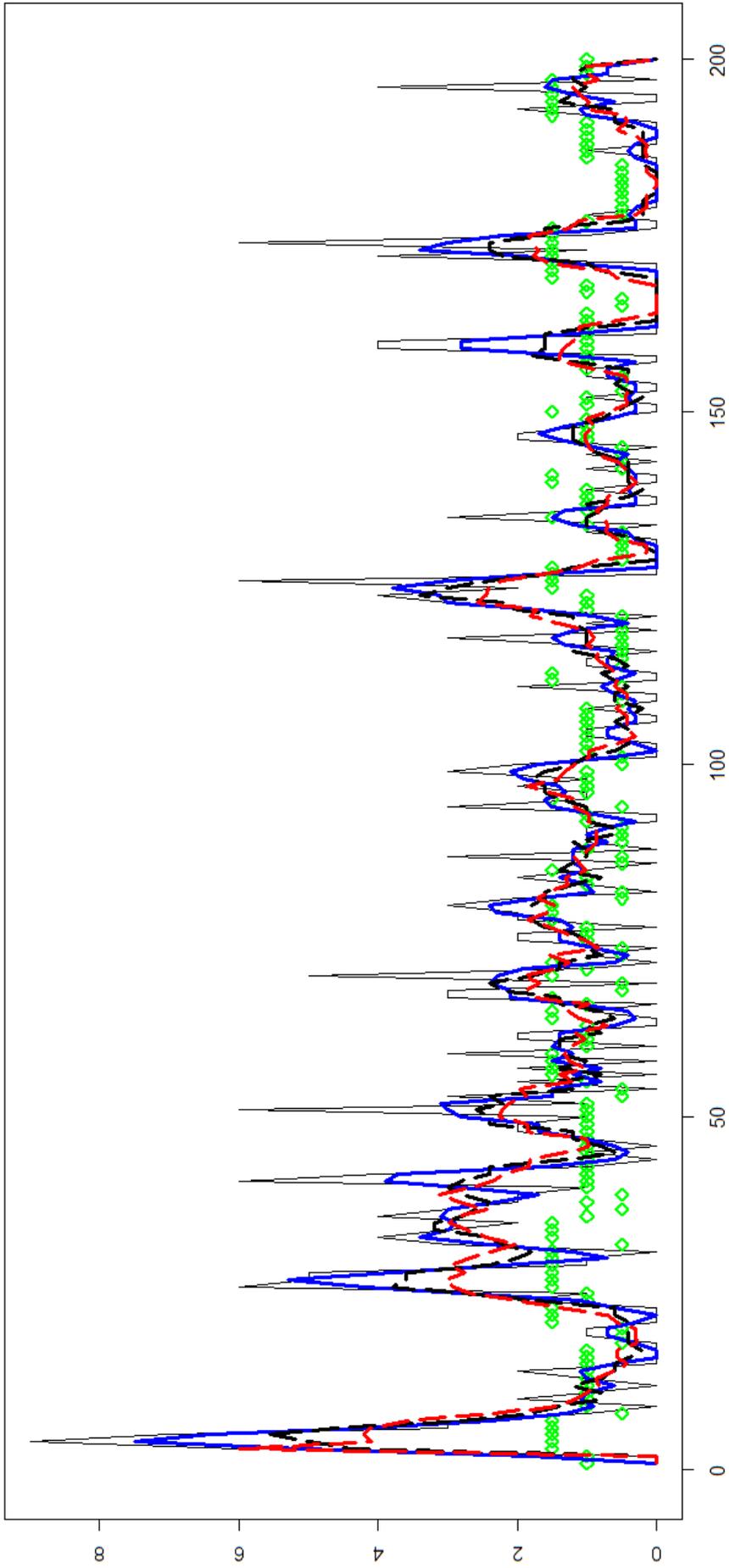


Figure B.1: Corresponding pre-estimates of $\{\mu_n\}$ obtained for various selections of the vector \mathbf{v}_m in the case of simulated $R3NGINAR_{max}(2, 4, 2)$ models: green – sequence of exact means $\{\mu_n\}$; black (regular) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = 1$; blue (thick) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$.

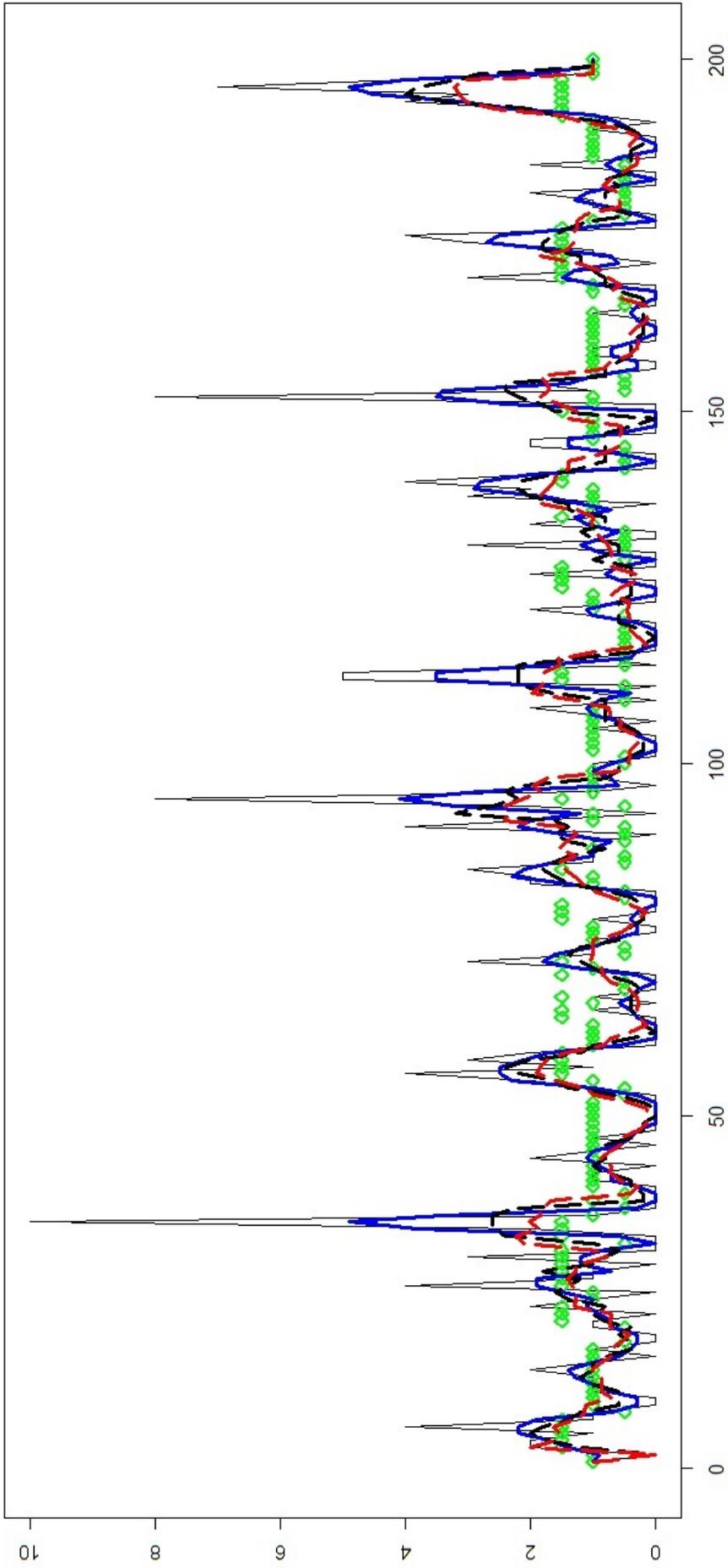


Figure B.2: Corresponding pre-estimates of $\{\mu_n\}$ obtained for various selections of the vector \mathbf{v}_m in the case of simulated $R3NGINAR_1(2, 4, 2)$ models: green – sequence of exact means $\{\mu_n\}$; black (regular) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = 1$; blue (thick) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ for $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$.

($\Delta_p = 1.415$), while in the case of $R3NGINAR_1(2, 4, 2)$ simulation optimal result is obtained for $d_p = 18$ ($\Delta_p = 1.407$).

Table B.1: Values of Δ_p for various selections of d_p .

$R3NGINAR_{max}(2, 4, 2)$		$R3NGINAR_1(2, 4, 2)$		$R3NGINAR_{max}(2, 4, 5)$		$R3NGINAR_1(2, 4, 5)$	
d_p	Δ_p	d_p	Δ_p	d_p	Δ_p	d_p	Δ_p
5	1.570	5	1.604	5	1.924	5	1.880
6	1.568	6	1.694	6	1.914	6	1.905
7	1.572	7	1.581	7	1.911	7	1.917
8	1.522	8	1.579	8	1.881	8	1.931
9	1.567	9	1.568	9	1.822	9	1.886
10	1.555	10	1.595	10	1.784	10	1.860
11	1.519	11	1.564	11	1.748	11	1.849
12	1.502	12	1.571	12	1.730	12	1.889
13	1.503	13	1.535	13	1.773	13	1.880
14	1.463	14	1.473	14	1.806	14	1.904
15	1.424	15	1.473	15	1.797	15	1.910
16	1.429	16	1.455	16	1.809	16	1.915
17	1.415	17	1.473	17	1.822	17	1.949
18	1.417	18	1.407	18	1.843	18	1.943
19	1.419	19	1.454	19	1.853	19	1.915
20	1.417	20	1.432	20	1.874	20	1.883

For fixed optimal value of d_p , a corresponding vector \mathbf{v}_p have to be provided. Figure B.3 and Figure B.4 provide sequences of pre-estimates $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ obtained for various selections of \mathbf{v}_p . The sequence of exact orders $\{P_n\}$ will be the benchmark here. According to figures, pre-estimate sequences show similar behavior when $k = 2$, $k = 3$ and $k = 4$. Each of them offers similar chance to obtain the correct clustering of the particular observation. Also, each of those sequences is a way better then the one obtained for $k = 1$, having no frequent and sharp ups and downs, which would lead to the erroneous clustering result. Due to the simplicity of the method, $k = 2$ is preferred, i.e. $\mathbf{v}_p = (0.4, 0.3)$ is taken in both cases.

Provided $\{T(\tilde{\mu}_n, \mathbf{v}_m)\}$ and $\{\tilde{P}_n\}$ enable one to determine $\tilde{\alpha}_n$, $n \in N$, using (4.5). Same as before, a suitable form of the vector \mathbf{v}_a can significantly improve such obtained pre-estimates. Figure B.5 and Figure B.6 show sequences of pre-estimates $\{T(\tilde{\alpha}_n, \mathbf{v}_a)\}$ obtained for various selections of \mathbf{v}_a . The sequence of exact thinning parameters $\{\alpha_n\}$ is chosen to be a benchmark.

As can be noticed, a sequence of pre-estimates obtained for $k = 4$ shows better fitting compared to other sequences. It oscillates between α_1 and α_3 most of the time, with particular good assessment of α_2 . Just a few sharp jumps are detected. In other words, pre-estimates rarely exceed α_3 (the highest thinning parameter value), and even when they do, overruns are bearable. The same holds for both simulations, $R3NGINAR_{max}(2, 4, 2)$ and $R3NGINAR_1(2, 4, 2)$. Thus, $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$ is taken in both cases.

To sum up, Table B.2 contains the *RENES* method parameters d_p , \mathbf{v}_m , \mathbf{v}_a and \mathbf{v}_p related to the simulated $R3NGINAR(2, 4, 2)$ time series. It remains to determine C_m, C_a and

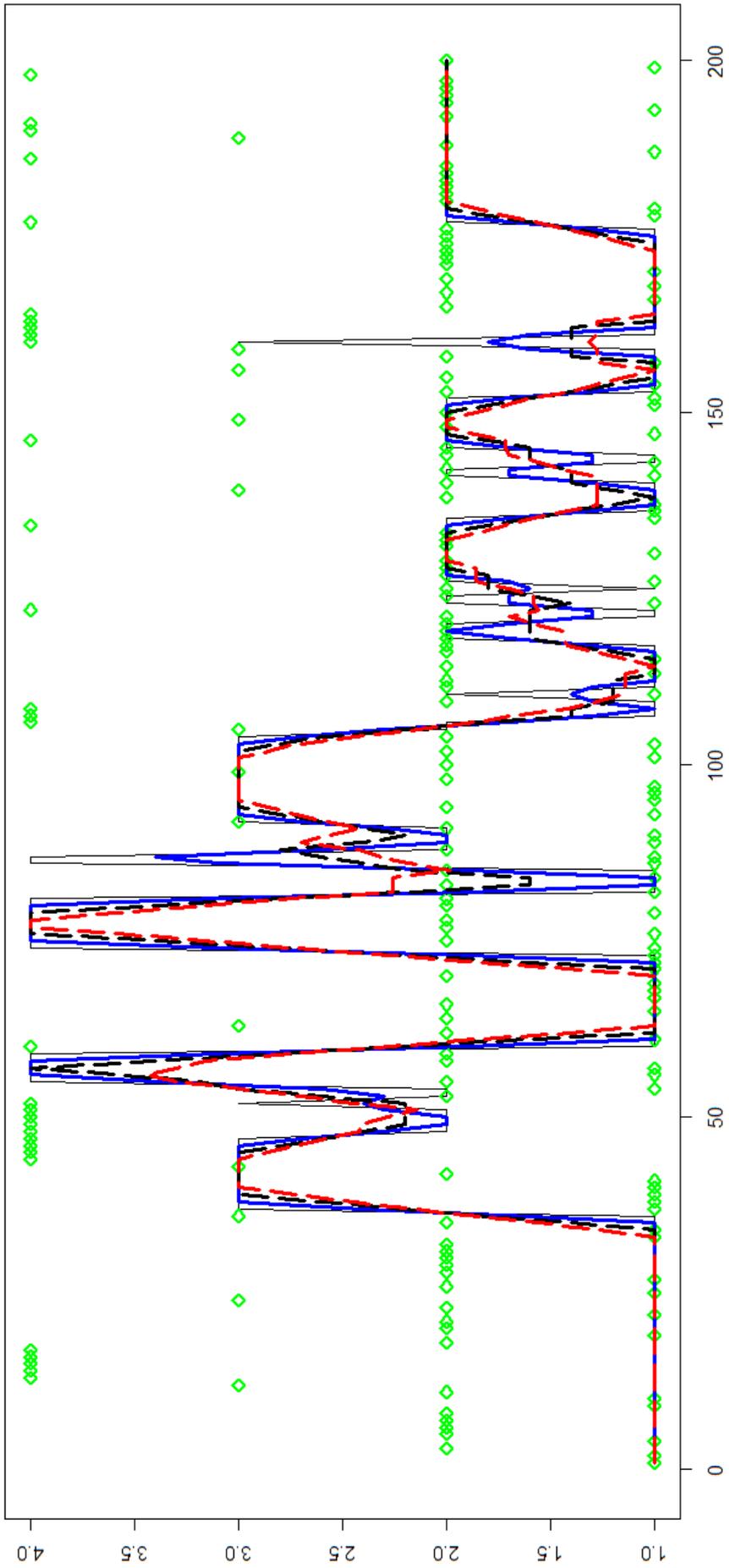


Figure B.3: Corresponding pre-estimates of $\{P_n\}$ obtained for various selections of the vector \mathbf{v}_p in the case of simulated $R3NGINAR_{v_{max}}(2, 4, 2)$ models: green – exact order sequence $\{P_n\}$; black (regular) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = 1$; blue (thick) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.4, 0.3)$; black (dashed) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\tilde{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14)$.

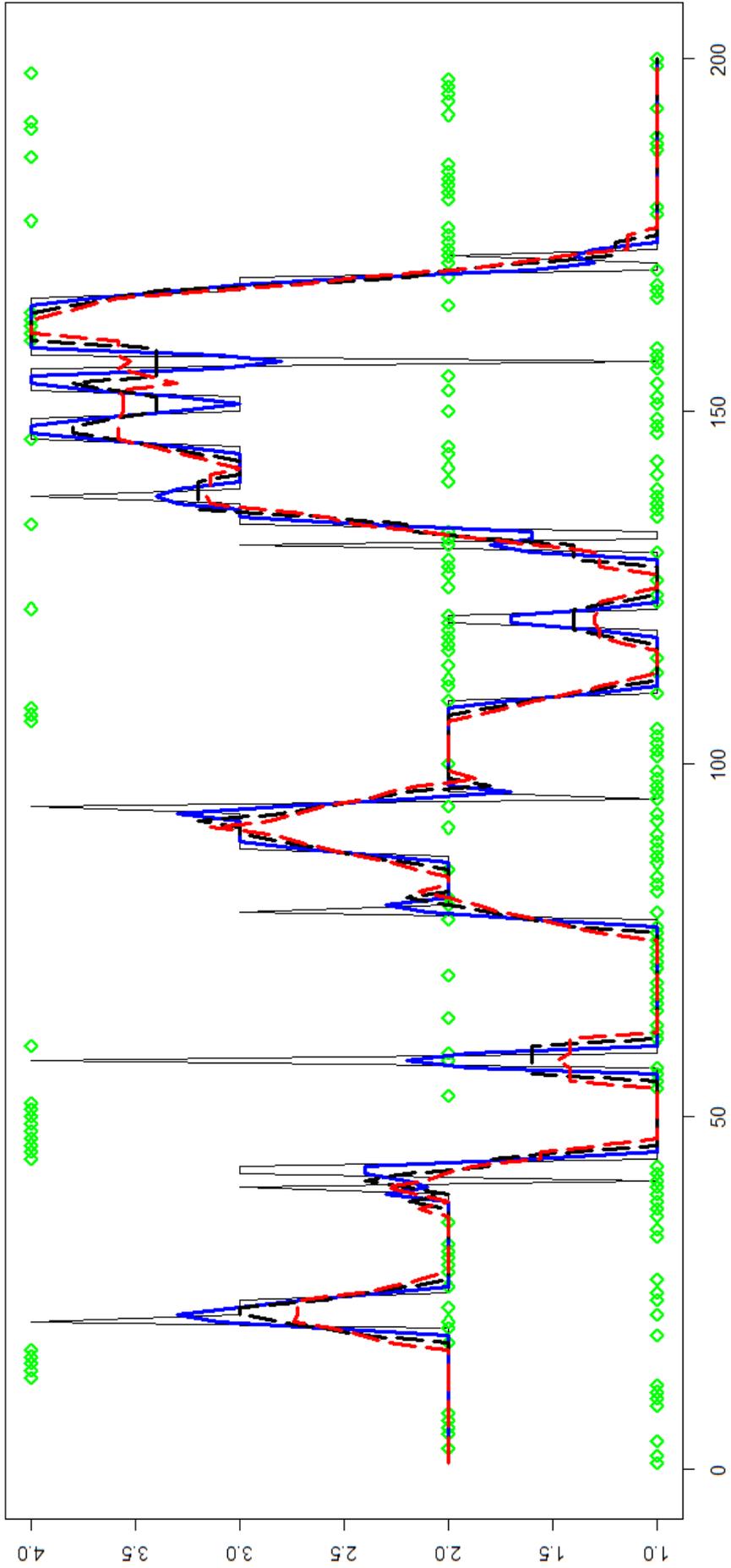


Figure B.4: Corresponding pre-estimates of $\{P_n\}$ obtained for various selections of the vector \mathbf{v}_p in the case of simulated $R3NGINAR_1(2, 4, 2)$ models: green – exact order sequence $\{P_n\}$; black (regular) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = 1$; blue (thick) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.4, 0.3)$; black (dashed) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\hat{P}_n, \mathbf{v}_p)\}$ for $\mathbf{v}_p = (0.16, 0.14, 0.14, 0.14)$.

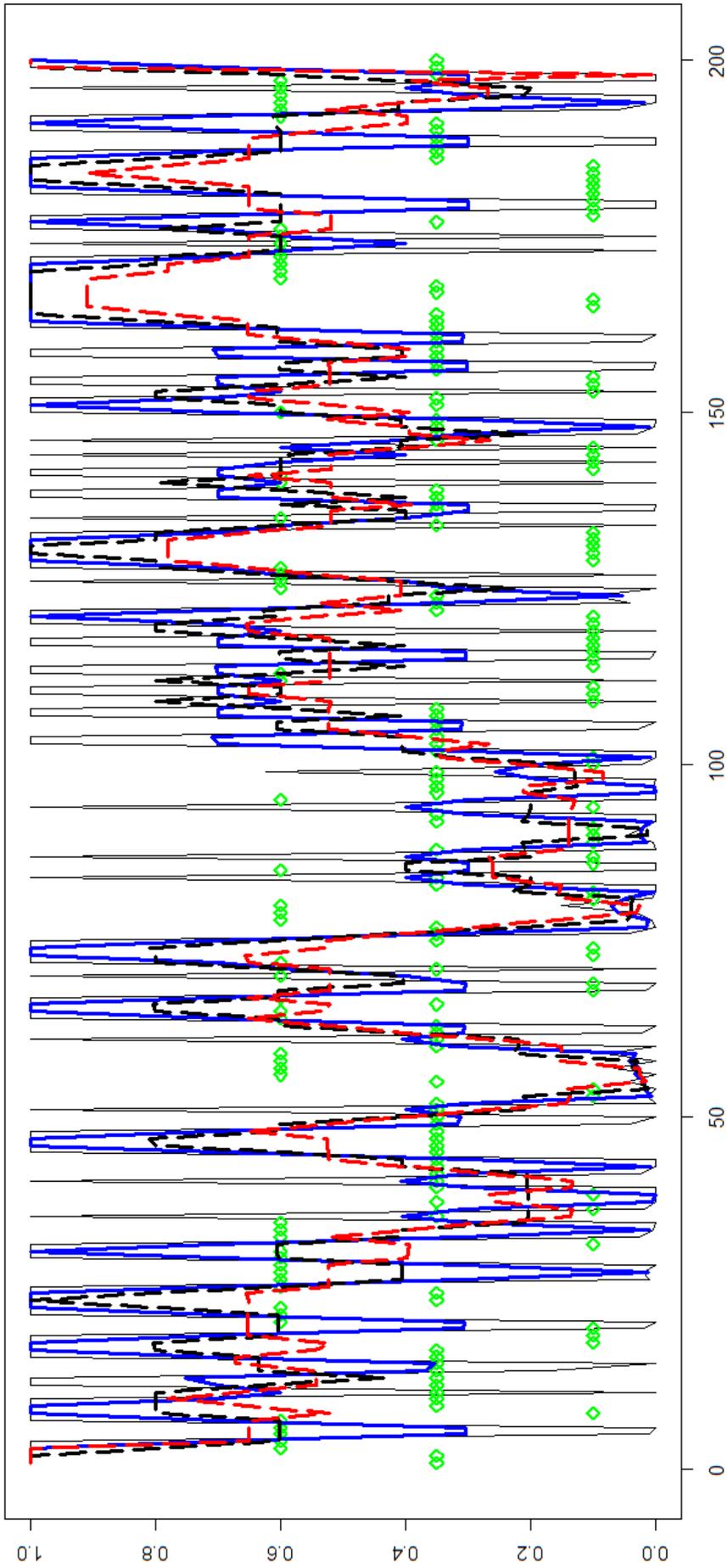


Figure B.5: Corresponding pre-estimates of $\{\alpha_n\}$ obtained for various selections of the vector \mathbf{v}_a in the case of simulated $R3NGINAR_{max}(2, 4, 2)$ models: green – exact thinning parameters sequence $\{\alpha_n\}$; black (regular) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = 1$; blue (thick) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.4, 0.3)$; black (dashed) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$.

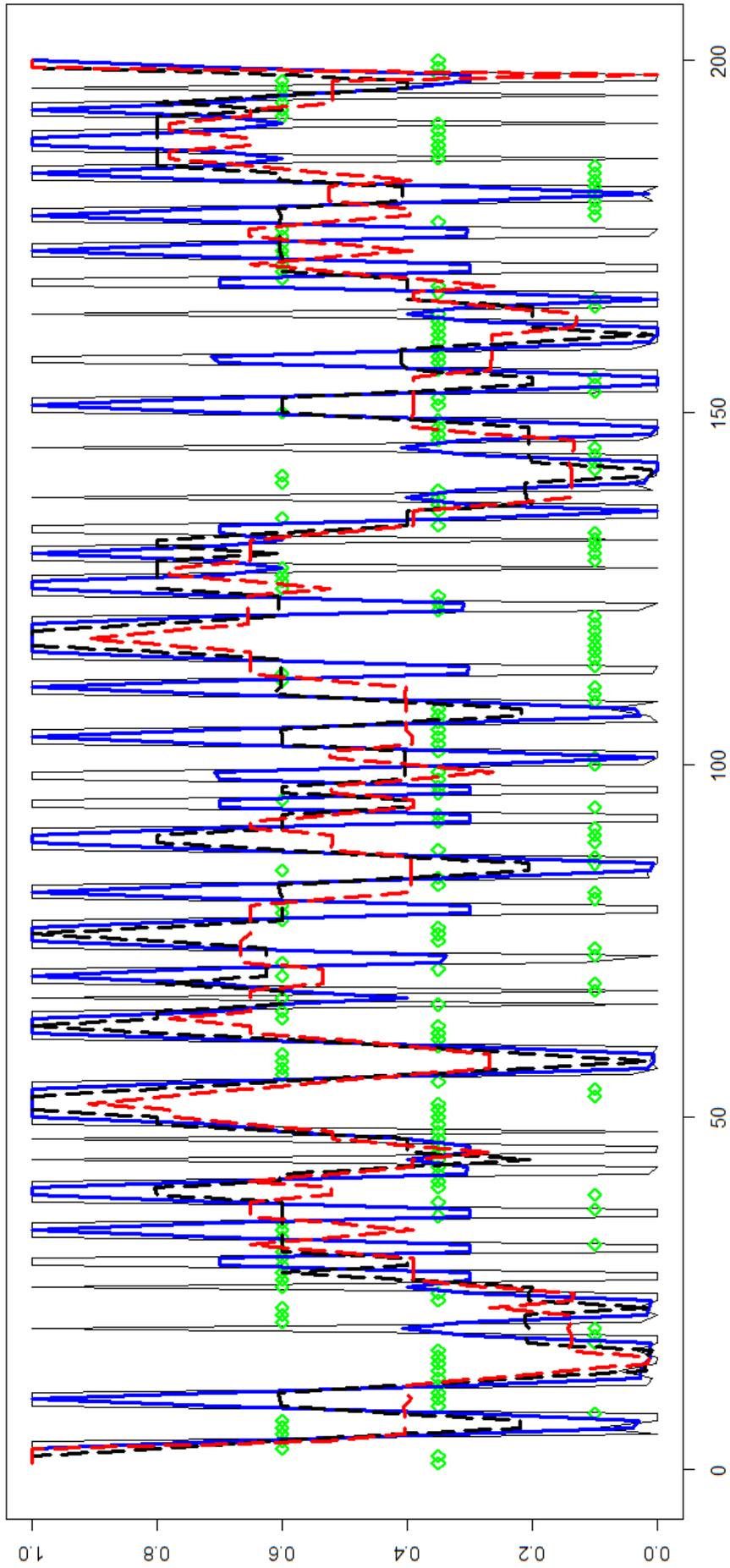


Figure B.6: Corresponding pre-estimates of $\{\alpha_n\}$ obtained for various selections of the vector \mathbf{v}_a in the case of simulated $R3NGINAR_1(2, 4, 2)$ models: green – exact thinning parameters sequence $\{\alpha_n\}$; black (regular) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = 1$; blue (thick) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.4, 0.3)$; black (dashed) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.2, 0.2, 0.2)$; red (dashed) – sequence $\{T(\hat{\alpha}_n, \mathbf{v}_a)\}$ for $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$.

Table B.2: Optimal values of the constant d_p and vectors \mathbf{v}_m , \mathbf{v}_a , \mathbf{v}_p , related to the simulated $R3NGINAR(2, 4, 2)$ time series.

$R3NGINAR_{max}(2, 4, 2)$			
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p
17	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.4,0.3)
$R3NGINAR_1(2, 4, 2)$			
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p
18	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.4,0.3)

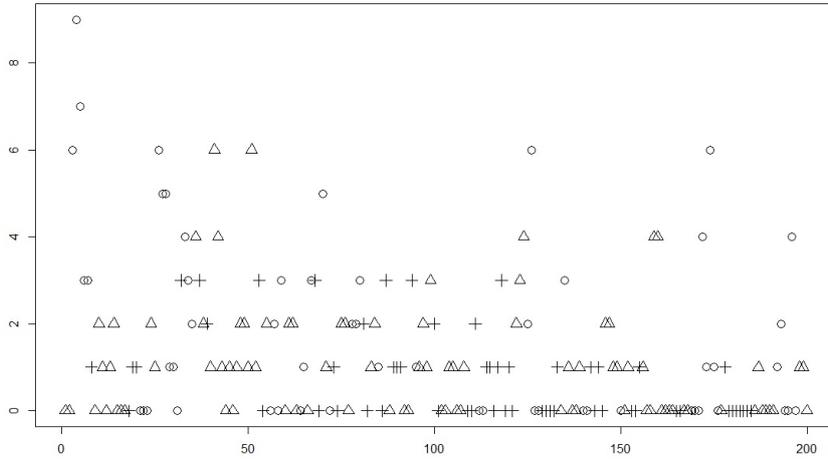
C_p , by which one controls the level of impact each $\{S(\tilde{\mu}_n, \mathbf{v}_m)\}$, $\{S(\tilde{\alpha}_n, \mathbf{v}_a)\}$, $\{S(\tilde{P}_n, \mathbf{v}_p)\}$ has on the clustering procedure. For that purpose, the clustering of

$$\{(C_m S(\tilde{\mu}_n, \mathbf{v}_m), C_a S(\tilde{\alpha}_n, \mathbf{v}_a), C_p S(\tilde{P}_n, \mathbf{v}_p))\}$$

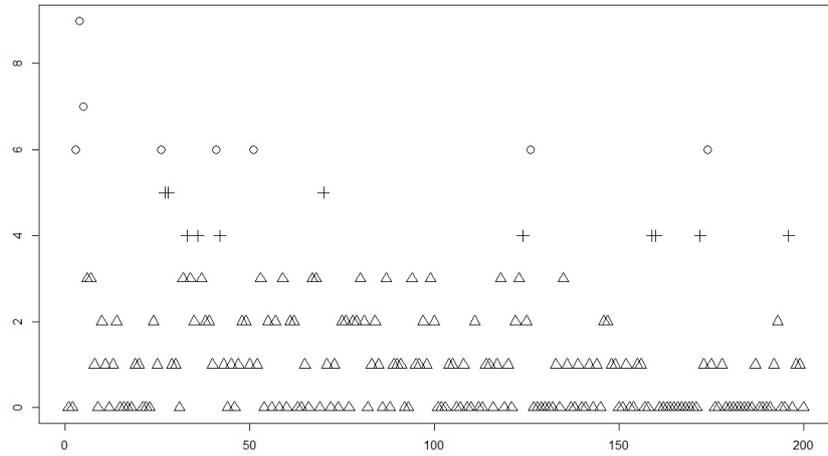
is performed for all $C_m, C_a, C_p = 1, 2, \dots, 10$. Thousand different estimates of the environment state sequence $\{z_n\}$ are provided in this way. In the case of $R3NGINAR_{max}(2, 4, 2)$ simulation, the best estimation result is obtained for $C_m = 9$, $C_a = 7$ and $C_p = 2$ with 209 estimated states equal to corresponding exact states. On the other hand, the standard K-means managed to reach 155 exactly estimated states, which is a significant deterioration. Figure B.7 offers a comparative overview of exact states, states obtained by standard K-means and states obtained by usage of the newly proposed *RENES* method. The same path is followed for obtaining C_m, C_a and C_p in the case of $R3NGINAR_1(2, 4, 2)$ simulation. Values $C_m = 6$, $C_a = 1$ and $C_p = 8$ provide the best estimation result with 216 estimated states equal to corresponding exact states. Compared to that, the standard K-means reached only 153 exactly estimated states. Figure B.8 gives a comparative overview of exact states, states obtained by standard K-means and states obtained by usage of the *RENES* method.

As in the case of simulations with 2 environment states, some improvements that *RENES* method offers in regard to the standard K-means are noticed. The higher number of exactly estimated states has been already discussed. Beside this, one might notice the newly proposed *RENES* method creates much longer sequences of consecutive elements corresponding to the same state. This improvement implies fruitful application of all *RrINAR* models, including $RrINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$. Further, it is possible to find very high or very low values in any of three given states. This feature allows the *RENES* method to properly cluster even those realizations with uncommon behavior in regard to the state in which they find themselves. Furthermore, a possibility of having equal realizations in different states is one of the improvements. This property additionally increases the flexibility of the *RENES* method and makes it more applicable in regard to the standard K-means.

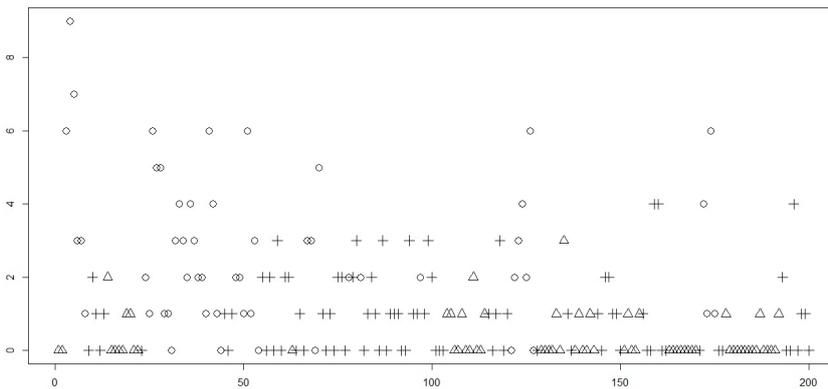
It is left to check out the level of benefit one can get by applying the *RENES* method instead of K-means. Unused replications of the simulated $R3NGINAR_{max}(2, 4, 2)$ and $R3NGINAR_1(2, 4, 2)$ time series are exploited for that purpose. Same as before, the standard K-means and the *RENES* method are both applied on the data and corresponding unknown model parameters are obtained with the help of *CML* procedure. For each clustering result singularly, corresponding $R3NGINAR_{max}(2, 4, 2)$ or $R3NGINAR_1(2, 4, 2)$



Exact states of the $R3NGINAR_{max}(2, 4, 2)$ simulation

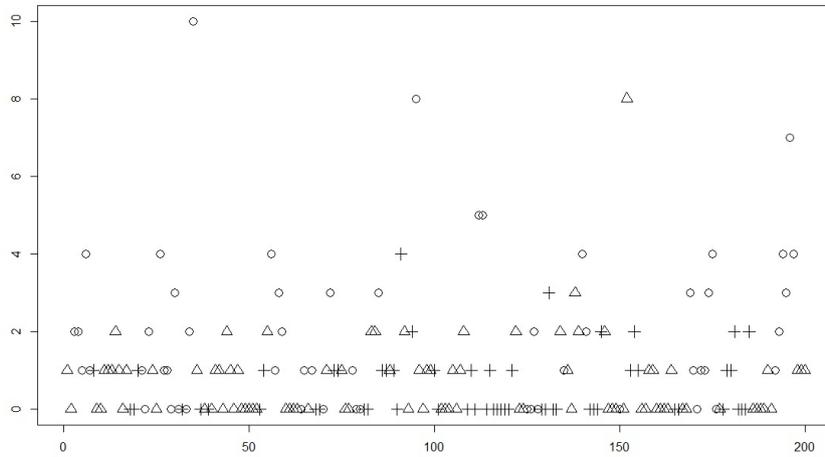


States obtained by standard K-means clustering method

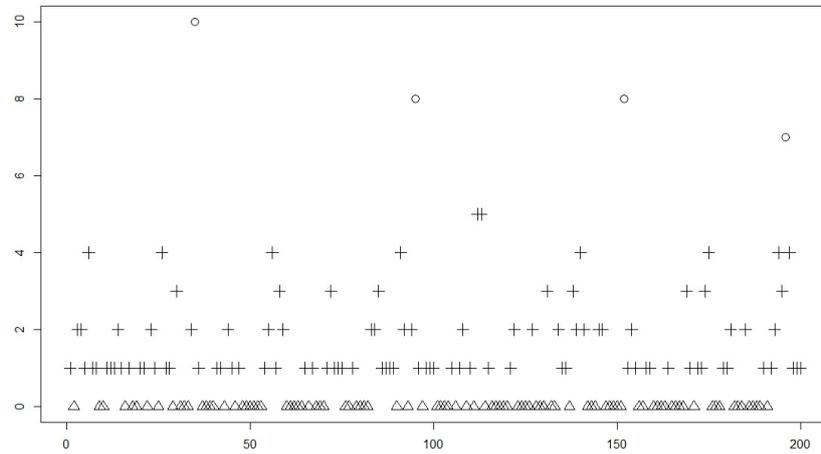


States obtained by *RENES* method for $d_p = 17$, $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_p = (0.4, 0.3)$, $C_m = 9$, $C_a = 7$, $C_p = 2$

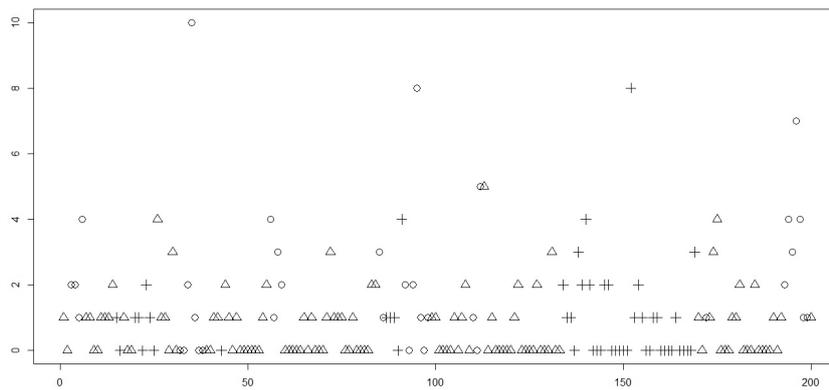
Figure B.7: Environment states of the $R3NGINAR_{max}(2, 4, 2)$ simulation.



Exact states of the $R3NGINAR_1(2, 4, 2)$ simulation



States obtained by standard K-means clustering method



States obtained by *RENES* method for $d_p = 18$, $\mathbf{v}_m = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_a = (0.16, 0.14, 0.14, 0.14)$, $\mathbf{v}_p = (0.4, 0.3)$, $C_m = 6$, $C_a = 1$, $C_p = 8$

Figure B.8: Environment states of the $R3NGINAR_1(2, 4, 2)$ simulation.

model is utilized to reconstruct unused replications. A fitting quality of each reconstruction is measured by calculating the *RMS* of differences between simulated data and modeled data. Such obtained results are provided in Table B.3. Obviously, a benefit of applying the *RENES* method is still perceptible, although significantly lesser than in the case of simulations with 2 environment states. Modeled data based on the standard K-means clustering method produced the following *RMS*-s: $RMS = 1.284$ in the case of $R3NGINAR_{max}(2, 4, 2)$ simulation and $RMS=1.470$ in the case of $R3NGINAR_1(2, 4, 2)$ simulation. On the other hand, modeled data based on the newly proposed *RENES* method provided the following: $RMS=1.148$ in the case of $R3NGINAR_{max}(2, 4, 2)$ simulation and $RMS=1.369$ in the case of $R3NGINAR_1(2, 4, 2)$ simulation. The lesser benefit is expected actually, bearing in mind that none of the clustering methods reached even half of exactly estimated states. Regarding the model parameters, estimates are quite similar actually. An exception is the vector of means, which is more accurately estimated after the *RENES* clustering is performed.

Table B.3: *RMS* values and *CML* parameter estimates obtained after reconstruction of unused data sequences corresponding to the $R3NGINAR_{max}(2, 4, 2)$ and $R3NGINAR_1(2, 4, 2)$ time series.

	$R3NGINAR_{max}(2, 4, 2)$		$R3NGINAR_1(2, 4, 2)$	
Clustering	<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
Regular K-means	$\widehat{\mathcal{M}} = (0.323, 2.307, 5.277)$ $\widehat{\mathcal{A}} = (0.052, 0.187, 0.218)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.893 & 0.107 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 \\ 0.329 & 0.332 & 0.339 & 0 \\ 0.221 & 0.201 & 0.241 & 0.337 \end{bmatrix}$ $\widehat{\phi}_3 = \begin{bmatrix} 1 & 0 \\ 0.001 & 0.999 \end{bmatrix}$	1.284	$\widehat{\mathcal{M}} = (0.522, 3.557, 7.778)$ $\widehat{\mathcal{A}} = (0.051, 0.202, 0.436)$ $\widehat{\phi}_1 = (0.999, 0.001)$ $\widehat{\phi}_2 = (0.248, 0.202, 0.240, 0.310)$ $\widehat{\phi}_3 = (0.001, 0.999)$	1.470
<i>RENES</i>	$\widehat{\mathcal{M}} = (0.502, 1.202, 1.501)$ $\widehat{\mathcal{A}} = (0.198, 0.327, 0.320)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.982 & 0.018 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 \\ 0.325 & 0.333 & 0.342 & 0 \\ 0.242 & 0.202 & 0.241 & 0.315 \end{bmatrix}$ $\widehat{\phi}_3 = \begin{bmatrix} 1 & 0 \\ 0.001 & 0.999 \end{bmatrix}$	1.148	$\widehat{\mathcal{M}} = (0.502, 1.160, 1.448)$ $\widehat{\mathcal{A}} = (0.091, 0.212, 0.367)$ $\widehat{\phi}_1 = (0.753, 0.247)$ $\widehat{\phi}_2 = (0.255, 0.210, 0.248, 0.287)$ $\widehat{\phi}_3 = (0.140, 0.860)$	1.369

An effectiveness of the introduced *RENES* method is additionally confirmed by utilizing $R3NGINAR_{max}(2, 4, 5)$ and $R3NGINAR_1(2, 4, 5)$ simulations, based on the second combination of model parameters from Appendix A. Each simulation is created in two replications. The first replication is used to provide the *RENES* method parameters. The same procedure described in previous paragraphs is followed again. Table B.4 offers such obtained optimal values of d_p , \mathbf{v}_m , \mathbf{v}_a , \mathbf{v}_p , C_m , C_a and C_p .

Optimally selected parameters d_p , \mathbf{v}_m , \mathbf{v}_a , \mathbf{v}_p , C_m , C_a and C_p make the *RENES* method totally ready to use. Alongside with K-means, *RENES* is applied on unused replications. Reconstructions of those unused replications have taken place for each clustering result, using corresponding $R3NGINAR_{max}(2, 4, 5)$ or $R3NGINAR_1(2, 4, 5)$ model. A fitting quality of each reconstruction is measured in the same way as earlier and presented in

Table B.4: Optimal values of the *RENES* method parameters related to the *R3NGINAR*(2, 4, 5) time series.

<i>R3NGINAR</i> _{max} (2, 4, 5)						
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p	C_m	C_a	C_p
12	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.4,0.3)	10	3	1
<i>R3NGINAR</i> ₁ (2, 4, 5)						
d_p	\mathbf{v}_m	\mathbf{v}_a	\mathbf{v}_p	C_m	C_a	C_p
11	(0.16,0.14,0.14,0.14)	(0.16,0.14,0.14,0.14)	(0.4,0.3)	7	5	2

Table B.5. A level of benefit perceived after application of the *RENES* method is trully satisfactory, bearing in mind the benefits given in Table B.3. Benefits are mostly generated by more accurate estimates of the mean values.

Table B.5: *RMS* values and *CML* parameter estimates obtained after reconstruction of unused data sequences corresponding to the *R3NGINAR*_{max}(2, 4, 5) and *R3NGINAR*₁(2, 4, 5) time series.

Clustering	<i>R3NGINAR</i> _{max} (2, 4, 5)		<i>R3NGINAR</i> ₁ (2, 4, 5)	
	<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
Regular K-means	$\widehat{\mathcal{M}} = (0.759, 4.419, 11.409)$ $\widehat{\mathcal{A}} = (0.176, 0.301, 0.298)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.009 & 0.991 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.967 & 0.033 & 0 & 0 \\ 0.297 & 0.301 & 0.402 & 0 \\ 0.298 & 0.301 & 0.301 & 0.100 \end{bmatrix}$ $\widehat{\phi}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 & 0 \\ 0.398 & 0.398 & 0.204 & 0 & 0 \\ 0.298 & 0.301 & 0.298 & 0.103 & 0 \\ 0.198 & 0.198 & 0.200 & 0.202 & 0.202 \end{bmatrix}$	2.108	$\widehat{\mathcal{M}} = (0.769, 5.059, 14.624)$ $\widehat{\mathcal{A}} = (0.163, 0.201, 0.301)$ $\widehat{\phi}_1 = (0.008, 0.992)$ $\widehat{\phi}_2 = (0.303, 0.301, 0.300, 0.096)$ $\widehat{\phi}_3 = (0.201, 0.201, 0.200, 0.203, 0.195)$	2.352
<i>RENES</i>	$\widehat{\mathcal{M}} = (2.497, 4.498, 6.510)$ $\widehat{\mathcal{A}} = (0.297, 0.301, 0.423)$ $\widehat{\phi}_1 = \begin{bmatrix} 1 & 0 \\ 0.001 & 0.999 \end{bmatrix}$ $\widehat{\phi}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.478 & 0.522 & 0 & 0 \\ 0.297 & 0.298 & 0.405 & 0 \\ 0.297 & 0.301 & 0.298 & 0.104 \end{bmatrix}$ $\widehat{\phi}_3 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.001 & 0.999 & 0 & 0 & 0 \\ 0.398 & 0.400 & 0.202 & 0 & 0 \\ 0.298 & 0.301 & 0.301 & 0.100 & 0 \\ 0.198 & 0.198 & 0.200 & 0.202 & 0.202 \end{bmatrix}$	1.718	$\widehat{\mathcal{M}} = (2.368, 4.425, 6.494)$ $\widehat{\mathcal{A}} = (0.161, 0.197, 0.472)$ $\widehat{\phi}_1 = (0.448, 0.552)$ $\widehat{\phi}_2 = (0.293, 0.197, 0.260, 0.250)$ $\widehat{\phi}_3 = (0.238, 0.161, 0.250, 0.162, 0.189)$	1.976

Appendix C. Modeling results obtained using various *INAR* models with stationary or non-stationary nature

Results of applying numerous *INAR* models with stationary or non-stationary nature can be found within this appendix. Many of them have been already introduced in Chapter 1.

Like $R2NGINAR(2, 4)$ and $R2NGINAR(2, 5)$, all these models are applied to the data that represents the number of newly detected COVID-19 cases on daily basis in Mauritius between March 18, 2020 and April 25, 2021. For each model, RMS of differences between real data and modeled data is calculated. As for stationary models, the following options are considered: $PoINAR(1)$ model given in [2], $GPQINAR(1)$ model from [4], $GINAR(1)$ model presented in [3], $NGINAR(1)$ model described in [43], $NGINAR(p)$ model ($p = 2, 3, 4, 5$) shown in [40] and $NBRCINAR(1)$ defined by [50]. Beside these, several non-stationary models are taken into account: $R2NGINAR(1)$ model discussed in [38] and $R2NGINAR(p)$ models given in [39], whereby $p = 2, 3, 4, 5$.

Table C.1 and Table C.2 offer RMS values calculated for each given model. RMS -s that correspond to stationary models are presented in Table C.1. In addition, the non-stationary $R2NGINAR(1)$ model is also included in this table. Table C.2 provides RMS -s acquired for $R2NGINAR_{max}(p)$ and $R2NGINAR_1(p)$ models of various orders. As can be seen, stationary models provide the highest RMS values. A significant progress is achieved by involving the concept of random environment with two environment states. This supports a hypothesis that observed time series really took place in two environment states. However, neither $R2NGINAR(1)$ model nor $R2NGINAR(p)$ models (for $p = 2, 3, 4, 5$) managed to achieve lower RMS -s than those provided by $R2NGINAR(2, 5)$ (see Table 4.6). Hence, it can be concluded that $R2NGINAR(\mathcal{M}, \mathcal{A}, \mathcal{P})$ models are indeed the best for chosen real-life data.

Table C.1: RMS values and CML parameter estimates obtained after reconstruction of the chosen real-life data using various $INAR$ models.

Model	CML	RMS	Model	CML	RMS
$PoINAR(1)$	$\hat{\lambda} = 2.061$ $\hat{\alpha} = 0.308$	6.903	$GPQINAR(1)$	$\hat{\lambda} = 0.421$ $\hat{\theta} = 0.824$ $\hat{\rho} = 0.196$	7.096
$GINAR(1)$	$\hat{q} = 0.828$ $\hat{\alpha} = 0.285$	7.027	$NGINAR(1)$	$\hat{\mu} = 4.572$ $\hat{\alpha} = 0.366$	6.922
$NGINAR(2)$	$\hat{\mu} = 4.572$ $\hat{\alpha} = 0.013$ $\hat{p} = 0.183$	8.637	$NGINAR(3)$	$\hat{\mu} = 4.572$ $\hat{\alpha} = 0.012$ $\hat{p} = 0.185$	8.678
$NGINAR(4)$	$\hat{\mu} = 4.572$ $\hat{\alpha} = 0.012$ $\hat{p} = 0.140$	8.680	$NGINAR(5)$	$\hat{\mu} = 4.572$ $\hat{\alpha} = 0.017$ $\hat{p} = 0.138$	8.670
$NBRCINAR(1)$	$\hat{p} = 0.153$ $\hat{\rho} = 0.492$ $\hat{n} = 0.512$	7.260	$RrNGINAR(1)$	$\hat{\mathcal{M}} = (1.843, 10.946)$ $\hat{\alpha} = 0.144$	5.744

Table C.2: *RMS* values and *CML* parameter estimates obtained after reconstruction of the chosen real-life data using $R2NGINAR_{max}$ and $R2NGINAR_1$ models of various orders.

$R2NGINAR_{max}(2)$		$R2NGINAR_1(2)$	
<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
$\widehat{\mathcal{M}} = (0.911, 7.203)$ $\widehat{\alpha} = 0.106$ $\widehat{\phi} = \begin{bmatrix} 1 & 0 \\ 0.426 & 0.574 \end{bmatrix}$	6.407	$\widehat{\mathcal{M}} = (0.912, 7.203)$ $\widehat{\alpha} = 0.106$ $\widehat{\phi} = (0.426, 0.574)$	6.408
$R2NGINAR_{max}(3)$		$R2NGINAR_1(3)$	
<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
$\widehat{\mathcal{M}} = (0.722, 30.148)$ $\widehat{\alpha} = 0.008$ $\widehat{\phi} = \begin{bmatrix} 1 & 0 & 0 \\ 0.998 & 0.002 & 0 \\ 0.246 & 0.366 & 0.388 \end{bmatrix}$	4.205	$\widehat{\mathcal{M}} = (0.832, 29.111)$ $\widehat{\alpha} = 0.009$ $\widehat{\phi} = (0.351, 0.438, 0.211)$	4.203
$R2NGINAR_{max}(4)$		$R2NGINAR_1(4)$	
<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
$\widehat{\mathcal{M}} = (0.703, 30.181)$ $\widehat{\alpha} = 0.008$ $\widehat{\phi} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0.969 & 0.031 & 0 & 0 \\ 0.331 & 0.388 & 0.281 & 0 \\ 0.332 & 0.234 & 0.432 & 0.002 \end{bmatrix}$	4.211	$\widehat{\mathcal{M}} = (0.709, 30.007)$ $\widehat{\alpha} = 0.008$ $\widehat{\phi} = (0.264, 0.451, 0.130, 0.155)$	4.201
$R2NGINAR_{max}(5)$		$R2NGINAR_1(5)$	
<i>CML</i>	<i>RMS</i>	<i>CML</i>	<i>RMS</i>
$\widehat{\mathcal{M}} = (0.661, 30.180)$ $\widehat{\alpha} = 0.008$ $\widehat{\phi} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0.821 & 0.179 & 0 & 0 & 0 \\ 0.331 & 0.394 & 0.275 & 0 & 0 \\ 0.238 & 0.195 & 0.334 & 0.233 & 0 \\ 0.001 & 0.245 & 0.161 & 0.345 & 0.248 \end{bmatrix}$	4.172	$\widehat{\mathcal{M}} = (0.668, 30.178)$ $\widehat{\alpha} = 0.008$ $\widehat{\phi} = (0.257, 0.433, 0.130, 0.141, 0.039)$	4.168

Biography

Bogdan Pirković was born in Kragujevac, Serbia, on September 1st, 1988. He graduated from elementary school "Milutin i Draginja Todorović" in Kragujevac, Serbia, class of 2003, as a bearer of Vuk Karadžić award (perfect GPA and awards in mathematics competitions). He graduated from high school "Prva kragujevačka gimnazija" in Kragujevac, class of 2007, module for mathematics and natural sciences.

On fall 2007, Bogdan enrolled in Bachelor programme in "Mathematics" at Faculty of Science, University of Kragujevac, and graduated from the university in September 2010, with the GPA 8,52. On fall, he enrolled in Master programme in Mathematics, majoring in theoretical mathematics, at the same faculty and passed all the exams and finals provided by the curriculum in two years period, with the GPA 9.69. In academic 2017/2018, Bogdan enrolled in PhD programme within Doctoral School of Mathematics, module analysis, at Faculty of Science in Kragujevac and passed all the exams provided by the curriculum with the GPA 9.86. He has broadened his field of research to mathematical statistics, statistical modeling and their applications. So far, he has published two scientific papers related to his doctoral dissertation field of research in journals of international importance. In addition, one more paper related to the same field of research is waiting to appear.

During academic 2013/2014, Bogdan was engaged in the high school "Prva kragujevačka gimnazija" in working with students of the special mathematics department. Since October 2014, he has been employed as a teaching assistant at Department of Mathematics and Informatics, Faculty of Science, University of Kragujevac, where he has participated in teaching in several subjects in the field of probability and statistics.

List of publications:

1. Djordjević, M. S., Ristić, M. M., **Pirković, B. A.** (2021) Identifying latent components of the TINAR(1) model, *Filomat* 35(13), 4469-4482. (ISSN:0354-5180, **M22**, **IF2020**=0.844, 197/330, **Category**: Mathematics)
DOI:10.2298/FIL2113469D
2. **Pirković, B. A.**, Laketa, P. N., Nastić, A. S. (2021) On generalized random environment *INAR* models of higher order: estimation of random environment states, *Filomat* 35(13), 4545-4576. (ISSN:0354-5180, **M22**, **IF2020**=0.844, 197/330, **Category**: Mathematics)
DOI:10.2298/FIL2113545P

Conferences:

1. **Pirković, B. A.**, Laketa, P. N., Nastić, A. S. (2021) Random environment estimation method for generalized random environment *INAR* models of higher order, International Workshop on Nonlinear Analysis and its Applications, Niš, October 13-16, 2021. (Book of abstracts-p. 40)

ИЗЈАВА АУТОРА О ОРИГИНАЛНОСТИ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Ја, Богдан А. Пирковић, изјављујем да докторска дисертација под насловом:

CONTRIBUTION TO THE THEORY OF RANDOM ENVIRONMENT
INTEGER-VALUED AUTOREGRESSIVE PROCESSES

која је одбрањена на Природно-математичком факултету
Универзитета у Крагујевцу представља *оригинално ауторско дело* настало као резултат *сопственог истраживачког рада*.

Овом Изјавом такође потврђујем:

- да сам *једини аутор* наведене докторске дисертације,
- да у наведеној докторској дисертацији *нисам извршио/ла повреду* ауторског нити другог права интелектуалне својине других лица,
- да умножени примерак докторске дисертације у штампаној и електронској форми у чијем се прилогу налази ова Изјава садржи докторску дисертацију истоветну одбрањеној докторској дисертацији.

У Крагујевцу, 1.6.2022. године,

Богдан Пирковић
потпис аутора

ИЗЈАВА АУТОРА О ИСКОРИШЋАВАЊУ ДОКТОРСКЕ ДИСЕРТАЦИЈЕ

Ја, Богдан А. Пировић,

дозвољавам

не дозвољавам

Универзитетској библиотеци у Крагујевцу да начини два трајна умножена примерка у електронској форми докторске дисертације под насловом:

CONTRIBUTION TO THE THEORY OF RANDOM ENVIRONMENT
INTEGER-VALUED AUTOREGRESSIVE PROCESSES

која је одбрањена на Природно-математичком факултету

Универзитета у Крагујевцу, и то у целини, као и да по један примерак тако умножене докторске дисертације учини трајно доступним јавности путем дигиталног репозиторијума Универзитета у Крагујевцу и централног репозиторијума надлежног министарства, тако да припадници јавности могу начинити трајне умножене примерке у електронској форми наведене докторске дисертације путем *преузимања*.

Овом Изјавом такође

дозвољавам

не дозвољавам¹

¹ Уколико аутор изабере да не дозволи припадницима јавности да тако доступну докторску дисертацију користе под условима утврђеним једном од *Creative Commons* лиценци, то не искључује право припадника јавности да наведену докторску дисертацију користе у складу са одредбама Закона о ауторском и сродним правима.

припадницима јавности да тако доступну докторску дисертацију користе под условима утврђеним једном од следећих *Creative Commons* лиценци:

- 1) Ауторство
- 2) Ауторство - делити под истим условима
- 3) Ауторство - без прерада
- 4) Ауторство - некомерцијално
- 5) Ауторство - некомерцијално - делити под истим условима
- 6) Ауторство - некомерцијално - без прерада²

у Крагујевцу _____, 1.6.2022. године,


потпис аутора

² Молимо ауторе који су изабрали да дозволе припадницима јавности да тако доступну докторску дисертацију користе под условима утврђеним једном од *Creative Commons* лиценци да заокруже једну од понуђених лиценци. Детаљан садржај наведених лиценци доступан је на: [http://creativecommons.org/rs/](http://creativecommons.org.rs/)